ISSN 1304-7191
E-ISSN 1304-7205

sigma

Journal of Engineering and Natural Sciences

Mühendislik ve Fen Bilimleri Dergisi

Volume 42
Number 4
Year 2024

YTU PRESS

www.sigma.yildiz.edu.tr

E-ISSN 1304-7205

sigma
Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Volume 42 Number 4 Year 2024 - August

E-ISSN 1304-7205

sigma
Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Volume 42 Number 4 Year 2024 - August

E-ISSN 1304-7205

Sigma
Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Volume 42 Number 4 Year 2024 - August

Abstracting and Indexing:
- Directory of Open Access Journals (DOAJ)
- EBSCO Host Online Research Databases
- Emerging Sources Citation Index (ESCI)
- Scopus
- MIAR
- Scilit
- OUCI
- Research4Life
- Ideal Online

Journal Description: The journal is supported by Yıldız Technical University officially, and is a blind peer-reviewed free open-access journal, published bimontly (February-April-June-August-October-December).
Owner: Yıldız Technical University
Editor-in-Chief: Ahmet Selim Dalkiliç
Language of Publication: English
Frequency: Bimontly (6 issues per year)
Publication Type: Online e-version
Publisher: Kare Publishing

E-ISSN 1304-7205

sigma
Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Volume 42 Number 4 Year 2024 - August

CONTENTS

E-ISSN 1304-7205

sigma
Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Volume 42 Number 4 Year 2024 - August

E-ISSN 1304-7205

Sigma
Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Volume 42 Number 4 Year 2024 - August

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00082

Research Article

Leaders of change: Children and the climate change movement

Beyza KARACAOĞLU[1], Mehmet Fatih AKBABA[2,*]

[1]*Department of Bioengineering, Yildiz Technical University, 34220, İstanbul, Türkiye*
[2]*Turkish and Foreign Languages Application and Research Center, Yildiz Technical University, 34220, İstanbul, Türkiye*

ARTICLE INFO

Article history
Received: 04 April 2024
Revised: 19 April 2024
Accepted: 04 June 2024

Keywords:
Awareness; Children; Climate Change; Communication; Education

ABSTRACT

Climate change is a global crisis associated with the increase in greenhouse gas emissions and is associated with many environmental, psychological, and health problems today. One of the problems of the past, present, and future is children's limited knowledge about climate change and their inability to take an active role as participants in the struggle. Children are the greatest pioneers of change, and their ability to change the world is undeniable. Children need to know the facts about climate change, make their voices heard in climate crisis management, and be involved in the struggle. In this regard, it is crucial to conduct various sustainability trainings and activities, both in school and after school, using effective communication strategies. These strategies should include explaining and conveying climate change using positive language, utilizing methods such as localization, visualization, and storytelling with children. In the meantime, it is very important that educators, families, and society also gain awareness and become equipped about climate change, and that children take responsibility and become active participants, because the fight against climate change is intergenerational and climate change has mutual effects on attitude and behavior change. The scope of this review is to examine the role of children in combating climate change and to compile literature studies and real-life examples to understand the importance of the subject. This study will make an original contribution to different communication strategies and educational programs that present solutions to the climate crisis, as well as lead the way for families and communities to guide children.

Cite this article as: Karacaoğlu B, Akbaba MF. Leaders of change: Children and the climate change movement. Sigma J Eng Nat Sci 2024;42(4):957–964.

INTRODUCTION

The climate change challenge, which is primarily caused by greenhouse gases and land use, can directly constrain the availability of raw material resources, lead to food insecurity, and consequently have adverse impacts on public health. These implications can arise from both the deterioration of the environment and financial meltdowns [1,2]. For example, Springmann et al. [3] predict that climate change-related deaths attributable to food insecurity alone will reach 529,000 by 2050. Children are experiencing the effects of climate change in the form of asthma [4], atopic

*Corresponding author.
*E-mail address: fatih.akbaba@yildiz.edu.tr

This paper was recommended for publication in revised form by Editor in Chief Ahmet Selim Dalkilic

dermatitis [5], and allergies [6]. Additionally, due to their awareness of the destructive effects of climate change and extreme weather events such as hunger and food insecurity, water scarcity, and emerging diseases [7], and growing up in uncertainty, children also experience anger problems due to their knowledge about this issue and have less effectiveness in finding solutions due to their helplessness, anxiety, and fear [8]. The results suggest that early childhood exposure to climate change impacts is also associated with lifelong migration, possibly for psychosocial reasons [9]. Furthermore, it is worth noting that young children may exhibit distinct thermoregulation mechanisms compared to adults, rendering them more susceptible to variations in ambient conditions. Hence, it is advisable to engage young children in everyday activities to promote sufficient physical fitness and enhance their immunity and psychological well-being [10].

Governments also play a crucial role in implementing policies that address climate change and adequately prioritize the interests of children [11]. Despite their perceived vulnerability to climate change, children actually contribute significantly to sustainable living as part of the biosphere [12]. Children are endowed with excellent adaptability and are considered change agents with curiosity for exploring the future, environmental consciousness, and the potential to spread it around, making them valuable leaders in addressing the climate change crisis [13,14]. Therefore, an approach that evaluates children as change leaders in climate change crisis management will advance societies [15], focusing on fostering their sense of ownership and responsibility rather than concealing the reality of climate change from them. The need for children to have a voice in climate crisis management, to showcase their talents, and to be involved in world organizations should not be ignored [16]. Moreover, children's involvement in the fight against climate change can influence adults' behavior, emotions, and actions on climate change [17].

As evidence of the importance of children and youth in combating climate change, the United Nations hosted the "Youth Climate Summit" in 2019. At this summit, young climate advocates between the ages of 18 and 30 explored ways to fulfill the commitments of the Paris Agreement and provided a platform for them to discuss their ideas with peers and world leaders [18]. In Africa, there are projects and organizations focused on raising awareness and consciousness among youth about climate change, as well as initiatives for climate action among youth, such as training young farmers and agricultural experts in combating climate change through initiatives like the "Climate Smart Agriculture Youth Network" [19–21]. Additionally, there are various educational programs [22], environmental art and craft activities [23] targeting younger children, and initiatives such as Eco-Schools [24] and Nature Play [25] aimed at raising awareness about climate change and sustainability.

This review focuses on the role children play in climate change, analyzing what kids know and perceive about climate change, the communication strategies used to raise awareness and its impact, and successful communication methods on climate change. It also focuses on school and community-based climate change education programs, campaigns, and events for children, and finally, how parents and educators can educate their children about climate change. This review is important because it shows that children play a key role in the fight against climate change, demonstrates the importance of successful communication strategies and education, and creates effective plans to raise awareness and fight against climate change today and in the future.

The objective of this article is to examine the role of children in efforts to tackle the climate crisis, review literature and real-life examples to understand and communicate the importance of the subject to readers and raise awareness of the need for children to take responsibility for community action. Furthermore, this review is constructive in terms of different communication strategies, educational programs, and the role of parents and communities in engaging children in the struggle with climate change.

Awareness and Communication Strategies for Children About Climate Change

Although children are sometimes characterized as naive, innocent, and passive, they should be seen as the creators of the future and agents of change. Scientific research on children should be accelerated, and efforts to support their development should be increased [26]. Numerous studies should be conducted to understand what children know about climate change and how they perceive it. Effective communication methods should be utilized to address their concerns, such as helplessness, anxiety, and fear, and to provide accurate awareness about climate change. Children should be pedagogically guided to deeply understand the concepts of sustainability, climate change, and global warming and to effect change. Their relationship with family, school, and society should be determined, and they should be encouraged to learn [27].

It was observed that children's evaluation of environmentally harmful actions varied according to their age group and their judgments about environmental damage varied according to the target victim (tree, plant, animal, etc.) [28]. Hahn & Garrett [29] emphasize in their research that early children evaluate actions that harm the environment as morally wrong and that moral judgments change between the ages of 3 and 4. Therefore, effective communication methods and climate change education are very important in changing the environmental behaviors and attitudes of early children. These effective communication strategies include localization [30], visualization [31], storytelling [32], and the use of positive and inspirational language [33]. However, in the future, it will be critical to identify and diversify specifically effective environmental

education components and approaches, as well as broaden the scope of targeted environmental behaviors [34]. To support children's participation in the climate challenge, it is necessary to identify communication strategies that promote peace and equality, hope, and a determined action plan to protect children, prioritize children, and leave them with a better future [35].

Given the complexity of comprehending and acquiring knowledge about climate change, particularly during early childhood, it is possible to explain the concept to children by emphasizing ecology and pro-environmental topics. Young children can engage in localization activities through interactive games that emphasize culture, nature, and practical problem-solving. These activities aim to enhance their understanding of climate change in a precise and accessible manner [36]. Studies have shown that it is important to establish local and place-based connections and transfer them to children and use nature-based materials, both in terms of climate change and age-appropriateness [37]. Art has the potential to guide children and support meaning-making through photography, visualization, and metaphor-making [23].

An art-based 15-week after-school program for children to strengthen their climate change sensitivity and activism was examined [38]. This program utilized ecosystems, climate-weather, and climate change education, along with outdoor activities and the photovoice process, to assist children in comprehending the challenges and potential solutions to local issues. At the end of the program, it was observed that the children gained awareness about sustainable transformation, had fun, especially with various teachings and games, and realized their activist skills. Post humanist studies that aim to connect children with nature and raise awareness about climate change focus on the concept of ecological aesthetics [39]. For instance, within the scope of the Climate Change and Me project, children and young people in New South Wales, Australia, took photographs to express their feelings about climate change. This study analyzed children's photographs, demonstrating the application of aesthetic framing and demonstrating how childhood experiences can reveal the potential for new environmental thoughts and behaviors [40].

One way to transition the concept of climate change from problem-based to action-based is to present a wide range of stories of people taking positive action, because the ability to take actions is the precursor to a movement, and presenting real stories to the public is thought to be much more effective for community action [41]. Hence, modern storytelling techniques are considered a creative way to interest children and influence their environmental behavior and sustainable attitudes [42]. In early environmental education, digital storytelling is also believed to stimulate children's emotions [43]. In a study, it was observed that children who were presented with positive role models through stories before playing showed more sustainable behavior during play [44]. Hahn & Garrett [29] also demonstrated in their study that the behaviors of fictional characters in books morally influence children's environmental behaviors and shape their future behavior patterns. The main reasons for this can be attributed to the triggering of emotional connection and empathy in readers through storytelling and the use of a collective discussion approach that is compatible with cultural characteristics [45,46].

Games developed on environmental and climate change issues allow children to visualize the effects of climate change and encourage environmentally sustainable behaviors [47]. A study in a Ukrainian school concluded that playing climate change games with children directly increased their understanding of climate change threats and solutions [48]. Alternative reality games, social media, and games connected to the real world are effective and innovative strategies used in communicating climate change [49]. Games emerge as a standalone motivational tool for combating climate change because they utilize entertainment to reach a wide audience [50]. One example of games designed for young children is Éco Héros, which allows children to protect their animals, engage in forest-friendly crafts, and safeguard marine animals from pollution within a digital game environment. Games such as Fate of the World and Anno 2070 are cited as examples of games focusing on environmental sustainability and climate change, where players simulate the effects of climate change and gain knowledge about topics such as renewable energy sources [51].

These effective communication strategies are critical not only for children's perception and learning about climate change but also for enabling children and youth to voice their concerns about climate change to the world and express themselves properly. One of the most significant examples of the role of children and youth in climate change is Greta Thunberg, who initiated the'school strike for climate' in front of the Swedish parliament in 2018. The event received a variety of positive and negative reactions from society and sparked widespread discussions on social media and various communication channels, with debates on this issue still ongoing [52]. Greta Thunberg's impact on rewriting the roles of children and young people in climate change is inevitable, but the communication language she used has also led to her being the target of criticism. It is clear from this incident that while conducting climate change education and awareness-raising activities for children, the communication strategies and language of the educators are also very important so that the child will tend to communicate in a similar way. In addition to educators, the communication strategies and language used by activists also hold significant importance. Greta Thunberg's inconsistent use of English and Swedish in her speeches can be interpreted as communication errors and misunderstandings. Additionally, her over-dramatization of her speeches has led to negative and pessimistic thoughts among the public, which contrasts with her positive and inspirational approach. Nevertheless, it is an inescapable fact that this event has helped raise global awareness and put pressure on

politicians, so Thunberg's courage and determination have elevated the role of children and young people in climate change to a new level [53,54].

Education and Awareness-Raising Programs

Education programs and activities on climate change, the environment, and sustainability should start at an early age in order to combat climate change, raise awareness in society, ensure that it is permanent and long-term, and leave a sustainable planet for future generations. Thanks to global initiatives such as the UNESCO Education for Sustainable Development Vision [55], which combines children's curiosity about living things with education, and national political support such as the Early Years Learning Framework for Australia [56] early years sustainability education is gaining momentum worldwide, and recently, in December 2023, at the Dubai Conference of the Parties (COP) 28 [57], was carried out and decided to transition to a greener education. When education on combating climate change is started at a young age, especially, it shows that children who grow up in close contact with nature become individuals who are aware of their responsibilities towards the environment and exhibit environmentally friendly and respectful behaviors towards nature in their later years [58].

Exposure to nature is often associated with a greater awareness of environmental harm [28]. Therefore, it is essential to integrate nature-based education with basic subjects such as literacy and mathematics in order to promote both sustainable living and ethical practices in the face of the climate change crisis [59]. Ginsburg & Audley [59], conducted a study on the content of sustainable education in a nature-based preschool in the Northeast region of the United States and the behavior of teachers towards children. As a result of the research, it was emphasized that it is important for children to receive sustainability education in nature and to integrate the water cycle, trees and nature with literacy and mathematics skills. However, it was reported that the education given to children should not be limited to school, there is a need for political and cultural systemic change, and in this way, permanent learning can be achieved in the fight against climate change. Children who participated in an after-school activity intended to promote children's behavior regarding climate change and the environment at the household and community levels reduced their energy consumption and waste and shared their knowledge of climate change with their family and other kids their age, thereby facilitating intergenerational learning, according to a study [60]. In their study, Jones & Davison [61] emphasized that climate change education should not only involve learning facts but also encompass children's emotional responses. This is because children's learning pathways focus on sensory and emotional perceptions, unlike adults' abstract and logical ways [62]. It was also underlined that emotions related to climate change should not be categorized as positive or negative, and that children should be allowed to experience a variety of emotions on this issue. According to another study [63], educating young children to observe the weather every morning in schools and experience how their clothing and outdoor play routines can alter based on the weather, thereby linking human and weather events with sensory and affective effects, serves as an educational strategy to counteract and increase awareness of the impacts of climate change. The idea that accessible information can only become permanent through emotion informs this approach.

Successful education and awareness-raising of children about climate change are associated with appropriate educational methods, as well as educators possessing adequate knowledge, skills, and metacognitive and pedagogical abilities in environmental, sustainability, and climate change issues [64]. All teachers and school administrations have responsibilities beyond merely educating students; they should also create environments for discussions on climate change, make students feel empowered to make a difference, and promote student activism and interstudent communication and connections [15]. It is important to note that educational methods and communication strategies may also vary according to the cultures of educators and children. For example, in a study examining early childhood sustainability education in Japan, Australia and Korea [65], Australian and Korean educators focused on conservation of natural resources and environmental issues, while Japanese educators were reluctant to do so. The study also highlighted the importance of globalizing sustainability education and providing educators with standardized, quality, and professional training on climate change. A study of pre-school and primary school teachers in Greece found that gender and age affect educators' performance and environmental attitudes, that they mostly use the media to learn about environmental issues, and that school curricula are incomplete [66]. Although the importance of climate change is evident in research, there is still a need for research and support for climate change education and awareness-raising programs due to limited programs to train educators at the appropriate level, insufficient time allocated to climate change, superficial coverage of climate change issues, and a lack of adequate financial and moral support [67]. Educators should have access to professional development opportunities that foster a deeper understanding of sustainability and its implementation in the school environment. These opportunities should extend beyond merely spending time in the garden or relying solely on books for instruction, but should also introduce more innovative ideas [65].

The Role of Family and Community

Children and families play an important role in creating conscious, fair, and peaceful societies [27]. Children's attitudes and actions towards mitigating climate change are influenced by changes in their daily activities within the family and the debates they observe within the family and in society [60]. The environmental perspectives of

parents or other family members have a significant impact on children's comprehension of climate change and future actions because family decisions impact children and children mimic their parents' education and worldview [14]. Moreover, not only the family influences the children, but also the children influence the family, as evidenced by the fact that children have been shown to influence parents' views on gender orientation [68]. Therefore, communication between family and child is very important and they influence each other through intergenerational learning.

Active listening, demonstrating empathy, spending meaningful time together, imparting problem-solving skills, maintaining transparency and honesty, and setting clear boundaries are the foundations of effective communication between families and children [69–71]. Raising awareness and educating children about climate change can be challenging for families, but mutual benefits can be achieved through various activities, such as the whole family participating in gardening projects together, caring for plants at home, consuming food grown by the family, recycling at home, and encouraging the reuse of recycled materials [27]. Furthermore, researchers determined the impact of The Little Explorers Playgroup, an Australian playgroup that allows children and parents to participate together, on their environmental attitudes and behaviors. The study found that parents and children who participated in the playgroup exhibited stronger environmental attitudes, behaviors, and responsibilities than those who did not [58].

Every action in society has a vital role in promoting intellectual curiosity among children and young people, increasing their knowledge of the urgent requirement to address climate change, and empowering them to take on socio-political responsibilities [72,73]. As emphasized in the rest of the article, there is intergenerational commitment in the fight against climate change, and studies underscore the importance of children making real and potential contributions to creating sustainable futures [74]. Encouraging children and young people to play an active role in climate change movements, strikes, and organizations to combat climate change, and ensuring that their voices are heard by the government and society—all these actions seem to increase and decrease the feelings of hope and optimism for the future in children and even adults [75]. For instance, Plant-for-the-Planet is an organization that began in 2007 as a fourth-grade school project with the goal of planting 1 million trees in every country, where children around the world work to plant forests and stop deforestation [76]. School strikes and the Fridays for Future movement that have occurred around the world since 2018 also enable children and young people to actively participate in society [77].

CONCLUSION

Climate change is one of the most fundamental problems of today and the future, and it requires an urgent action plan. Sustainable development plans and actions continue to be implemented worldwide to support efforts to use renewable energy resources and reduce greenhouse gas emissions in the fight against climate change. However, children, as future creators, also play an important role in this struggle to make these scientific, social, and awareness studies permanent. Fostering an early education that educates children about climate change and cultivates a sense of accountability regarding this matter will establish a foundation for subsequent generations to inhabit more environmentally conscious societies. Providing children with education about climate change and fostering a sense of responsibility from an early age will pave the way for more informed societies and sustainable environments for future generations. Children can learn about sustainability through various communication strategies like nature-based education, books, and games. In this process, educators, families, and society play a crucial role, and effective communication with children requires a solid foundation in sustainable practices. Raising awareness and educating children about climate change opens the door for them to take action towards a sustainable future. This movement can also occur through changes in their attitudes and thought processes within the home, as well as through their contributions to global reforestation efforts or future career choices. Therefore, it is crucial to establish climate change communication in children, implement effective communication strategies, and provide them with relevant training, as the lessons they learn at a young age will shape their future. In addition, it is anticipated that awareness will increase among parents and society who witness children's struggle with climate change due to intergenerational commitment. It is recommended that sustainability practices be expanded and repeated with innovative perspectives in order to obtain more accurate results about the role of children in climate change and the training provided. Lastly, it is crucial to adopt an international approach and raise local people's awareness in children's climate change awareness studies, as well as in all social organizations in which children participate

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] McCool WC, Anderson AS, Baide AJ, Gonzalez T, Codding BF. Evaluating the relationships between climate change, population pressure, economic intensification, and childhood stress in the Prehispanic Nasca region of Peru. Quaternary Int 2024;689-690:120–134. [CrossRef]

[2] Venegas Hargous C, Strugnell C, Allender S, Orellana L, Corvalan C, Bell C. Double- and triple-duty actions in childhood for addressing the global syndemic of obesity, undernutrition, and climate change: A scoping review. Obesity Rev 2023;24:e13555. [CrossRef]

[3] Springmann M, Mason-D'Croz D, Robinson S, Garnett T, Godfray HCJ, Gollin D, et al. Global and regional health effects of future food production under climate change: a modelling study. Lancet 2016;387:1937–1946. [CrossRef]

[4] Sheffield PE, Knowlton K, Carr JL, Kinney PL. Modeling of regional climate change effects on ground-level ozone and childhood asthma. Am J Prevent Med 2011;41:251–257. [CrossRef]

[5] Wang, SP, Stefanovic N, Orfali RL, Aoki V, Brown SJ, Dhar S, et al. Impact of climate change on atopic dermatitis: A review by the International Eczema Council. Allergy 2024;79:1455–1469. [CrossRef]

[6] Wright RJ. Influences of climate change on childhood asthma and allergy risk. Lancet Child Adolesc Health 2020;4:859–860. [CrossRef]

[7] Hanna R, Oliva P. Implications of climate change for children in developing countries. Future Child 2016;26:115–132. [CrossRef]

[8] Martin G, Reilly KC, Gilliland JA. Impact of awareness and concerns of climate change on children's mental health: a scoping review protocol. JBI Evid Synth 2020;18:516. [CrossRef]

[9] Thiede BC, Randell H, Gray C. The childhood origins of climate-induced mobility and immobility. Popul Dev Rev 2022;48:767–793. [CrossRef]

[10] Morrison SA. Moving in a hotter world: Maintaining adequate childhood fitness as a climate change countermeasure. Temperature 2023;10:179–197. [CrossRef]

[11] Currie J, Deschênes O. Children and climate change: Introducing the issue. Future Child 2016;26:3–9. [CrossRef]

[12] Lee N. Childhood and Biopolitics: Climate Change, Life Processes and Human Futures. Manhattan, New York City: Springer; 2013. p. 190.

[13] Godden NJ, Farrant BM, Yallup Farrant J, Heyink E, Carot Collins E, Burgemeister B, et al. Climate change, activism, and supporting the mental health of children and young people: Perspectives from Western Australia. J Paediatr Child Health 2021;57:1759–1764. [CrossRef]

[14] Hahn ER. The developmental roots of environmental stewardship: Childhood and the climate change crisis. Curr Opin Psychol 2021;42:19–24. [CrossRef]

[15] Grauer SR. Climate change: The thief of childhood. Phi Delta Kappan 2020;101:42–46. [CrossRef]

[16] Zhanda K, Dzvimbo MA, Chitongo L. Children climate change activism and protests in africa: reflections and lessons from greta thunberg. Bullet Sci Technol Soc 2021;41:87–98. [CrossRef]

[17] Hickman C. Children and Climate Change: Exploring Children's Feelings About Climate Change Using Free Association Narrative Interview Methodology. In: Hoggett P, editor. Climate Psychology: On Indifference to Disaster. Cham: Springer International Publishing; 2019. p. 41–59. [CrossRef]

[18] Han H, Ahn SW. Youth mobilization to stop global climate change: Narratives and impact. Sustainability 2020;12:4127. [CrossRef]

[19] Mungai C, Muchaba T, Szilagyi L, Radeny MAO, Atakos V, Ntiokam D. Youth Engagement in Climate-Smart Agriculture in Africa: Opportunities and Challenges. 2018 May 31; Available from: https://hdl.handle.net/10568/92979 Accessed on Jun 04, 2024.

[20] Harris C. Looking to the future? Including children, young people and future generations in deliberations on climate action: Ireland's Citizens'Assembly 2016-2018. Innov Eur J Sci Res 2021;34:677–693. [CrossRef]

[21] Hilder C, Collin P. The role of youth-led activist organisations for contemporary climate activism: the case of the Australian Youth Climate Coalition. J Youth Stud 2022;25:793–811. [CrossRef]

[22] Larson LR, Castleberry SB, Green GT. Effects of an environmental education program on the environmental orientations of children from different gender, age, and ethnic groups. J Park Recreat Adm 2010;28:95–113.

[23] Bentz J. Learning about climate change in, with and through art. Clim Change 2020;162:1595–1612. [CrossRef]

[24] Cincera J, Kroufek R, Simonova P, Broukalova L, Broukal V, Skalík J. Eco-School in kindergartens: the effects, interpretation, and implementation of a pilot program. Environ Educ Res 2017;23:919–936. [CrossRef]

[25] Leopardi V, Chang YM, Pham A, Luo J, Garden OA. A Systematic review of the potential implication of infectious agents in myasthenia gravis. Front Neurol 2021;12:618021. [CrossRef]

[26] Raby R, Sheppard LC. Constructs of childhood, generation and heroism in editorials on young people's climate change activism: Their mobilisation and effects. Child Soc 2021;35:380–394. [CrossRef]

[27] MacDonald M. Early childhood education and sustainability: A living curriculum. Child Educ 2015;91:332–341. [CrossRef]
[28] Collado S, Sorrel MA. Children's environmental moral judgments: Variations according to type of victim and exposure to nature. J Environ Psychol 2019;62:42–48. [CrossRef]
[29] Hahn ER, Garrett MK. Preschoolers' moral judgments of environmental harm and the influence of perspective taking. J Environ Psychol 2017;53:11–19. [CrossRef]
[30] Richards DP. Not a cape, but a life preserver: the importance of designer localization in interactive sea level rise viewers. Commun Des Q Rev 2018;6:57–69. [CrossRef]
[31] Lumley S, Sieber R, Roth R. A framework and comparative analysis of web-based climate change visualization tools. Comput Graph 2022;103:19–30. [CrossRef]
[32] Fish C. Storytelling for Making Cartographic Design Decisions for Climate Change Communication in the United States. Cartographica 2020;55:69–84. [CrossRef]
[33] Schneider CR, Zaval L, Markowitz EM. Positive emotions and climate change. Curr Opin Behav Sci 2021;42:114–120. [CrossRef]
[34] van de Wetering J, Leijten P, Spitzer J, Thomaes S. Does environmental education benefit environmental outcomes in children and adolescents? A meta-analysis. J Environ Psychol 2022;81:101782. [CrossRef]
[35] Balvin N, Christie DJ, editors. Children and Peace: From Research to Action. Cham: Springer International Publishing; 2020. [CrossRef]
[36] Smith GA. Place-based education: Learning to be where we are. Phi Delta Kappan 2002;83:584–594. [CrossRef]
[37] Beaver BC, Borgerding LA. Climate Change Education in Early Childhood Classrooms: A Nature-Based Approach. Int J Early Child Environ Educ 2023;11:3.
[38] Trott CD. Reshaping our world: Collaborating with children for community-based climate change action. Action Res 2019;17:42–62. [CrossRef]
[39] Rousell D, Cutter-Mackenzie-Knowles A. Uncommon Worlds: Toward an Ecological Aesthetics of Childhood in the Anthropocene. In: Cutter-Mackenzie-Knowles A, Malone K, Barratt Hacking E, editors. Research Handbook on Childhoodnature : Assemblages of Childhood and Nature Research. Cham: Springer International Publishing; 2020 p. 1657–1679. [CrossRef]
[40] Cutter-Mackenzie A, Rousell D. Education for what? Shaping the field of climate change education with children and young people as co-researchers. Child Geogr 2019;17:90–104. [CrossRef]
[41] Meyer KD, Coren E, McCaffrey M, Slean C. Transforming the stories we tell about climate change: from 'issue' to 'action'. Environ Res Lett 2020;16:015002. [CrossRef]
[42] Vaughan-Lee C. The Power of Immersive Storytelling: A tool for transformative learning. Child Educ 2019;95:23–31. [CrossRef]
[43] Byman J, Kumpulainen K, Wong CC, Renlund J. Children's emotional experiences in and about nature across temporal-spatial entanglements during digital storying. Literacy 2022;56:18–28. [CrossRef]
[44] Ebersbach M, Brandenburger I. Reading a short story changes children's sustainable behavior in a resource dilemma. J Exp Child Psychol 2020;191:104743. [CrossRef]
[45] Bloomfield EF, Manktelow C. Climate communication and storytelling. Clim Change 2021;167:34. [CrossRef]
[46] Daigle JJ, Michelle N, Ranco DJ, Emery MR. Traditional lifeways and storytelling: tools for adaptation and resilience to ecosystem Change. Hum Ecol 2019;47:777–784. [CrossRef]
[47] Douglas BD, Brauer M. Gamification to prevent climate change: a review of games and apps for sustainability. Curr Opin Psychol 2021;42:89–94. [CrossRef]
[48] Khalaim O. Climate Change Games as an Effective Tool for ESD Practices. Stud. Perieget 2017.
[49] Wu JS, Lee JJ. Climate change games as tools for education and engagement. Nat Clim Change 2015;5:413–418. [CrossRef]
[50] Mazur-Stommen S, Farley K. Games for Grownups: The Role of Gamification in Climate Change and Sustainability. Indicia Consulting LLC. 2016;405.
[51] Abraham BJ, Jayemanne D. Where are all the climate change games? Locating digital games' response to climate change. 2017 Nov 8; Available from: https://opus.lib.uts.edu.au/handle/10453/121664
[52] Spyrou S. Children as future-makers. Childhood 2020;27:3–7. [CrossRef]
[53] Sabherwal A, Ballew MT, van Der Linden S, Gustafson A, Goldberg MH, Maibach EW, et al. The Greta Thunberg Effect: Familiarity with Greta Thunberg predicts intentions to engage in climate activism in the United States. J Appl Soc Psychol 2021;51:321–333. [CrossRef]
[54] Haugseth JF, Smeplass E. The Greta Thunberg effect: A study of Norwegian youth's reflexivity on climate change. Sociology 2023;57:921–939. [CrossRef]
[55] Catana MM, Brilha JB. The role of UNESCO global geoparks in promoting geosciences education for Sustainability. Geoheritage 2020;12:1. [CrossRef]
[56] Cheeseman S, Sumsion J, Press F. Infants of the knowledge economy: the ambition of the Australian Government's Early Years Learning Framework. Pedagogy Cult Soc 2014;22:405–424. [CrossRef]

[57] Kidman G, Chang CH. Sustainability education: meeting the demands of climate change aspirations Gillian Kidman and Chew-Hung Chang. Int Res Geograph Environ Educ 2024;33:1–5. [CrossRef]
[58] Mintoff Z, Andersen P, Warren J, Elliott S, Nicholson C, Byfield-Fleming H, et al. The Effectiveness of a Community-Based Playgroup in Inspiring Positive Changes in the Environmental Attitudes and Behaviours of Children and their Parents: A Qualitative Case Study. Aust J Environ Educ 2024;40:22–34. [CrossRef]
[59] Ginsburg JL, Audley S. "You don't wanna teach little kids about climate change": Beliefs and Barriers to Sustainability Education in Early Childhood. Int J Early Child Environ Educ 202;73:42.
[60] Trott CD. Youth-led climate change action: multi-level effects on children, families, and communities. Sustainability 2021;13:12355. [CrossRef]
[61] Jones CA, Davison A. Disempowering emotions: The role of educational experiences in social responses to climate change. Geoforum 2021;118:190–200. [CrossRef]
[62] Rule AC, Zhbanova KS. Guardians of the Earth: Teaching Children to Care for All Living Things. In: Renck Jalongo M, editor. Teaching Compassion: Humane Education in Early Childhood [Internet]. Dordrecht: Springer Netherlands; 2014. p. 197–211. [CrossRef]
[63] Rooney T. Weather worlding: learning with the elements in early childhood. Environ Educ Res 2018;24:1–12. [CrossRef]
[64] Andrea V, Petkou D. Exploring the attitudes and views of pre-primary and primary school teachers for climate change education. J Int Bus Entrep Dev 2022;14:287–303. [CrossRef]
[65] Inoue M, O'Gorman L, Davis J, Ji O. An international comparison of early childhood educators' understandings and practices in education for sustainability in Japan, Australia, and Korea. Int J Early Years Educ 2017;49:353–373. [CrossRef]
[66] Petkou D, Andrea V, Anthrakopoulou K. The impact of training environmental educators: environmental perceptions and attitudes of pre-primary and primary school teachers in Greece. Educ Sci 2021;11:274. [CrossRef]
[67] Winter V, Kranz J, Möller A. Climate change education challenges from two different perspectives of change agents: Perceptions of school students and pre-service teachers. Sustainability 2022;14:6081. [CrossRef]
[68] Baldwin C, Pickering G, Dale G. Knowledge and self-efficacy of youth to take action on climate change. Environ Educ Res 2023;29:1597–1616. [CrossRef]
[69] Hall C, Slembrouck S. Communication with parents in child welfare: skills, language and interaction. Child Fam Soc Work 2009;14:461–470. [CrossRef]
[70] Howells R, Lopez T. Better communication with children and parents. Paediatr Child Health 2008;18:381–385. [CrossRef]
[71] King G, Desmarais C, Lindsay S, Piérart G, Tétreault S. The roles of effective communication and client engagement in delivering culturally sensitive care to immigrant parents of children with disabilities. Disabil Rehabil 2015;37:1372–1381. [CrossRef]
[72] O'Brien K, Selboe E, Hayward BM. Exploring youth activism on climate change: dutiful, disruptive, and dangerous dissent. Ecol Soc 2018;23:42. [CrossRef]
[73] Stratford E. 'Dear Prime Minister ...' Mapping Island Children's Political Views on Climate Change. In: Children, Young People and Critical Geopolitics. London: Routledge; 2016. [CrossRef]
[74] Walker C. Embodying 'the Next Generation': children's everyday environmental activism in India and England. Contemp Soc Sci 2017;12:13–26. [CrossRef]
[75] Cloughton I. Global Youth Activism on Climate Change. Social Work & Policy Studies: Social Justice, Practice and Theory 2021;41:1–12.
[76] Goymer P. A trillion trees. Nat Ecol Evol 2018;2:208–209. [CrossRef]
[77] Sporre K. Young people - citizens in times of climate change? A childist approach to human responsibility. HTS Teol Stud 2021;77:1–8. [CrossRef]

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 965–972, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2023.00146

Research Article

Investigation of the cytotoxic effects of juglone on C6 glioma cell line

Ekin BEKTAŞ[1,*], Mine KUÇAK[1], Firuze ÜNLÜ BEKTAŞ[1], Tolga ÖNTÜRK[1], Muhammed Hamza MÜSLÜMANOĞLU[1]

[1]Department of Molecular Biology and Genetics, Yıldız Technical University, İstanbul, 34220, Türkiye

ARTICLE INFO

Article history
Received: 19 October 2022
Revised: 17 December 2022
Accepted: 12 March 2023

Keywords:
Cancer; Cytotoxicity; Glioma; Juglone; RT-qPCR

ABSTRACT

Juglone is an organic compound found naturally in the leaves, roots, bark, and fruits of plants belonging to the Juglandaceae family. It has toxic and growth inhibiting features in many plant species. Glioma is a type of tumorigenesis that occurs when glial cells begin to proliferate uncontrollably because of various reasons. C6 glioma cell line is a tumour line derived from brain glial cells of *Rattus norvegicus*. In this study, after juglone application to C6 glioma cell line, MTT and WST-1 tests were performed and IC_{50} values were found 6.666 µM and 5.646 µM, respectively. The data obtained showed that the results of two tests were closely related and the cytotoxic effects of juglone were present in this cell line. Furthermore, colony formation assay showed a decrease in colony formation abilities of the cells after juglone treatment. These observations were reinforced by wound healing assay and similarly, the migration ability of the cells decreased with juglone application. Morphological examinations of the cell groups were performed under fluorescent microscopy by DAPI staining, which showed a reduction in cell number. As a final step, the expression levels on the cancer related genes were determined by RT-qPCR. While a significant increase was observed in the expression levels of pro-apoptotic genes compared to the control group, a significant decrease was observed in the expression of the anti-apoptotic gene and the gene involved in proliferation. In line with the data presented, the cytotoxic effects of juglone were determined in the C6 glioma cell line.

Cite this article as: Bektaş E, Kuçak M, Ünlü Bektaş F, Öntürk T, Müslümanoğlu MH. Investigation of the cytotoxic effects of juglone on C6 glioma cell line. Sigma J Eng Nat Sci 2024;42(4):965–972.

INTRODUCTION

Juglone, also called 5-hydroxy-1,4-naphthaledione, is an organic compound with the molecular formula $C_{10}H_6O_3$. It is found naturally in the leaves, roots, bark, fruit, and bark of plants in the Juglandaceae family, particularly the black walnut, and is toxic or growth retarding for many plant species [1]. Juglone is not a substance found or newly discovered in the 21st century, however; research on the effects of juglone on cancer has recently gained momentum. Juglone has cytotoxic effects in a variety of human cancer lines such as prostate cancer, breast cancer, skin, and lung cancer [2,3]. In addition, there are effects of juglone not only on human cancer lines but also on rat cancer cell

*Corresponding author.
*E-mail address: ekinbektas@hotmail.com

This paper was recommended for publication in revised form by Editor in-Chief Ahmet Selim Dalkilic

lines. The action mechanism of juglone and how it works are tried to be clarified [4]. Juglone has cytotoxic and genotoxic effects on rat melanoma cells [5].

Gliomas are tumours originating from glial cells and are common tumours of the central nervous system [6]. Many model cells are used to understand glioma cell biology. One of the most well-known and widely used among these is C6 glioblastoma cells. C6 glioblastoma cells were first obtained from Wistar furth rat brains by Benda et al. (1968) by forming glioblastoma with N-nitroso methyl urea [7]. C6 glioblastoma cells exhibit fibroblast-like morphology, are in polygonal form and multiply well on the surface they attach to [8]. C6 glioma cells are frequently used for glioblastoma research since they have high mitotic activity as well as various malignant glioblastoma features such as intratumor haemorrhage, nuclear polymorphism, and tumour necrosis foci [9]. C6 cell line was determined to be the cell line most similar to the mechanism of human brain tumours [9].

In this study, it was aimed to demonstrate and observe the cytotoxic effects of juglone, a secondary metabolite, in C6 glioma cancer cell line. Juglone is advantageous agent because it is a substance that can be obtained naturally and has been used in various industries for a long time [10,11]. This study was carried out to observe its effect on brain cancer, which is one of the biggest problems of today. Therefore, it was aimed to use juglone in further studies and to show whether it is a potential therapeutic agent.

MATERIALS AND METHODS

Cell Culture

The C6 glioma cells were obtained from Yildiz Technical University (Istanbul, Turkey). The C6 glioma cells were cultured in Dulbecco's Modified Eagle Medium-High glucose (DMEM) (Sigma) supplemented with 10% fetal bovine serum (FBS) (Thermo Fisher) and 1% penicillin/streptomycin (Capricorn). The cells were grown in a 5% CO_2 incubator at 37°C in a humidified atmosphere. All cell culture was limited to twenty-five passages (P25).

Cell Viability Assays

The viability of C6 glioma cells was first determined using the MTT test. For 24 hours, the cells (10^4) were plated in 96-well culture plates. For another 24 hours, the experimental cells were exposed to juglone (Across) doses of 5, 10, 15, 20, 40, and 80 µM. After 24 hours of treatment, 10 µL MTT (Goldbio) was applied to the cell and incubated for 2 hours at 37°C. Then, 100 µL dimethyl sulfoxide was added to dissolve the formazan crystals and reveal the cells. The absorbance at 570 nm was measured, and the IC_{50} of juglone was computed using GraphPad Prism 9.3.0.

The viability of the cells was also tested with the WST-1 method using given procedure. The cells (10^4) were seeded in 96-well culture plates overnight. The experimental cells were treated with 5, 10, 20, and 40 µM concentrations of juglone for 24 hours. These doses were determined according IC_{50} value of juglone in MTT test. After treatment, 10 µL WST-1 solution (Cayman) was added into cells and incubated at 37°C for 2 hours. Optical density was measured at 450 nm absorbance and the IC_{50} of juglone was calculated via GraphPad Prism 9.3.0.

Considering the IC_{50} value of both cell viability assays, two promising doses were chosen for the next tests. Doses one above and one below of the IC_{50} value were selected to observe whether there was a steady dose-dependent decrease when juglone was applied to the cells.

Colony Formation Assay

In this method, (2×10^5) cells were seeded in each well of a 6-well plate and incubated for 24 hours. The first well was determined as the control group and no treatment was applied. 5 µM and 10 µM juglone doses were added to the other wells, respectively, and incubated for 24 hours. After 24 hours of incubation, untreated and juglone treated cells were replated in a 6-well plate at a density of 250 cells per well. They were incubated for 10 days at 37°C and 5% CO_2. After 10 days, the images of the colonies formed after the cells were stained with crystal violet were taken. The colony numbers of control and experimental groups determined with image processing program ImageJ.

DAPI Staining

The cells were plated in T25 flasks and cultured to 70-80% confluency. They were subsequently cultured for 24 hours in the presence of 5 µM, 10 µM juglone and in the absence of juglone. The cells were rinsed with 1X PBS (Wisent) and fixed for 10 minutes in ice cold methanol (Merck). Following fixation, the nuclei were stained with 850 nM DAPI staining solution (Cayman) and kept in dark. Finally, the DAPI solution was withdrawn from the cells, and they were rinsed with 1X PBS. A fluorescent microscope was used to image the cells (Zeiss). Image processing program ImageJ was used to count the cells.

Wound-Healing Assay

The cells (5×10^4) were seeded onto 24-well culture plates and cultured to 80% confluency. Using sterile 100 µL pipette tips, a wound was scraped. The cells were subsequently treated with 5 µM and 10 µM juglone. Images of the cells were obtained under a light microscope at 0 and 24 hours (Zeiss). Microscopy image analysis program Wimasis was used to compute the scratch areas.

Reverse Transcription-Quantitative PCR Assay

Total RNA of cells was isolated with RiboEx™ solution (Geneall) using the manufacturer's instructions and the concentration of total RNA was determined via 260/280 nm absorbance via a NanoDrop spectrophotometer (Thermo Scientific). Reverse transcription reactions were performed with a smART First Strand cDNA Synthesis kit (EUR_X) by

Table 1. Primer sequences

Genes	Forward sequence (5'3')	Reverse sequence (5'3')
β-actin	CTCTGTGTGGATTGGTGGCT	GCAGCTCAGTAACAGTCCGC
MMP8	TGGAGTGTGCCATCAACCCTGAC	CACCATGGTCTCTTGAGACGAAAGC
BAX	GAGGACTCCAGCCACAAAGA	CGAGCTGATCAGAACCATCA
BCL-2	TATATGGCCCCAGCATGCGA	GGGCAGGTTTGTCGACCTCA
Caspase3	GGAGCTTGGAACGCGAAGAA	ACACAAGCCCATTTCAGGGT
PARP1	TCTACTTTGCTGATATGGTGTCC	TGGGTAACTTGCTGATGTGAG

Table 2. qPCR protocol

Step	Cycles	Temperature	Time
Polymerase activation	1	95 °C	2 min
Denaturation	40	95 °C	5 s
Annealing		60 °C	10 s
Extension		72 °C	30 s

given protocol. The quantitative PCR reaction performed on Applied Biosystem 7500 with the total volume of 20 µL. The sequences of specific PCR primers are shown in Table 1 and PCR reaction are set as in Table 2. β-actin primer was used as a housekeeping gene. The relative quantification of the mRNA was calculated using $2^{-\Delta\Delta Ct}$ method.

Statistical Analyses

The tests were repeated three times. Comparisons among groups were performed using one-way ANOVA by GraphPad Prism 9.3.0. The significant level was set at $p < 0.05$.

RESULTS AND DISCUSSION

Juglone Inhibits Cell Viability of C6 Glioma Cells

By MTT assay, the effects of various juglone doses on cell viability were measured for 24 hours. After treatment, juglone inhibited proliferation of C6 glioma cells in a dose-dependent manner (Figure 1A). The IC_{50} value was calculated as 6.666 µM (Figure 1B). Additionally, R^2 value was found as 0.9953 (Figure 1A). In addition, IC_{50} value was measured with the WST-1 test. This value was calculated as 5.646 µM and the R^2 value of the result was determined as 0.9504 (Figure 1C&D). The calculated IC_{50} values were found to be statistically significant. The results of two cytotoxicity tests show that, juglone remarkably reduces the survival rate of C6 glioma cells. According to the IC_{50} value determined by MTT Assay and WST-1 test, the doses to be used in the continuation of the study were determined as 5 µM and 10 µM. Choosing two doses, above and below the average of the IC_{50} values, will be beneficial for observing the dose-dependent effect of juglone in the later steps.

Juglone Decreases Colony Formation Capability of C6 Glioma Cells

Colony formation experiment was performed to validate the cytotoxic effects of juglone on C6 glioma cells as well as to observe the long-term effects of juglone. The test was done using 5 µM and 10 µM doses to compare with the control group so as to determine the effects of juglone on cells gradually. In treated C6 glioma cells, the number of colonies decreased compared to the control group. (Figure 2A). The colony formation test shows that juglone significantly reduces the colony forming ability of the cells, depending on the dose.

Juglone Causes Apoptotic Bodies in C6 Glioma Cells

The number of DAPI-stained nuclei was counted, and the number of nuclei was decreased as shown in Figure 2B&2C for juglone-treated cells. In addition to a dose-dependent decrease in the number, it was observed that apoptotic structures were formed, especially when 10 µM dose was applied because of the cytotoxic effects of juglone. Furthermore, when compared with the control group, it was observed that nuclei of the cells treated with juglone were smaller and their structure was deteriorated.

Juglone Inhibits Migration of C6 Glioma Cells

Cell migration is an important part of metastasis. Thus, the ability of C6 glioma cells to migrate after being treated with juglone was assessed using a wound-healing test. In the absence or presence of juglone, 80% confluent cells were scratched using a sterile pipette tip. Scratch areas were larger in juglone-treated cells than in the untreated group after 24 hours of treatment (Figure 3A), showing that juglone caused a decrease in cell-cell communication. Juglone application for 24 hours inhibited the ability of C6 glioma cells to migrate, according to the findings. The efficiency of migration was reduced by 31,94% and 75,694%, respectively (Figure 3B).

Juglone Alters the Expression Levels of MMP8, BAX, BCL2, CASPASE3 and PARP1

β-actin has been used as a housekeeping gene. A significant increase was observed in the expression level of BAX gene, a gene with a pro-apoptotic role in the apoptosis mechanism [12], at a dose of 5 µM compared to the control group. A significant decrease was observed in the expression

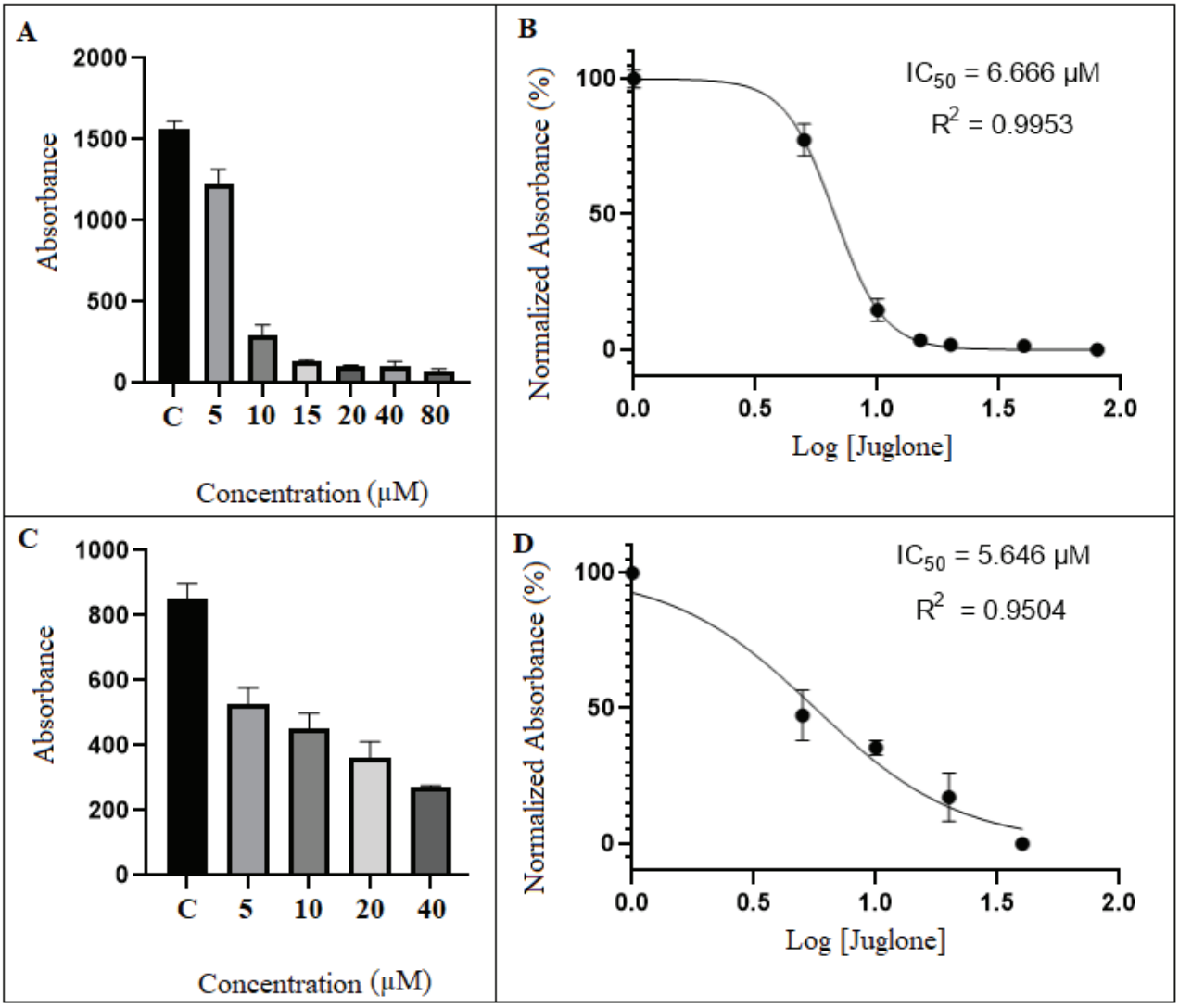


Figure 1. Juglone inhibits cell viability of C6 glioma cells. **(A)** MTT assay, C6 glioma cells were treated with juglone for 24 hours. The cell viability reduced in a dose-dependent manner ($P < 0.05$). **(B)** IC_{50} value was found with MTT assay as 6.666 µM via GraphPad Prism 9.3.0. **(C)** WST-1 assay, C6 glioma cells were treated with juglone for 24 hours. The cell viability reduced in a dose-dependent manner ($P < 0.05$). **(D)** IC_{50} value was found with WST-1 assay as 5.646 µM via GraphPad Prism 9.3.0.

level of the BCL2 gene, a gene with an anti-apoptotic role in the apoptosis mechanism [13], at a dose of 5 µM compared to the control group. A significant decrease was observed in the expression level of the MMP8 gene, a gene that has a role in the proliferation mechanism [14], at a dose of 5 µM compared to the control group. A significant increase was observed in the expression level of Caspase3, a gene with a proapoptotic role in the mechanism of apoptosis [15], at a dose of 5 µM compared to the control group. An increase in the expression level of the PARP1 gene, a gene that has a role in the DNA repair mechanism [16], was observed at a dose of 5 µM compared to the control group, but it was not found significant (Figure 3C).

There are findings showing that juglone has an effect on prostate, colon, lung, breast and skin cancer cell lines [3]. Juglone has been shown to have an inhibitory activity on tumour growth as the dose increases in the mouse model [17]. Based on these findings, this study was carried out to show that juglone also affects glioma cells and finally to suggest the use of juglone as an effective agent. Based on Zhang et al.'s[3] and our findings, the IC_{50} value differs depending on the cell line and the cytotoxicity method used. Similarly, Karki et al. [18] showed that cytotoxic effect of juglone changes according to the type of cancer cell line used and they also found that the IC_{50} value of juglone in MIA Paca-2 cells was 5.05 µM after

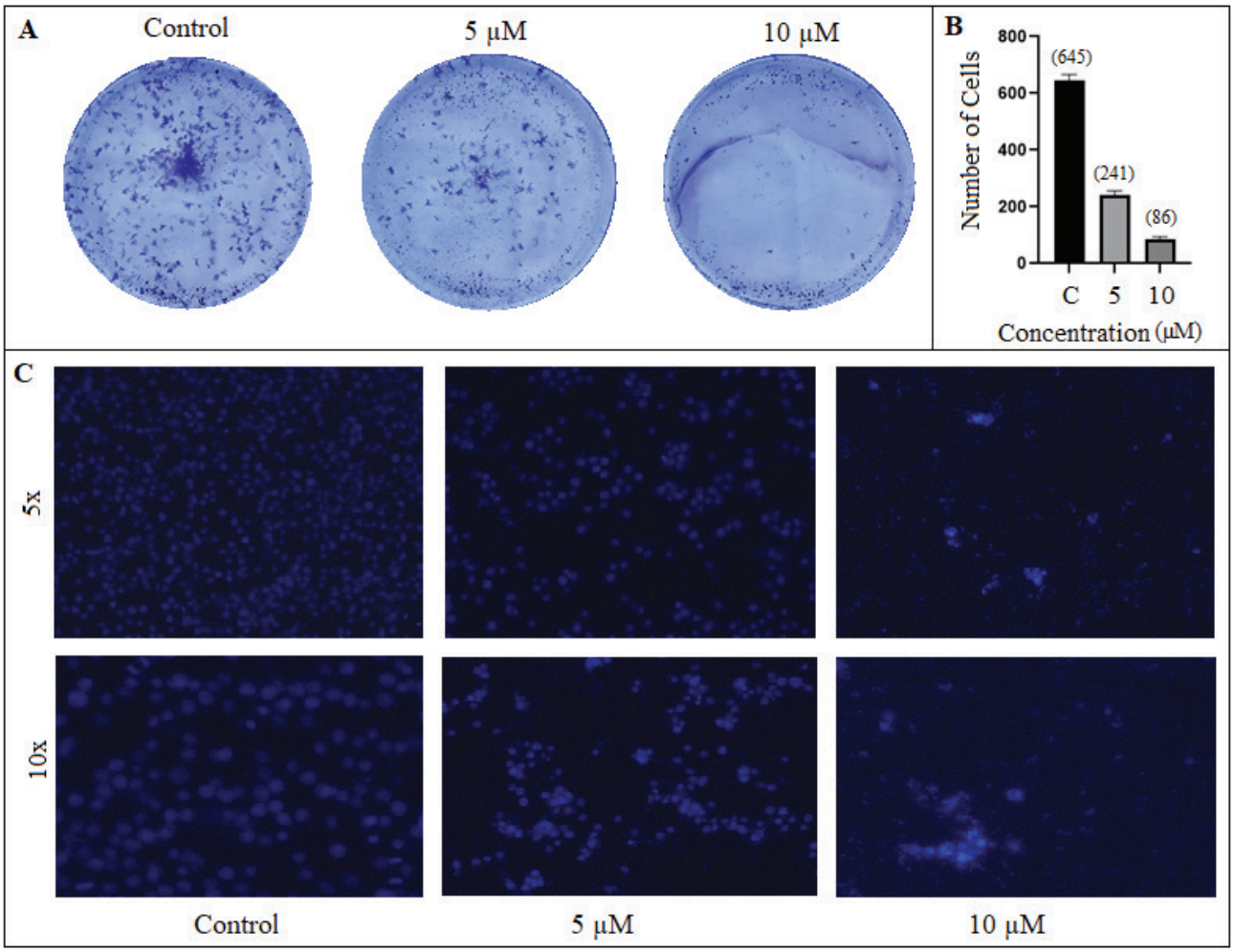


Figure 2. Juglone reduces colony formation capability of C6 glioma cells and the number of cells **(A)** Juglone significantly reduced the number of colonies in both treated groups, 5 μM and 10 μM ($P < 0.05$). **(B)** The number of nuclei reduced in 5 μM and 10 μM juglone treated cells ($P < 0.05$). **(C)** The image of DAPI stained nucleus at 5x and 10x.

24 hours application. Furthermore, another study was demonstrated that IC_{50} value for juglone was 9 μM on the HaCat cell line [1]. Mahdavi et al. [19] demonstrated that juglone has cytotoxic and apoptotic effects on PC3 and DU145 cell lines at a dose of 10 μg/mL. In addition, they showed that the cell death rate increased as the exposure time increased, although there was no increase in the dose amount depending on the dosing time [19]. Accordingly, the IC_{50} value of juglone decreases as the dosing time increases. Based on our findings in this study, it is possible to suggest that there is a decrease in cell number due to the increase in the amount of juglone.

As Aithal et al. [20] demonstrated in B16F1 cell line, juglone treatment is the reason of decrease in number of colonies, which is in line with our findings. Therefore, juglone shows anti-proliferative effects in cell lines [20]. Inbaraj et al. [1] demonstrated that on the HaCat cell line, juglone and another quinone, plumbagin, had cytotoxic effects and therefore they detected a significant decrease in the number of colonies. Karki et al. [18] used the same amount of juglone, which are 5 and 10 μM, on MIA Paca-2 cells and showed that juglone treatment causes to loss of colony forming ability after 6 hours. The similar effects of juglone in different cell lines make it possible to suggest that juglone is an effective cytotoxic agent. The decrease in colony forming ability is due to the decreased ability of cells to divide and migrate after administration of juglone.

It has been reported by Wang et al. [21] that on U-251 cells, apoptotic effects of juglone lead to a decrease in cell migration and a decrease in wound healing [21]. Wound healing assay's and MTT assay's results and changes in Caspase3 expression, which are the experimental methods used in this study, are found to be similar. Obtaining similar results in our study supports that juglone causes decrease on wound healing, reduces cell migration and creates changes in gene expression levels. Shi et al. [22] showed that juglone has apoptotic effects on OVCAR-3 cell line via DAPI

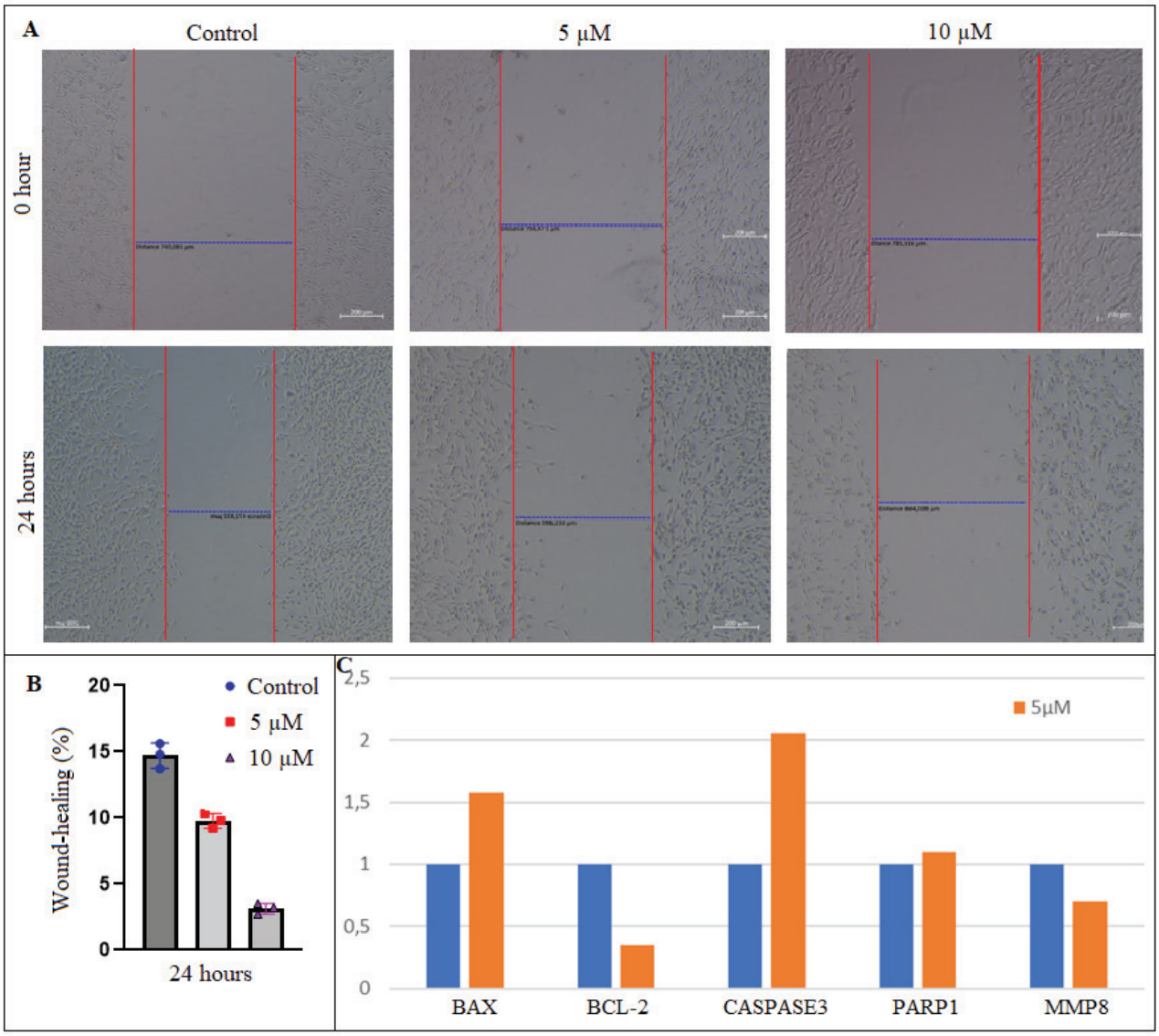


Figure 3. Juglone inhibits migration of rat glioma cells. Juglone alters the expression levels of MMP8, BAX, BCL2, Caspase3 and PARP1. **(A)** Wound areas of 5 µM and 10 µM juglone treated cells captured after 24 hours. **(B)** After 24-hour treatment with juglone migration efficiencies were reduced by 31,94% and 75,694%. **(C)** Expression differences of MMP8, BAX, BCL-2, Caspase-3 and PARP1 between untreated and 5 µM treated C6 glioma cells.

staining method [22]. Staining the cells with this method allowed to observe the change in the treated groups to be clearly understood. Another study showed that juglone application for 24 hours causes small, condensed and bright nuclei and apoptotic bodies in MIA Paca-2 cells, which was observed by Hoechst staining method [18]. Similarly, in the current study, juglone causes deformation in cell morphology and formation of apoptotic bodies.

It was stated in the study of Lu et al. [23] that the apoptotic and anti-proliferative effects created by juglone were determined by using the c-Jun N-terminal kinase/c-Jun pathway. Furthermore, there was an increase in BAX and Caspase3 gene expressions after juglone treatment [23]. These results show that the expression of BAX and Caspase3, which are pro-apoptotic genes, underwent a change compared to the control group after juglone administration, and the decrease in the cell number may have been caused by apoptosis due to the change in these gene expression levels. In parallel with our study, Xu et al. showed that juglone application increased Caspase3 gene expression, increased BAX gene expression and decreased BCL-2 gene expression of HL-60 cells compared to the control group [24]. The observed changes in gene expression levels provide information on how juglone alters the expression of certain genes in C6 glioma cells.

CONCLUSION

The experiments of juglone performed on many different cell lines in many studies show slight differences depending on the experimental and the methods used. However, the common point observed in all studies is that juglone has cytotoxic effects on many cell lines, as we observed in our study. We applied cytotoxicity tests to determine the cytotoxicity of juglone on C6 glioma cells and supported these results with the colony formation assay and the wound-healing assay. Similarly, we examined cells under fluorescence microscope to visualize these effects. Finally, we observed how juglone affects expression levels in certain genes. Considering the cytotoxic effects of juglone on the C6 glioma cell line, it is possible to suggest juglone as an effective agent. It can be suggested that juglone affects C6 glioma cells, especially based on our results of the RT-qPCR experiment performed with 5μM juglone application.

Further studies can be carried out with Western Blot to show how the effects of juglone are related to protein modifications and post-translational modifications. RT-qPCR can be done using more cancer-related genes to observe overall effect of juglone on C6 glioma cell and further, testing different cell lines would help to understand effects of juglone on cancer cells. In order to confirm the determination of apoptosis; TUNEL test, which is used to detect DNA fragments formed during apoptosis, can be applied. After enough *in vitro* studies, this hypothesis can be supported by *in vivo* studies. It is possible to determine whether juglone could potentially be used as a therapeutic agent in the treatment of glial cancer, but further research is required to achieve this.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Inbaraj JJ, Chignell CF. Cytotoxic action of juglone and plumbagin: A mechanistic study using HaCaT keratinocytes. Chem Res Toxicol 2004;17:55–62. [CrossRef]

[2] Bayram D, Özgöçmen M, Armagan I, Sevimli M, Türel GY, Şenol N. Investigation of apoptotic effect of juglone on CCL-228-SW 480 colon cancer cell line. J Cancer Res Ther 2019;15:68–74. [CrossRef]

[3] Zhang XB, Zou CL, Duan YX, Wu F, Li G. Activity guided isolation and modification of juglone from Juglans regia as potent cytotoxic agent against lung cancer cell lines. BMC Complement Altern Med 2015;15:396. [CrossRef]

[4] Sugie S, Okamoto K, Rahman KM, Tanaka T, Kawai K, Yamahara J, et al. Inhibitory effects of plumbagin and juglone on azoxymethane-induced intestinal carcinogenesis in rats. Cancer Lett 1998;127:177–183. [CrossRef]

[5] Aithal BK, Kumar MR, Rao BN, Udupa N, Rao BS. Juglone, a naphthoquinone from walnut, exerts cytotoxic and genotoxic effects against cultured melanoma tumor cells. Cell Biol Int. 2009 Oct;33(10):1039–1049. [CrossRef]

[6] Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization classification of tumors of the central nervous system: A summary. Acta Neuropathol 2016;131:803–820. [CrossRef]

[7] Benda P, Lightbody J, Sato G, Levine L, Sweet W. Differentiated rat glial cell strain in tissue culture. Science 1968;161:370–371. [CrossRef]

[8] Kłopocka W, Korczyński J, Pomorski P. Cytoskeleton and Nucleotide Signaling in Glioma C6 Cells. In: Barańska J, editor. Glioma signaling. Cham, Switzerland: Springer; 2020. p. 109–128. [CrossRef]

[9] Grobben B, De Deyn PP, Slegers H. Rat C6 glioma as experimental model system for the study of glioblastoma growth and invasion. Cell Tissue Res 2002;310:257–270. [CrossRef]

[10] Khalafy J, Bruce JM. Oxidative dehydrogenation of 1-tetralones: Synthesis of juglone, naphthazarin, and α-hydroxyanthraquinones. J Sci IRI 2002;13:131–139.

[11] Cosmulescu S, Gheorghe A, Botu M, Baciu A, Gruia M. Phenolics of green husk in mature walnut fruits. Not Bot Horti Agrobo 2010;38:53–56.

[12] Gross A, Jockel J, Wei MC, Korsmeyer SJ. Enforced dimerization of BAX results in its translocation, mitochondrial dysfunction and apoptosis. EMBO J 1998;17:3878–3885. [CrossRef]

[13] Vaux DL, Cory S, Adams JM. Bcl-2 gene promotes haemopoietic cell survival and cooperates with c-myc to immortalize pre-B cells. Nature 1988;335:440–442. [CrossRef]

[14] Kessenbrock K, Plaks V, Werb Z. Matrix metalloproteinases: Regulators of the tumor microenvironment. Cell 2010;141:52–67. [CrossRef]

[15] Alnemri ES, Livingston DJ, Nicholson DW, Salvesen G, Thornberry NA, Wong WW, et al. Human ICE/CED-3 protease nomenclature. Cell 1996;87:171. [CrossRef]

[16] Pascal JM. The comings and goings of PARP-1 in response to DNA damage. DNA Repair (Amst) 2018;71:177–182. [CrossRef]

[17] Wang H, Zou C, Zhao W, Yu Y, Cui Y, Zhang H, et al. Juglone eliminates MDSCs accumulation and enhances antitumor immunity. Int Immunopharmacol 2019;73:118–127. [CrossRef]

[18] Karki N, Aggarwal S, Laine RA, Greenway F, Losso JN. Cytotoxicity of juglone and thymoquinone against pancreatic cancer cells. Chem Biol Interact 2020;327:109142. [CrossRef]

[19] Mahdavi M, Azadbakht M, Vahdati A, Shokrzadeh M, Farhadi A. Cytotoxic effects of juglone and *pterocarya fraxinifolia* on prostate cancer cells. J Pharm Bioallied Sci 2019;11:195–204. [CrossRef]

[20] Aithal KB, Kumar S, Rao BN, Udupa N, Rao SB. Tumor growth inhibitory effect of juglone and its radiation sensitizing potential: In vivo and in vitro studies. Integr Cancer Ther 2012;11:68–80. [CrossRef]

[21] Wang J, Liu K, Wang XF, Sun DJ. Juglone reduces growth and migration of U251 glioblastoma cells and disrupts angiogenesis. Oncol Rep 2017;38:1959–1966. [CrossRef]

[22] Shi JY, Huang ZR, Gao HY, Xu XL. Anticancer effects of juglone in OVCAR-3 human ovarian carcinoma are facilitated through programmed cell death, endogenous ROS production, inhibition of cell migration and invasion and cell cycle arrest. J BUON 2021;26:1188.

[23] Lu Z, Chen H, Zheng XM, Chen ML. Experimental study on the apoptosis of cervical cancer Hela cells induced by juglone through c-Jun N-terminal kinase/c-Jun pathway. Asian Pac J Trop Med 2017;10:572–575. [CrossRef]

[24] Xu HL, Yu XF, Qu SC, Zhang R, Qu XR, Chen YP, et al. Anti-proliferative effect of Juglone from Juglans mandshurica Maxim on human leukemia cell HL-60 by inducing apoptosis through the mitochondria-dependent pathway. Eur J Pharmacol 2010;645:14–22. [CrossRef]

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 973–987, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2023.00031

Research Article

Calculation of buckling loads of IPE-section bending members based on optimization of analytical formulations

Ahmet ÖZBAYRAK[1,*], Mohammed Kamal ALI[2], Hatice ÇITAKOĞLU[1]

[1]Department of Civil Engineering, Faculty of Engineering, Erciyes University, Kayseri, 38039, Türkiye
[2]Department of Civil Engineering,College of Engineering, University of Kirkuk, Kirkuk, 36001, Iraq

ARTICLE INFO

Article history
Received: 09 November 2022
Revised: 25 December 2022
Accepted: 14 February 2023

Keywords:
Analytical Calculation; Buckling Load; Calibration; Cantilever Beam; Finite Element Method

ABSTRACT

The critical lateral buckling load of cantilever beams with IPE cross-section was calculated using analytical closed-form equations and numerical finite element analyses within the scope of the research. The equations suggested in the specifications for simply supported beams were used to calculate the buckling load of cantilever beams. The rationality of the values calculated due to this is not fully known. In the research, a single loading was made to the shear center at the free end of the cantilever beam. Cantilever length and section height were kept variable. As a result, it has been determined that there are partial differences in the analysis result obtained from the elastic stability theory and finite element method. Accordingly, the results obtained from ANSYS and SAP2000 analyses confirm each other. On the other hand, the results obtained using the formulation of Timoshenko and Gere, the calculation results made according to the AISC and DCCPSS regulations, and the results obtained from the LTBeam program confirm each other. However, it differs from the FEA analysis due to the cantilever beam length's shortening and the section height increase. Thus, to obtain accurate and reliable results in the buckling load calculation of cantilever beams, the equations used in analytical calculations were optimized according to finite element analysis (FEA) results. As a result of the study conducted according to the error criteria, it was determined that the updated equation results gave similar results to the FEA results.

Cite this article as: Özbayrak A, Ali MK, Çıtakoğlu H. Calculation of buckling loads of IPE-section bending members based on optimization of analytical formulation. Sigma J Eng Nat Sci 2024;42(4):973–987.

INTRODUCTION

In the lateral buckling calculations of steel beams, the design methods of simply supported and cantilever beams given in the regulation are the same. However, due to the different end support conditions in cantilever beams, the maximum displacement and buckling angle occur at the free ends instead of the middle of the span. The buckling modes obtained as a result of this situation are different from each other. Therefore, the recommended methods for simple support beams are not suitable for cantilever beams

***Corresponding author.**
*E-mail address: ozbayrak@erciyes.edu.tr

This paper was recommended for publication in revised form by Editor in-Chief Ahmet Selim Dalkilic

[1]. AISC or DCCPSS regulations do not guide the lateral buckling of cantilever beams [2,3]. This research aims to provide rational information against lateral buckling in the design of steel cantilever beams. In general, the concepts of lateral buckling of beams and lateral-torsional buckling are explained in many books in the literature. Accordingly, the elastic lateral torsion buckling load under the bending effect of simple supported beams can be solved with the help of closed-form equations [4–7]. However, analytical solutions become very complex when beam end conditions differ from simple support. Therefore, numerical approximations such as the finite element method are needed to solve basic differential equilibrium equations [8–10]. The load-displacement relationships of I-section steel cantilevers were investigated by [11]. Under the effect of a single load at the free end, numerical and experimental results were compared. Accordingly, estimating the buckling load by numerical analysis with the ABAQUS program during the design phase was considered an acceptable method. Studies on unsymmetrical I-section cantilever beams were carried out by Samanta and Kumar [12]. In the studies, single load, distributed load, and moment were affected at the beam end. With the help of the ABAQUS program, the buckling load was investigated by giving lateral support to the top flange, bottom flange, and both. Accordingly, it has been found that if loading is made to the lower flange, the side support position does not significantly affect the cantilever beam buckling capacity. Özbaşaran et al. presented alternative design methods for calculating the buckling load and movement of I-section cantilever beams under the effect of lateral-torsional buckling [13]. In the critical elastic lateral-torsional buckling load calculation, the results of the closed-form equations, the analysis made by ABAQUS, and the experimental findings were found in accordance with the results of the proposed design method. Ma et al. conducted a study on elastic lateral buckling of unsymmetrical I-section cantilever beams [14]. According to the Rayleigh-Ritz method, while the profile flanges remain linear during buckling, it is assumed that the web part is susceptible to distortion. The accuracy of the proposed method has been verified with the help of the NASTRAN program, which calculates according to the finite element method. The elastic lateral torsional buckling behavior of tapered beams with different support conditions has been investigated by Andrade et al. [15]. A better understanding of the tapered beam behavior was provided by providing concrete explanations for some of the results regarded as illogical in the research. Zhang et al. conducted studies on the lateral-torsional buckling behavior of I-section cantilever beams with stiffening plates [16]. An analytical solution of the dimensionless buckling equation of these beams was obtained with the help of dimensionless parameters. The dimensionless critical moment formula developed with the help of mathematical optimization analysis software (1stOPT) has been verified with ADINA finite element software. A simple and useful calculation method for practical engineering calculations is presented in the research. The results of the finite element analysis of the elastic lateral-torsional buckling strength of light steel cantilever beams under the effect of transverse loading were shared by Kurniawan and Mahendran [17]. Accordingly, the applicability of modification factors in various steel design codes was reviewed, and the design approach in the AS4100 code was proposed for light steel cantilever beams subjected to transverse loading. The study carried out by Trahair stated that the lateral buckling formulations suggested in the design regulations for simply supported beams with uniformly distributed loads are not suitable for cantilever beams [7]. His study aimed to develop simple approximate methods in the design of cantilever beams against inelastic lateral buckling. Within the scope of the research conducted by Yılmaz and Kıraç [18], an equation that can be used to calculate the critical torsional buckling load of the IPE and IPN simple support beams in European norms was presented [18]. The slenderness of the profile section and the effect of loading positions were taken into account in their study. Consistent results were obtained among analytical, parametric, and numerical solutions. It has been found that the lateral torsional buckling load of European IPE and IPN beams can be determined by the presented equation and used safely in design procedures. I-section composite beams are discussed by Prombut and Anakpotchanakul [19]. It has been observed that the bending results obtained from the shear deformation theory and finite element analysis in beams under uniformly distributed load applied to the upper flange are compatible with each other. It is stated that thanks to the validated finite element procedure, realistic results can be obtained based on curvature, taper, and buckling along the length of an I-section. Özbaşaran and Yılmaz introduced shape optimization for symmetrical I-section beams with tapered flanges and/or web [20]. The optimization procedure was created using the Big Bang - Big Crunch algorithm and Deb's constraint handling method. The designs made were verified by finite element analysis. It has been shown that tapering in absolute conditions may not significantly affect the material economy. Trahair stated that the design methods given in regulations such as AS4100, BS595, Eurocode3, and AISC for lateral buckling of cantilever beams are modifications of the rules introduced according to simple support beams [1]. The accuracy of these modifications was found to be questionable, and it was emphasized that they could not fully guide the design. A different method has been developed, and the solution has been summarized with examples.

Minimizing the cost and weight of products has been an area of interest for many industries. It is among these sectors in reinforced concrete and steel structures. Complex situations arise in reinforced concrete and steel structures design due to the nonlinear structure behavior and related design equations. In addition, the behavior of the designed sections under the effect of dynamic loads also creates complex situations. These problems are sizing optimization

problems [21]. Previous optimization studies on reinforced concrete and steel structures are based on weight and cost [22–30]. Shaqfa and Orbán improved the position of the upper and lower flexural member, simultaneously minimizing cost, weight, and cost-weight [31]. Hayalioğlu and Değertekin presented a genetic algorithm for designing the optimum cost of nonlinear steel frames with semi-rigid connections subject to the displacement and stress restrictions of the American Institute of Steel Construction-Allowable Stress Design (AISC-ASD) regulation [23]. As a result of their studies, they stated that more economical optimum frames could be obtained by adjusting the stiffness of the connections in frame systems. Omkar et al. used Particle Swarm Optimization (PSO) to minimize the weight and total cost of the composite component to achieve a certain strength of composite components [32]. Barraza et al. used Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) to minimize the structural weight of steel structures exposed to earthquake loads and to improve the structural performance of buildings [33]. As a result of their studies, they emphasized that they generally obtained better solutions with PSO in structural buildings compared to the GA approach. Another issue of sizing optimization is to maximize the cross-section against torsion and fracture [34–37]. Cho optimized the design of a composite cylindrical shell against buckling and fracture and stated that the optimized composite cylindrical shell exhibits significantly improved mechanical properties compared to the traditional design as result of the study [36]. Many optimization studies are also done in Excel–Solver [37–39]. Taki optimized the dimensions of the Z-hardened panel under compression load with Excel-solver to update Farrar's work. As a result of the work, he developed design charts for Z-hardened panels and produced a design guide [40]. Msabawy and Mohammad used the Generalized Reduced Gradient (GRG) algorithm in the Solver Add-on tool in Microsoft Excel to perform first-order elastic structural analysis of semi-rigid steel portal frames [37]. Msabawy and Mohammad used the GRG algorithm to optimize cross-sectional areas in cold-formed steel frames [39]. As a result of their studies, they stated that it proved the reliability and validity of the GRG algorithm in terms of the ability to obtain optimum configurations of optimized sections. In addition to the sizing optimization problem in reinforced concrete and steel structures, there are modification studies of theoretical equations. Perelmuter and Yurchenko determined the optimum height and weight of the tower by changing various equations depending on the capacity of the wind-powered generator's generated energy [30]. Based on the concepts of the Euler-Bernouli beam theory and fracture mechanics, Vosoughi reformulated the management equation using genetic algorithms (GA) and particle swarm optimization (PSO) techniques [41]. They showed the convergence, efficiency, and accuracy of the optimization method with the finite element method by solving different examples. Le et al. took into account the Adaptive Neuro-Fuzzy Inference System (ANFIS) using the GA and PSO to assess the buckling damage of steel columns subjected to axially compressive load [42]. They concluded that the ANFIS-PSO method significantly outperformed the ANFIS-GA method with a correlation factor of 0.929. Jung et al. working to assess the tensile characteristics of high strength steel, used Artificial Neural Networks (ANN) and back-propagated linear regression [43]. They asserted that using a deep learning system produced predictions of yield strength, yield ratio, and tensile strength with high accuracy. Cuong-Le et al. introduced a PSO-optimized Support Vector Machine (SVM) to identify deterioration in truss and frame constructions [44]. Additionally, they contrasted the suggested approach with ANN, Deep Neural Networks (DNN), and Adaptive Neuro-Fuzzy Inference System (ANFIS). They concluded that the damage and the degree of damage for truss and frame structures were successfully identified using the proposed strategy, outperforming the other techniques. Das and Das have used Random Forest Regressor (RFR) to evaluate the fundamental natural frequencies of isotropic plate structures [45]. They have been considered as square, rectangular, thin, and thick plates whose materials have been selected as Structural Steel, Aernet 100, Al 7108, and Al 2024 for the isotropic plates. They claimed that the suggested strategy accurately predicts the fundamental natural frequency and is an adequate model for such a scenario. Özbayrak et al. conducted buckling load calculations using ANSYS on European I-section cantilever beams reinforced with transverse stiffener plates at various intervals [46]. They have created formulations employing multiple linear regression analysis and multigene genetic programming techniques to estimate the found load values more effectively. According to their statement, the lateral buckling stress according to the transverse stiffener plate spacing for European I-section cantilever steel beams can be calculated with formulations created using computer technology.

In the construction literature, more studies use machine-learning models of steel I-beams and cantilever beams. Artificial intelligence has enabled the suggested formula to successfully forecast the residual lateral buckling capacity of steel I-beams, according to research on artificial neural networks [47]. In a different study, a deep learning classifier was used to determine the damage status of cantilever beams in an invasive-free manner with the maximum level of accuracy [48]. Artificial neural networks were used to assess the twisting performance of a steel I beam that was externally attached to sheets with polymer matrix reinforcement enhanced with various fibers to reduce the experimental work [49]. Additionally, the web-post buckling shear strength of cellular beams and the load-bearing capability of castellated steel beams were predicted using artificial neural network models [50,51]. Thanks to a database provided by a study that included 475 finite element models, the lateral torsional buckling strength was calculated using an artificial neural network and the multiple regression approach [52]. An adaptive neuro-fuzzy

inference system was used to develop empirical equations for estimating natural frequencies from a finite element dataset [53]. Additionally, form optimization makes use of artificial intelligence. Using a genetic algorithm, the stiffness of cold-formed steel sections was increased [54].

Within the scope of the research, analytical calculations were compared with the results of numerical analysis. It has been observed that the calculations made with the LTBeam program are compatible with the analytical calculation results. Critical lateral buckling loads found from analytical equations and LTBeam program results are consistent in this regard. However, it has been determined that there are some differences between these and the FEA results depending on the profile cross-section and length. In the studies in the literature, it is stated that the analysis made according to the finite element method with the help of developing computer technology is more accurate than the calculations made with closed-form equations. First, using two different FEA programs, lateral buckling loads calculated in the ANSYS program were verified with the help of the SAP2000 program. Later, studies were carried out to harmonize the results from the equations given in Timoshenko, Gere, and other Regulations with FEA results (ANSYS). Using optimization techniques, the equation given by Timoshenko and Gere and formulations given in AISC and DCCPSS regulations were successfully updated.

MATERIALS AND METHODS

In the case of a single load acting on the shear centre at the free end of the IPE section cantilever beam, the lateral buckling load was calculated and compared with five different methods. These are, respectively, elastic stability theory, regulation on design, calculation, and construction principles of steel structures (DCCPSS), LTBeam program, SAP2000, and ANSYS software (Figure 1). In the calculations, the material elasticity modulus was 210000 MPa, the shear modulus was 80769 MPa, and the Poisson ratio was 0.3. The section heights of the cantilever beams used in the study include all IPE profiles in the range of 100-600 mm. Cantilever beam lengths were evaluated in five different sizes: 1000 mm, 1500 mm, 2000 mm, 2500 mm, and 3000 mm.

Calculation According to Elastic Stability Theory

The critical value of the lateral buckling load for cantilever beams is calculated as given in Equation 1 by Timoshenko and Gere [6], depending on the boundary conditions of the beam endpoints.

$$P_{cr} = \gamma_2 \frac{\sqrt{E \cdot I_y \cdot C}}{L^2} \tag{1}$$

The factor γ_2 in this expression is a dimensionless coefficient obtained according to the ratio L^2C/C_1. The values of this coefficient are as given in Table 1.

As the L^2C/C_1 ratio increases, the γ_2 factor approaches the 4.013 limit value. This value corresponds to the critical load of thin rectangular beams. If ratio L^2C/C_1 takes values greater than 40, the approximate factor γ_2 is calculated as given in Equation 2.

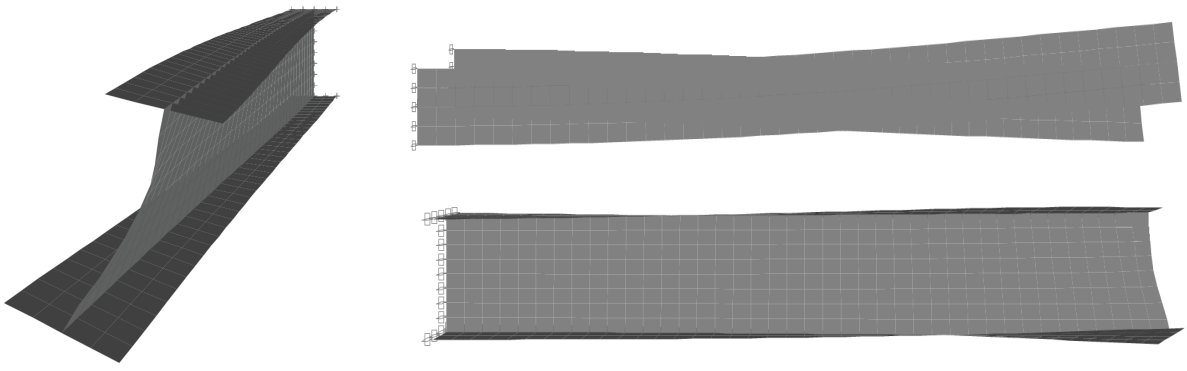


Figure 1. Lateral-torsional buckling condition.

Table 1. Factor γ_2 for I-section cantilever beams

L^2C/C_1	0.1	1	2	3	4	6	8
γ_2	44.3	15.7	12.2	10.7	9.76	8.69	8.03
L^2C/C_1	10	12	14	16	24	32	40
γ_2	7.58	7.20	6.96	6.73	6.19	5.87	5.64

$$\gamma_2 = \frac{4.013}{(1-\sqrt{C_1/L^2C})^2} \quad (2)$$

Calculation According to AISC and DCCPSS Regulations

In the case of lateral torsion buckling boundaries, the positive contribution of the bending moment distribution along the length between the points supported by the lateral stability connection is taken into account by the coefficient given in Equation 3. The regulations stipulate that this coefficient in cantilever beams is taken as C_b = 1 with an approach on the safe side. However, since it was determined that this approach has a limited contribution, the expression given in Equation 3 was used within the scope of the research.

$$C_b = \frac{12.5M_{maks}}{2.5M_{maks} + 3M_A + 4M_B + 3M_C} \quad (3)$$

The critical stress value of I-cross section elements with double symmetry axes, whose web and flange parts are compact and under the effect of bending around their strong principal axes, are calculated with the expression given in Equation 4 according to the lateral-torsional buckling.

$$F_{cr} = \frac{C_b \cdot \pi^2 \cdot E}{(L_b/i_{ts})^2}\sqrt{1 + 0.078\frac{J \cdot c}{W_{ex} \cdot h_o}\left(\frac{L_b}{i_{ts}}\right)^2} \quad (4)$$

The effective radius of inertia used in the critical stress value formulation is as given in Equation 5.

$$i_{ts}^2 = \frac{\sqrt{I_y \cdot C_w}}{W_{ex}} \quad (5)$$

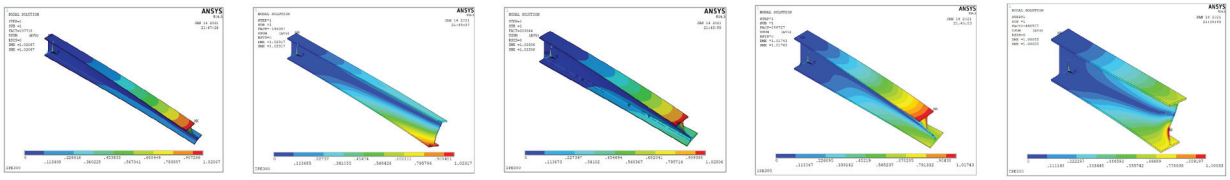
Accordingly, the critical value of the lateral-torsional buckling load of the I cross-section elements under the bending effect is calculated as given in Equation 6 in the regulations.

$$P_{cr} = \frac{F_{cr} \cdot W_{ex}}{L_b} \quad (6)$$

Calculation According to Finite Element Method

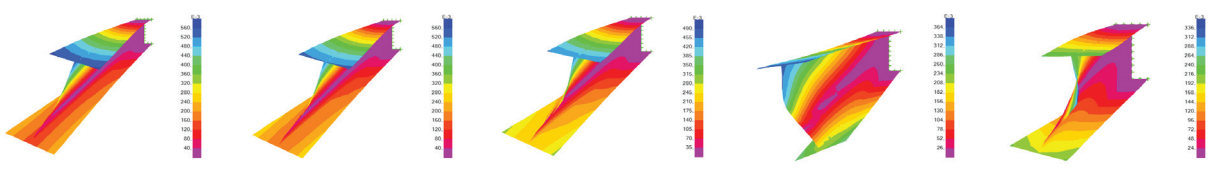
Finite element models of beams were created with the help of ANSYS, SAP2000 and LTBeam software. Critical lateral buckling load analysis was performed with the help of the created models. Accordingly, three-dimensional solid modelling of cantilever beams was created in the analysis made with ANSYS software (Figure 2) The material type of the created models was defined as SOLID187. In the analysis made according to linear elastic material properties, cantilever beams were divided into finite elements with an average range of 2.5 ~ 5 cm. Fixed support was defined at the nodal point on one side of the beam endpoints, and a 1 N unit loading was made to the shear centre on the other free end. The analysis type was selected as Eigen Buckling and the value calculated as buckling load factor at the end of the analysis gave the buckling load.

According to the analysis made by utilizing the SAP2000, the cantilever beam body and flange elements are defined using the Shell Element. (Figure 3). In the models, flange and web joints were combined at 90° angles. Cantilever beams with linear elastic material properties were divided into finite elements with an average range of 2.5 ~ 5 cm. Fixed support properties were assigned to the nodes on one side of the beam endpoints. The shear centre at the other free end was loaded with 1 N unit loading. P-Delta effects were taken into consideration by selecting the analysis type



a) L=3000 mm b) L=2500 mm c) L=2000 mm d) L=1500 mm e) L=1000 mm

Figure. 2. Cantilever beams modelled in ANSYS program.



a) L=3000 mm b) L=2500 mm c) L=2000 mm d) L=1500 mm e) L=1000 mm

Figure 3. Cantilever beams modelled in the SAP2000 program.

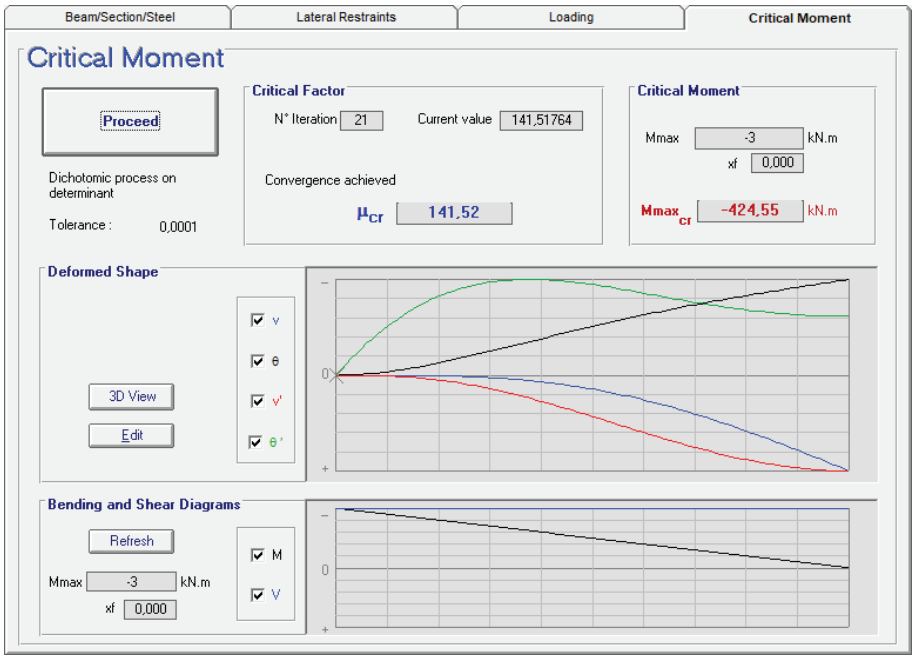


Figure 4. Cantilever beams modelled in LTBeam program.

as Buckling. As a result of the analysis, the value obtained as the buckling load factor gives the buckling load value.

LTBeam is free software developed by CTICM (Center Technique Industriel de la Construction Métallique) in France, used only for the calculation of critical moments [55]. Critical elastic lateral-torsional buckling loads can also be determined through one-dimensional finite element models, where beams are modeled according to their actual geometry using LTBeam software. (Figure 4).

The program can perform buckling analysis of both simply supported beams and cantilever beams. Limited documentation on LTBeam made interpretation of results difficult. However, several reliable sources Access Steel (2005) and ECCS (2006) refer to LTBeam as a useful program [56,57].

Optimization With Microsoft Excel Solver

Solver is the simplest and most understandable computer software used to find the optimal result. Solver is an add-in command available in Microsoft Excel. Although it is a command included in Excel, the user must enable this command. Solver is used to find the largest or smallest value of the target cell in a formula. Constraints can be developed to the values to be used in models to be developed with the Solver and these restrictions can be applied to cells [Excel - help]. Using the lateral buckling load values of the IPE profile obtained from ANSYS program with the help of Excel Solver;

- The factor γ_2 in Equation 1 given by Timoshenko and Gere was calibrated [6].
- Fixed coefficients in Equation 4 given in AISC and DCCPSS regulations were calibrated.

The steps taken for calibration were listed below.

1. The Solver command is opened from the Data tab.
2. The target cell is determined by choosing one of values such as mean absolute error (MAE), root mean square error (RMSE), and mean absolute relative error (MARE).
3. The Largest is chosen if the target cell value is desired to be as large as possible, and the Smallest is chosen if it is desired to be as small as possible. If certain value is desired to be obtained, the value option is selected, and its value is written in the box. Since the error rate was desired to be the least in the study, the smallest option and RMSE value were chosen.
4. A variable cell must be determined for each coefficient in the equations to be calibrated. Each variable cell must have a direct or indirect relationship with the target cell. In the study, cells containing the values of coefficients a, b and c were selected as variable cells. Before starting the optimization process, a random number must be defined to the coefficients a, b and c.
5. Solver is based on Nonlinear Generalized Restricted Gradient (GRG), Simple LP and expansion methods. The nonlinear Generalized Restricted Gradient (GRG) method was used in the study.
6. After clicking the Solve command, the equation was solved by the data solver and the expansion coefficients that give the smallest error value were calculated.

Error Criteria

Error criteria were used to test the accuracy of the calibrated equations for the estimation of the lateral buckling load values of the IPE profiles. Commonly used error

criteria in the literature are Mean Absolute Relative Error (MARE), mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), determination coefficient (R^2) [58,59]. In this study, MARE, MAE, MSE, RMSE and R^2 error criteria were used. The fact that MARE, MAE, MSE and RMSE values are closest to zero and R^2 value is closest to one reflects the accuracy and power of the prediction. In addition, the Nash-Sutcliffe efficiency coefficient (NSE), proposed by Nash and Sutcliffe [60], has been used in many studies to measure estimation accuracy. The variance of the estimated data compared to the variance of the observed data is a normalized statistic that determines the relative size. NSE expresses to what extent the observed and predicted data converge [60]. MARE, MAE, MSE, RMSE and NSE values were calculated from the formulas given in Equation 7-11.

$$MARE = 100 \cdot \left(\frac{1}{n} \sum_{i=1}^{n} \left| \frac{BL_{p,i} - BL_{ansys,i}}{BL_{ansys,i}} \right| \right) \tag{7}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| BL_{p,i} - BL_{ansys,i} \right| \tag{8}$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(BL_{p,i} - BL_{ansys,i} \right)^2 \tag{9}$$

$$RMSE = \frac{1}{n} \sum_{i=1}^{n} \sqrt{\left(BL_{p,i} - BL_{ansys,i} \right)^2} \tag{10}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n} \left(BL_{p,i} - BL_{ansys,i} \right)^2}{\sum_{i=1}^{n} \left(BL_{ansys,i} - \overline{BL_{ansys}} \right)^2} \tag{11}$$

RESULTS AND DISCUSSION

Analytical and Numerical Findings

The critical lateral buckling load of I-section cantilever beams was calculated using a total of five different methods. The calculated buckling loads were obtained by applying a single load to the centre of shear at the free end of the beam. Although there are differences when the profile section increases and cantilever length decreases, the buckling load values calculated using the closed form equations and LTBeam program are close to each other. On the other hand, when the profile section is reduced and the cantilever length is increased, differences occur in the FEA results. However, the buckling load values calculated with the help of FEA generally confirmed each other. As a result, the results of the first three methods and the last two methods given in Table 2 are quite different from each other. This can be clearly seen in Figure 5.

Table 2. Lateral buckling load values calculated with different methods

Profile Type	Beam Length (mm)	Timoshenko (kN)	AISC- DCCPSS (kN)	LTBeam (kN)	ANSYS (kN)	SAP2000 (kN)
IPE 100	**1000**	40.18	39.36	40.44	37.44	11.23
IPE 120		74.03	71.94	73.83	69.76	35.05
IPE 140		127.87	125.38	126.77	113.22	66.57
IPE 160		209.12	208.97	207.47	121.78	98.05
IPE 180		329.12	334.59	324.40	194.45	132.95
IPE 200		497.23	516.59	493.53	203.16	170.85
IPE 220		777.32	800.37	749.76	297.87	219.68
IPE 240		1147.63	1197.99	1110.30	319.21	273.74
IPE 270		2146.85	1961.40	1771.40	479.70	353.49
IPE 300		4032.70	3103.90	2755.10	492.58	459.98
IPE 330		5881.12	4446.00	3923.00	676.51	557.10
IPE 360		8524.97	6392.55	5608.20	732.06	688.53
IPE 400		11748.40	8958.72	7819.20	854.77	832.07
IPE 450		16266.63	12778.91	11093.00	1070.22	1029.68
IPE 500		22154.84	18114.12	15656.00	1279.80	1254.17
IPE 550		29518.50	24797.49	21391.00	1551.16	1479.64
IPE 600		39465.72	34300.18	29512.00	1974.66	1830.02

Table 2. Lateral buckling load values calculated with different methods (*continued*)

Profile Type	Beam Length (mm)	Timoshenko (kN)	AISC- DCCPSS (kN)	LTBeam (kN)	ANSYS (kN)	SAP2000 (kN)
IPE 100	**1500**	15.09	15.26	15.12	14.96	0.93
IPE 120		26.78	26.45	26.95	26.04	4.20
IPE 140		45.22	44.05	45.43	41.66	14.30
IPE 160		73.31	71.32	73.39	62.39	33.28
IPE 180		113.51	110.87	112.95	85.32	59.11
IPE 200		171.76	168.80	170.06	114.40	86.15
IPE 220		256.21	256.34	254.27	146.67	119.08
IPE 240		375.37	380.28	372.89	189.23	153.26
IPE 270		586.02	609.80	580.41	234.77	200.59
IPE 300		916.78	952.75	884.95	296.73	263.50
IPE 330		1287.18	1359.60	1251.80	366.42	324.36
IPE 360		1784.93	1945.82	1775.40	452.60	410.13
IPE 400		2835.20	2718.34	2458.80	556.62	506.80
IPE 450		4704.91	3860.99	3457.90	692.70	644.73
IPE 500		7111.55	5457.19	4850.20	861.62	814.93
IPE 550		9850.19	7460.46	6604.60	1076.95	985.93
IPE 600		13685.47	10302.54	9079.00	1330.55	1273.27
IPE 100	**2000**	7.68	8.10	7.71	8.48	0.15
IPE 120		13.44	13.63	13.47	13.46	0.68
IPE 140		22.29	22.02	22.38	21.88	2.55
IPE 160		35.55	34.89	35.83	34.14	7.62
IPE 180		54.40	52.93	54.61	49.01	19.25
IPE 200		81.83	79.58	81.76	69.19	37.20
IPE 220		121.28	118.55	121.17	93.14	62.89
IPE 240		178.48	174.29	176.73	125.52	90.56
IPE 270		273.25	273.28	270.66	157.69	127.37
IPE 300		413.59	420.70	406.92	203.64	172.04
IPE 330		579.28	597.66	572.20	250.62	215.13
IPE 360		821.91	850.50	806.09	315.28	275.33
IPE 400		1148.93	1183.51	1110.20	383.08	341.32
IPE 450		1596.85	1671.94	1548.30	501.56	435.68
IPE 500		2190.69	2354.40	2158.40	623.86	554.99
IPE 550		2930.69	3212.96	2928.90	756.28	678.74
IPE 600		4677.88	4427.48	4010.90	940.96	882.53
IPE 100	**2500**	4.62	5.04	4.64	4.66	0.04
IPE 120		5.01	8.32	7.98	8.06	0.16
IPE 140		13.03	13.20	13.09	13.14	0.61
IPE 160		20.70	20.60	20.79	20.36	1.90
IPE 180		31.10	30.68	31.44	30.21	5.34
IPE 200		46.58	45.66	46.85	43.58	12.55
IPE 220		68.88	66.93	69.00	60.83	27.17
IPE 240		100.34	97.60	100.30	83.97	47.49
IPE 270		153.44	149.81	152.05	111.37	78.71
IPE 300		227.62	227.15	226.39	145.40	115.46
IPE 330		320.72	321.15	317.08	183.84	150.84
IPE 360		448.53	454.21	444.46	230.65	198.14
IPE 400		619.07	629.30	609.04	287.02	248.05
IPE 450		847.93	883.59	843.75	357.03	318.18
IPE 500		1204.29	1238.94	1170.30	446.34	407.00
IPE 550		1637.31	1687.23	1583.30	562.43	500.25
IPE 600		2235.76	2319.16	2160.30	703.94	650.52

Table 2. Lateral buckling load values calculated with different methods (*continued*)

Profile Type	Beam Length (mm)	Timoshenko (kN)	AISC- DCCPSS (kN)	LTBeam (kN)	ANSYS (kN)	SAP2000 (kN)
IPE 100	**3000**	3.08	3.44	3.08	3.11	0.01
IPE 120		3.29	5.62	5.25	5.32	0.05
IPE 140		3.54	8.81	8.53	8.49	0.19
IPE 160		13.37	13.61	13.44	13.88	0.59
IPE 180		20.06	20.00	20.18	19.94	1.69
IPE 200		29.90	29.54	29.96	28.98	4.18
IPE 220		43.54	42.73	43.88	41.22	10.21
IPE 240		63.10	61.89	63.59	59.46	21.10
IPE 270		95.81	93.21	95.68	80.80	43.34
IPE 300		142.61	139.31	141.52	107.71	73.51
IPE 330		200.10	196.04	197.65	139.41	104.75
IPE 360		276.97	275.53	276.06	186.65	145.01
IPE 400		380.57	380.04	376.98	222.18	186.48
IPE 450		524.17	530.18	519.59	287.56	242.90
IPE 500		727.46	739.97	717.59	348.59	313.45
IPE 550		979.80	1005.44	968.32	452.78	387.04
IPE 600		1319.75	1378.19	1317.40	551.96	504.52

The existing analytical method formulations have been optimized according to the FEA analysis results obtained by utilizing the ANSYS program. There are two main reasons for choosing the buckling load values to be referenced in FEA calculations from the ANSYS program instead of SAP2000. Firstly, cantilever beams are modelled as Solid elements in the ANSYS program and as Shell elements in SAP2000. Secondly, in SAP2000, while the web and flanges are joined perpendicular to each other at an angle of 90°; The web and flange joints of the models in ANSYS are curvilinear and exactly the same as the real profile geometry. For these reasons, the ANSYS program was used for the buckling load values taken as a reference within the scope of the research. The comparisons of the buckling load values of the IPE series cantilever beams of five different lengths calculated by five different methods are as given in Figure 5.

Optimization Technique Findings with Microsoft Excel Solver

In the second stage of the study, the a and b coefficients in Equation 12 were calibrated with the help of Excel–Solver, keeping the relationship between the γ_2 factor and the L^2C/C_1 ratio. The analyses have been conducted on a computer with an AMD Ryzen 7 PRO 3700 8-Core 3.60 GHz processor and 8 GB RAM. In addition, the c, d and f coefficients in Equation 13 of AISC and DCCPSS Regulations were calibrated with the help of Excel–Solver. At this stage, two different calibration processes were carried out using two different equations. The graph of the relationship between the ratio L^2C/C_1 given in Table 1 and γ_2 is given in Figure 6. As can be seen in Figure 6, there is an exponential relationship between the ratio L^2C/C_1 and γ_2 as in Equation 12. Therefore, the coefficients of Equation 12 were optimized to determine the γ_2 factor. In the advancing age of science, there are new methods for calculating lateral buckling load values as well as the Timoshenko and Gere [6] equation. The c, d and f coefficients of Equation 13 were calibrated by adhering to AISC and DCCPSS Regulations and using the lateral buckling load values obtained from the ANSYS program.

$$\gamma_2 = a \cdot \left(L^2 \cdot \frac{C}{C_1}\right)^b \tag{12}$$

$$F_{cr} = \frac{C_b \cdot \pi^2 \cdot E}{\left(\frac{L_b}{i_{ts}}\right)^2}\left(c + d \cdot \frac{J_C}{W_{ex} \cdot h_o} \cdot \left(\frac{L_b}{i_{ts}}\right)^2\right)^f \tag{13}$$

c= 1, d= 0.078, f= 0.5

The coefficients of Equation 12 and Equation 13 were calibrated using Excel Solver so that the RMSE error criterion was the smallest. The coefficients of the calibrated equations are given in Table 3. As seen in Table 3, the new coefficients are different from each other. Comparison criteria for two different calibrated equations are given in Table 4. The approximate MSE values of the calibrated equations were obtained as 890, MARE values of 50 and

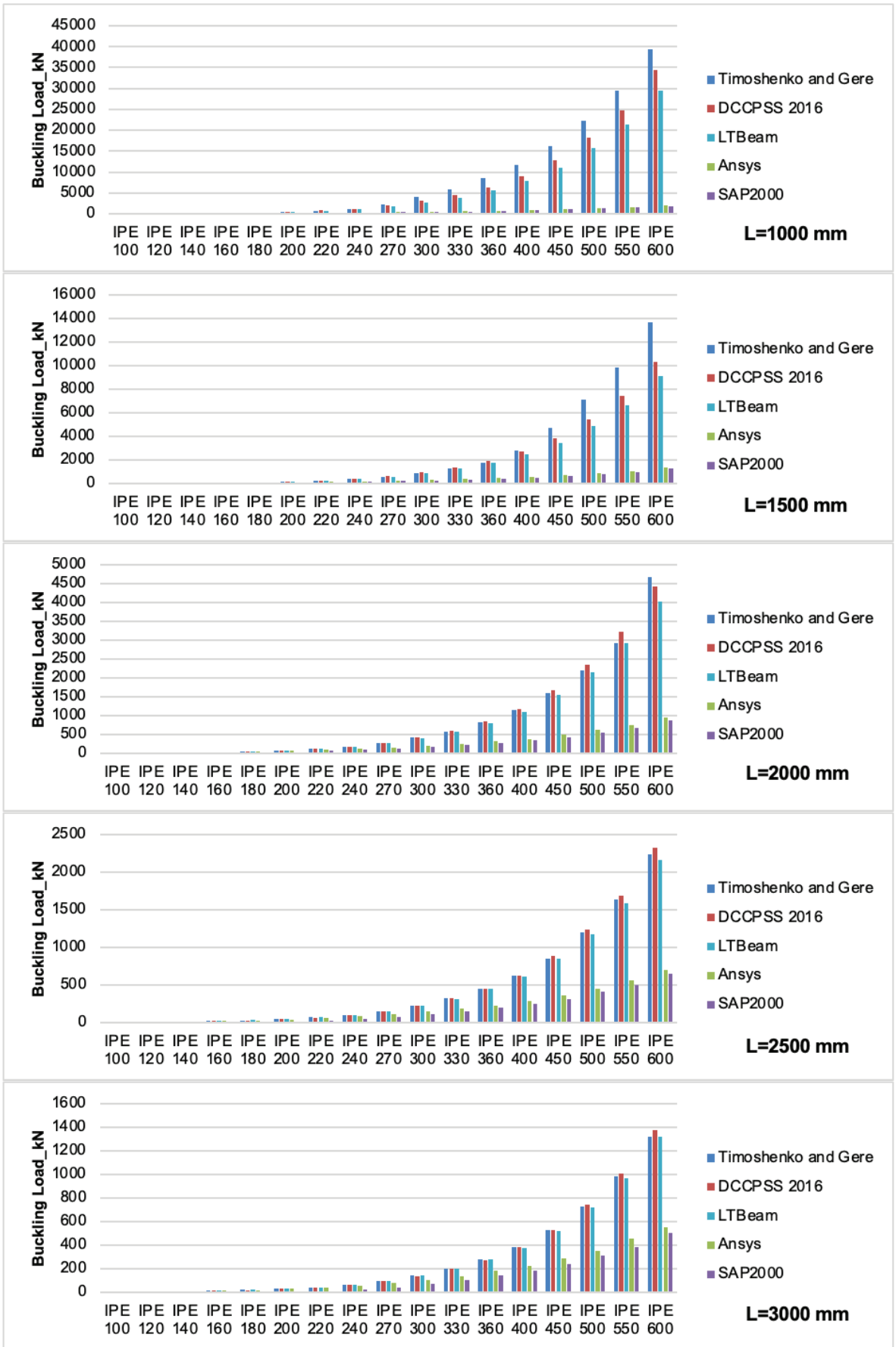


Figure 5. Lateral buckling loads of cantilever beams of different lengths.

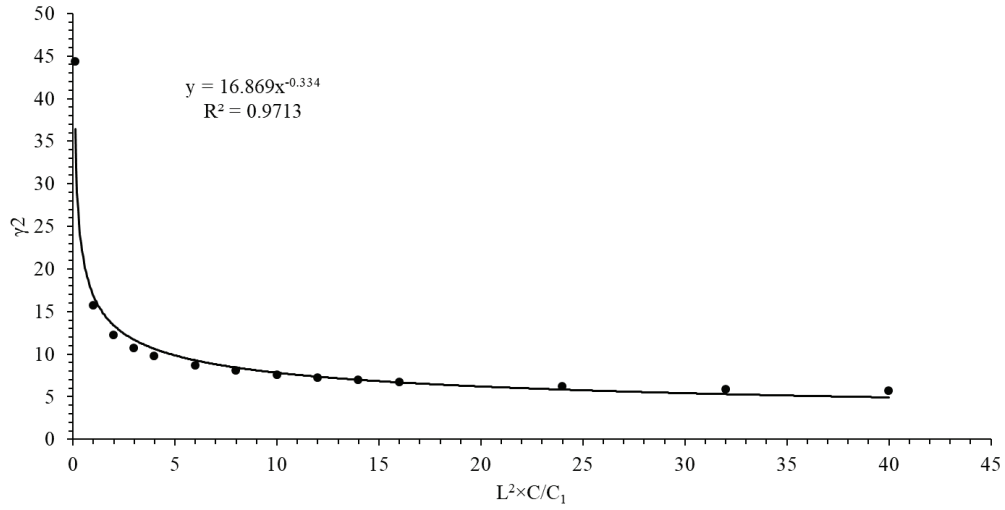


Figure 6. Relationship between L^2C/C_1ratio and γ2

Table 3. Coefficients for calibrated models

Models	**a**	**b**	
Caliber Model1	3.896068277	0.419127105	
	c	**d**	**f**
Caliber Model2	0	0.160835288821455	0.919080100568975

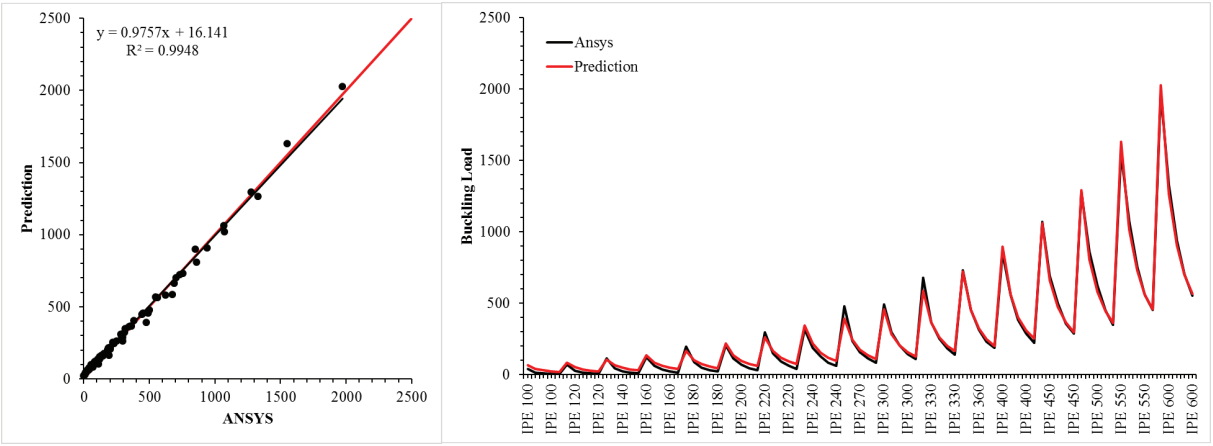


Figure 7. Comparisons of the prediction and ANSYS values of the buckling load for Calibration Timoshenko and Gere equation

MAE values of 24. NSE error value was used because of high MSE values. An NSE value greater than 0.9 indicates that the estimate is correct. The scatter plots in Figures 7-8 are quite good as there is no deviation from x=y (45°).

In recent years, machine learning research in the construction industry have used a maximum of 3 performance criteria [47,49,50,52,53,61–65]. The equations created utilizing the six performance criteria have undergone a thorough analysis in this study. Six performance parameters were used to objectively assess the Caliber models' performance.

The correlation coefficient R has been utilized as a performance criterion in machine learning studies in the field of construction the most frequently in recent years [49, 50, 52, 53, 63–67]. The R value for these studies in the literature was discovered to be somewhere in the range of 0.9, and

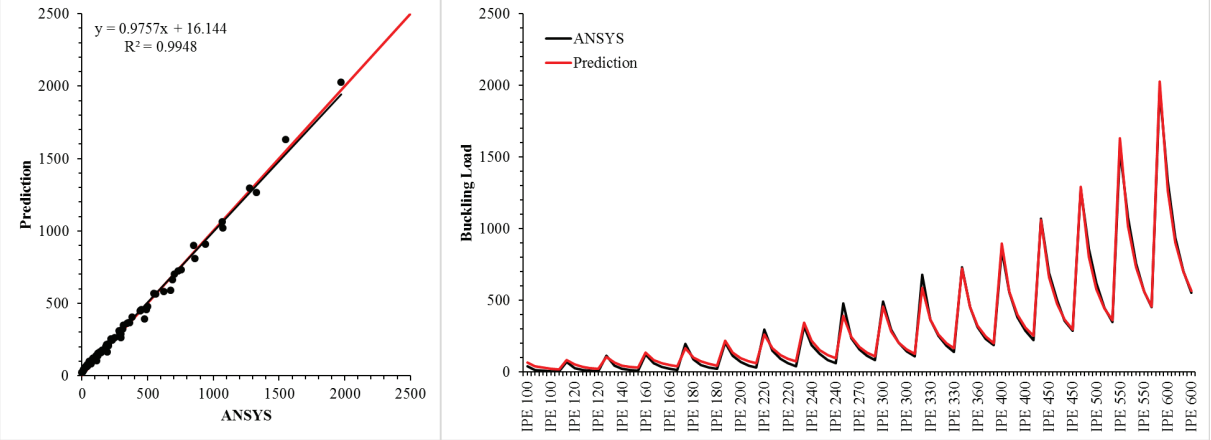


Figure 8. Comparisons of the prediction and ANSYS values of the buckling load for Calibration AISC - DCCPSS

Table 4. Comparison criteria of modified equations

	NSE	MAE	MARE	MSE	RMSE	R2
Caliber Model1	0.994	24.30	51.56	892.56	29.88	0.995
Caliber Model2	0.994	24.29	51.47	891.46	29.86	0.995

the correlation between the actual predicted value and the observed value was only somewhat stronger. Additionally, several researchers used the R^2 criterion to assess the effectiveness of the models they created for machine learning investigations [47,50,53,61]. The R^2 criteria was determined to be more than 0.9 in these investigations. This finding demonstrates that prediction and observation have a stronger link. The determination factor R^2 criteria was utilized in this work instead of the R coefficient, which would have been misleading for assessing the performance of the models. Tables 4 show that the Caliber models' findings meet the R^2 criteria of larger than 0.9 and that there is a stronger connection between prediction and observation.

The amount of time that passed while the suggested strategy was being used is another important finding. It took 0.173 seconds to optimize the coefficients. With the developing technology, the proposed approach for calculating the buckling load is not only accurate but also significantly faster. Thus, decreases the amount of computation necessary to carry out such an analysis.

Excel–Solver has been used in optimization studies in the field of construction in recent years [36,38–40]. However, it was used for the first time in the field of construction in the optimization of the coefficients of an equation in this study.

CONCLUSION

Analytical solutions become very complex when the beam end conditions are a cantilever beam with a fixed support as opposed to a simple support. Therefore, numerical approaches such as the finite element method are needed to solve the fundamental differential equilibrium equations. Despite having various types of smart automating, finite element types, and executed analyses, two identical models examined using the finite element method in two distinct verified software's should have produced findings that are equal. Because the mesh cannot be adjusted, the usage of isoperimetric components on SAP2000 can only be used to make initial estimates for early phases of study. Finite elements with more nodes or integration points are not an option in SAP2000. ANSYS software, which is more suitable for research than design, has been preferred in finite element analysis due to the advantage of the selection and full control of different finite element meshes, where different behaviour rules of materials can be applied, and possible complex analysis applications.

In today's steel specifications, no distinction is made in the design and calculation methods of simply supported beams and cantilever beams under the buckling effect. However, the buckling zones on the steel beams change according to the end support conditions. Considering these situations in differential equilibrium equations creates quite complex problems in generating analytical solutions. Instead, existing analytical formulas were optimized for cantilever beams with reference to finite element analysis. Thanks to the results obtained, the buckling load calculation of cantilever beams can be successfully solved with the help of renewed closed form equations. This renewed

formula can make very high accuracy predictions, which can be an alternative to finite element analysis.

The proposed approach in optimizing the coefficients has taken only 0.173 seconds, it can be concluded that employing Excel-Solver reduces the computational cost that is required to conduct for buckling loads of IPE-Section. 85 numerical calculations were conducted to determine the buckling load of IPE cantilever beams with varied length. Therefore, only cantilever beams with the IPE section are compatible with the calibrated equation created using the Excel-Solver. Other I-section cantilever beams require a significantly wider variety of numerical analysis. In further studies, experimentally validate the proposed method and its results is recommended for buckling load calculations.

NOTATION

P_{cr}	Critical buckling load
L	Beam length
E	Modulus of elasticity
I_y	Moment of inertia about the weak axis
C	Torsional stiffness
C_1	Distortion stiffness
C_b	Moment correction coefficient
F_{cr}	Critical Stress
L_b	Length of element not supported by stability joint
i_{ts}	Effective radius of inertia
J	Torsional constant
C_w	Distortion constant
W_{ex}	Elastic section modulus about the strong axis
h_o	The distance between the centres of gravity of the cross-section flanges
γ_2	a dimensionless coefficient
M_{maks}	The absolute value of the maximum bending moment along the length of the laterally unsupported beam
M_A	The absolute value of the bending moment at 1/4 point of the laterally unsupported beam length.
M_B	The absolute value of the bending moment at 1/2 point of the laterally unsupported beam length.
M_C	The absolute value of the bending moment at 3/4 point of the laterally unsupported beam length.
BL_p	Buckling load estimated by calibrated models
BL_{ansys}	Buckling load obtained from ANSYS analysis
n	Length of the series
BL_{ansys}	Average of the buckling load obtained from ANSYS analysis

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

AUTHOR CONTRIBUTION

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Ahmet Özbayrak, Mohammed Kamal Ali and Hatice Çıtakoğlu. The first draft of the manuscript was written by Ahmet Özbayrak, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

CONFLICT OF INTEREST

On behalf of all authors, the corresponding author states that there is no conflict of interest.

ETHICAL APPROVAL

This article does not contain any studies with human participants or animals performed by any of the authors.

FUNDING DETAILS

No funds, grants, or other support was received.

REFERENCES

[1] Trahair NS. Steel cantilever strength by inelastic lateral buckling. J Constr Steel Res 2010;66:993–999. [CrossRef]

[2] American Institute of Steel Construction. Specification for structural steel buildings. Chicago: American Institute of Steel Construction; 2011.

[3] Turkish Republic - Ministry of Envirounment and Urbanization. Çelik yapıların tasarım, hesap ve yapım esaslarına dair yönetmelikte değişiklik yapılmasına dair yönetmelik. Available at: https://www.resmigazete.gov.tr/eskiler/2018/02/20180215M1-4.htm. Accessed on Jun 12, 2024.

[4] Chen WF, Atsuta T. Theory of beam-columns: Space behavior and design. McGraw-Hill, New York: J. Ross Publishing; 1977.

[5] Chen WF. Structural stability: Theory and implementation. 1st ed. New York: Prentice Hall; 1987.

[6] Timoshenko SP, Gere JM, Prager W. Theory of elastic stability, second edition. J Appl Mech 1962;29:220–221. [CrossRef]

[7] Trahair NS. Inelastic lateral buckling of steel cantilevers. Eng Struct 2020;208:109918. [CrossRef]

[8] ANSYS. ANSYS Help. Release 145 Copyr 2012.

[9] Sapfire. SAP2000 Structural analysis and design. Available at: https://www.csiamerica.com/products/sap2000. Accessed on Jun 26, 2024.

[10] Gonenli C, Das O. Effect of crack location on buckling and dynamic stability in plate frame structures. J Braz Soc Mech Sci Eng 2021;43:311. [CrossRef]

[11] Demirhan AL, Eroğlu HE, Mutlu EO, Yılmaz T, Anil Ö. Experimental and numerical evaluation of inelastic lateral-torsional buckling of I-section cantilevers. J Constr Steel Res 2020;168:105991. [CrossRef]

[12] Samanta A, Kumar A. Distortional buckling in braced-cantilever I-beams. Thin-Walled Struct 2008;46:637–645. [CrossRef]
[13] Ozbasaran H, Aydin R, Dogan M. An alternative design procedure for lateral-torsional buckling of cantilever I-beams. Thin-Walled Struct 2015;90:235–242. [CrossRef]
[14] Ma M, McNatt T, Hays B, Hunter S. Elastic lateral distortional buckling analysis of cantilever I-beams. Ships Offshore Struct 2013;8:261–269. [CrossRef]
[15] Andrade A, e Costa PP, Camotim DRZ. Elastic lateral-torsional buckling of restrained web-tapered I-beams. Comput Struct 2010;88:1179–1196. [CrossRef]
[16] Zhang WF, Liu YC, Hou GL, Chen KS, Ji J, Deng Y, et al. Lateral-torsional buckling analysis of cantilever beam with tip lateral elastic brace under uniform and concentrated load. Int J Steel Struct 2016;16:1161–1173. [CrossRef]
[17] Winatama Kurniawan C, Mahendran M. Elastic lateral buckling of cantilever litesteel beams under transverse loading. Int J Steel Struct 2011;11:395–407. [CrossRef]
[18] Yılmaz T, Kiraç N. Analytical and parametric investigations on lateral torsional buckling of European IPE and IPN beams. Int J Steel Struct 2017;17:695–709. [CrossRef]
[19] Prombut P, Anakpotchanakul C. Deflection of composite cantilever beams with a constant I-cross section. IOP Conf Ser Mater Sci Eng 2019;501:012025. [CrossRef]
[20] Ozbasaran H, Yilmaz T. Shape optimization of tapered I-beams with lateral-torsional buckling, deflection and stress constraints. J Constr Steel Res 2018;143:119–130. [CrossRef]
[21] Rahami H, Kaveh A, Glolipour Y. Sizing, geometry and topology optimization of trusses via force method and genetic algorithm. Eng Struct 2008;30:2360–2369. [CrossRef]
[22] Camp CV, Bichon BJ, Stovali SP. Design of steel frames using ant colony optimization. J Struct Eng 2005;131:369–379. [CrossRef]
[23] Hayalioglu MS, Degertekin SO. Design of non-linear steel frames for stress and displacement constraints with semi-rigid connections via genetic optimization. Struct Multidiscip Optim 2004;27:259–271. [CrossRef]
[24] Lagaros ND, Papadrakakis M. Robust seismic design optimization of steel structures. Struct Multidiscip Optim 2007;33:457–469. [CrossRef]
[25] Lagaros ND, Papadrakakis M. Seismic design of RC structures: A critical assessment in the framework of multi-objective optimization. Earthq Eng Struct Dyn 2007;36:1623–1639. [CrossRef]
[26] Yassami M, Ashtari P. Using fuzzy genetic algorithm for the weight optimization of steel frames with semi-rigid connections. Int J Steel Struct 2014;15:63–73. [CrossRef]
[27] Dehghani S, Fathizadeh SF, Vosoughi AR, Farsangi EN, Yang TY, Hajirasouliha I. Development of a novel cost-effective toggle-brace-curved damper (TBCD) for mid-rise steel structures using multi-objective NSGA II optimization technique. Struct Multidiscip Optim 2020;63:661–688. [CrossRef]
[28] Prendes-Gero MB, Bello-García A, del Coz-Díaz JJ, Suárez-Domínguez FJ, Gero MBP. Optimization of steel structures with one genetic algorithm according to three international building codes. Rev Constr 2018;17:47–59. [CrossRef]
[29] Kaveh A, Moghanni RM, Javadi S. Optimum design of large steel skeletal structures using chaotic firefly optimization algorithm based on the Gaussian map. Struct Multidiscip Optim 2019;60:879–894. [CrossRef]
[30] Perelmuter A, Yurchenko V. Parametric optimization of steel shell towers of high-power wind turbines. Procedia Eng 2013;57:895–905. [CrossRef]
[31] Shaqfa M, Orban Z. Modified parameter-setting-free harmony search (PSFHS) algorithm for optimizing the design of reinforced concrete beams. Struct Multidiscip Optim 2019;60:999–1019. [CrossRef]
[32] Omkar SN, Khandelwal R, Ananth TVS, Naik GN GS. Quantum behaved Particle Swarm Optimization (QPSO) for multi-objective design optimization of composite structures. Expert Syst Appl 2009;36:11312–11322. [CrossRef]
[33] Barraza M, Bojórquez E, Fernández-González E, Reyes-Salazar A. Multi-objective optimization of structural steel buildings under earthquake loads using NSGA-II and PSO. Design Optim Appl Civil Eng 2017;21:488–500. [CrossRef]
[34] Błachut J, Magnucki K. Strength, stability, and optimization of pressure vessels: Review of selected problems. Appl Mech Rev 2008;61:060801. [CrossRef]
[35] Meddaikar YM, Irisarri FX, Abdalla MM. Laminate optimization of blended composite structures using a modified Shepard's method and stacking sequence tables. Struct Multidiscip Optim 2017;55:535–546. [CrossRef]
[36] Cho HK. Optimization of laminated composite cylindrical shells to maximize resistance to buckling and failure when subjected to axial and torsional loads. Int J Precis Eng Manuf 2018;19:85–95. [CrossRef]
[37] Msabawy A Mohammad F. Practical analysis procedures of steel portal frames having different connections rigidities using modified stiffness matrix and end-fixity factor concept. IOP Conf Ser Mater Sci Eng 2019;518:022037. [CrossRef]
[38] Kirsch U. A unified reanalysis approach for structural analysis, design, and optimization. Struct Multidiscip Optim 2003;25:67–85. [CrossRef]
[39] Msabawy A, Mohammad F. Continuous sizing optimization of cold-formed steel portal frames with semi-rigid joints using generalized reduced gradient algorithm. Mater Today Proc 2021;42:2290–2300. [CrossRef]

[40] Taki T. Optimization of flat Z-stiffened panel subjected to compression. Trans Japan Soc Aero Space Sci 2019;62:44–54. [CrossRef]
[41] Vosoughi AR. A developed hybrid method for crack identification of beams. Smart Struct Syst 2015;16:401–414. [CrossRef]
[42] Le LM, Ly H-B, Pham BT, Le VM, Pham TA, Nguyen D-H, et al. Hybrid artificial intelligence approaches for predicting buckling damage of steel columns under axial compression. Materials (Basel) 2019;12:1670. [CrossRef]
[43] Jung ID, Shin DS, Kim D, Lee J, Lee MS, Son HJ, et al. Artificial intelligence for the prediction of tensile properties by using microstructural parameters in high strength steels. Materialia 2020;11:100699. [CrossRef]
[44] Cuong-Le T, Nghia-Nguyen T, Khatir S, Trong-Nguyen P, Mirjalili S, Nguyen KD. An efficient approach for damage identification based on improved machine learning using PSO-SVM. Eng Comput 2022;38:3069–3084. [CrossRef]
[45] Daş O, Bağci Daş D. İzotropik plakaların regressif topluluk öğrenmesi kullanarak serbest titreşim analizi. Eur J Sci Technol 2022:428–434. [CrossRef]
[46] Özbayrak A, Ali MK, Çıtakoğlu H. Buckling load estimation using multiple linear regression analysis and multigene genetic programming method in cantilever beams with transverse stiffeners. Arab J Sci Eng 2023;48:5347–5370. [CrossRef]
[47] Sharifi Y, Tohidi S. Lateral-torsional buckling capacity assessment of web opening steel girders by artificial neural networks - Elastic investigation. Front Struct Civ Eng 2014;8:167–177. [CrossRef]
[48] Onchis DM, Gillich GR. Stable and explainable deep learning damage prediction for prismatic cantilever steel beam. Comput Ind 2021;125:103359. [CrossRef]
[49] Kamane SK, Patil NK, Patagundi BR. Prediction of twisting performance of steel I beam bonded exteriorly with fiber reinforced polymer sheet by using neural network. Mater Today Proc 2021;43:514–519. [CrossRef]
[50] Nguyen TA, Ly HB, Tran VQ. Investigation of ANN architecture for predicting load-carrying capacity of castellated steel beams 2021;6697923. [CrossRef]
[51] Limbachiya V, Shamass R. Application of artificial neural networks for web-post shear resistance of cellular steel beams. Thin Walled Struct 2021;161:107414. [CrossRef]
[52] Hosseinpour M, Rossi A, de Souza ASC, Sharifi Y. New predictive equations for LDB strength assessment of steel-concrete composite beams. Eng Struct 2022;258:114121. [CrossRef]
[53] Mohanty N, Sasmal SK, Uttam, Mishra UK, Sahu SK. Experimental and computational analysis of free in-plane vibration of curved beams. J Vib Eng Technol 2022;11:1777–1796. [CrossRef]
[54] Neves M, Basaglia C, Camotim D. Stiffening optimisation of conventional cold-formed steel cross-sections based on a multi-objective Genetic Algorithm and using Generalised Beam Theory. Thin Walled Struct 2022;179:109713. [CrossRef]
[55] Galéa Y.) LTBeam Version 1.0. 11. CTICM Fr 2012.
[56] Access Steel. NCCI: Elastic critical moment for lateral torsional buckling. Available at: https://www.steelconstruction.info/images/0/0f/SN003b.pdf. Accessed Jun 26, 2024.
[57] ECCS. Rules for Member Stability in EN 1993-1-1: Background documentation and design guidelines. Available at: https://store.steelconstruct.com/site/index.php?module=store&target=publicStore&id_category=13&id=19. Accessed Jun 26, 2024.
[58] Karunanithi N, Grenney WJ, Whitley D, Bovee K. Neural networks for river flow prediction. J Comput Civ Eng 1994;8:201–220. [CrossRef]
[59] Citakoglu H. Comparison of artificial intelligence techniques for prediction of soil temperatures in Turkey. Theor Appl Climatol 2017;130:545–556. [CrossRef]
[60] Nash JE, Sutcliffe JV. River flow forecasting through conceptual models part I - A discussion of principles. J Hydrol 1970;10:282–290. [CrossRef]
[61] Graciano C, Kurtoglu AE, Casanova E. Machine learning approach for predicting the patch load resistance of slender austenitic stainless steel girders. Structures 2021;30:198–205. [CrossRef]
[62] Ferreira FPV, Shamass R, Limbachiya V, Tsavdaridis KD, Martins CH. Lateral-torsional buckling resistance prediction model for steel cellular beams generated by Artificial Neural Networks (ANN). Thin Walled Struct 2022;170:108592. [CrossRef]
[63] Sharifi Y, Moghbeli A, Hosseinpour M, Sharifi H. Neural networks for lateral torsional buckling strength assessment of cellular steel I-beams. 2019;22:2192–2202. [CrossRef]
[64] Abambres M, Rajana K, Tsavdaridis KD, Ribeiro TP. Neural network-based formula for the buckling load prediction of I-section cellular steel beams. Computers 2019;8:2. [CrossRef]
[65] Moghbeli A, Sharifi Y. New predictive equations for lateral-distortional buckling capacity assessment of cellular steel beams. Structures 2021;29:911–923. [CrossRef]
[66] Hosseinpour M, Moghbeli A, Sharifi Y. Evaluation of lateral-distortional buckling strength of castellated steel beams using regression models. Innov Infrastruct Solut 2021;6:142. [CrossRef]

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 988–1008, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00084

Research Article

Automatic unit test generator software by collecting run-time data

Sevdanur GENÇ[1,*]

[1]*Department of Computer Technologies, Taşköprü Vocational School, Kastamonu University, Kastamonu, 37400, Türkiye*

ARTICLE INFO

Article history
Received: 17 November 2022
Revised: 05 January 2023
Accepted: 04 April 2023

Keywords:
Byte Code; Java Agent; Software Testing; The Opcode Parsing Method; Unit Test Generation

ABSTRACT

Quality and productivity needs are considered together in software. For this reason, any existing software should be tested automatically with test automation. Software test automations is automated software testing activities. Automating constantly manually written tests, on the other hand, saves time, reduces error rates, produces better quality software, and reduces costs. This study aims to produce an automatic unit testing framework that is planned to work in run-time on software products. This developed application performs unit test transformations that can respond to the desired test scenarios on the product being studied.. Java agent is used as the basis of all these transformations. All information about the objects, methods, and variables of the sample java classes to be worked on is converted into data in run-time using byte code. During this transformation, information is saved in the database, and unit tests are created automatically through the template engine. Compared to the products developed on automatic unit test generation in the literature, the opcode parsing method was developed for this study. This method reads a byte code at run-time, uses the properties of the java class it belongs to, and automatically creates the unit test class and test methods. The study can also examine different object definitions and conditional and loop structures within a method and produce alternative test scenarios. The automatic unit test scenario produced has been turned into a flexible framework that can encounter minimum errors at run-time.
Considering the scarcity of studies in the field of national software testing; It is thought that the automatic unit test generation product developed within the scope of this study, using byte code, will contribute to the work area.

Cite this article as: Genç S. Automatic unit test generator software by collecting run-time data. Sigma J Eng Nat Sci 2024;42(4):988–1008.

INTRODUCTION

The main purpose of designing the Java programming language, developed by Sun Microsystems and made available in 1995, was to develop a portable, easy-to-learn, general-purpose, platform-independent, object-oriented programming language. The Java compiler converts the source code into java bytecode, which is a platform-independent intermediate language. This code is then processed and run through the java virtual machine on each platform. Java agents are used to flexibly modify the application logic executed by the JVM at this run-time of the code. A java agent is a specially crafted jar file. This file uses the

***Corresponding author.**
*E-mail address: sgenc@kastamonu.edu.tr

This paper was recommended for publication in revised form by Editor in-Chief Ahmet Selim Dalkilic

Instrumentation API to replace the existing bytecode loaded in the JVM. The most important feature of Java agent tools is that they can redefine or modify classes at run-time. They can modify method bodies by redefining their constant and variable properties. They can also change the signatures or inheritance properties of methods.

Quality and productivity needs are considered together in software. For this reason, factors such as a fluent algorithm and strong risk management are needed to test the existing software's compliance with these criteria. In order to fulfill these factors, there are serious responsibilities in the test area. Fast tests and high accuracy of the results are the factors that make a difference in software development. In order to realize this factor, software test automation is needed.

The focus of projects on software is software testing processes. At the end of a successful test process, highly accurate software with the least errors is produced. Studies on software quality in Turkey show that the need for the test-driven software development process and test tools is increasing in our country [1]. In test-driven software development, the target is to write a testable code first, with the scenario belonging to this code, before writing the code that will do the necessary work. After various software development principles design this testable code, if it gives a result with high accuracy, that software has passed the test successfully. If the test results are unsuccessful, it is returned to the beginning, the code is examined, and the problem is tried to be corrected. In software projects, tests called unit tests are written to prove that each unit (class or method) works flawlessly. Unit tests facilitate and accelerate the software development process and ensure that each class and method works correctly.

There are two of the most well-known basic testing frameworks on Java platforms. These are JUnit and TestNG [2]. Both are powerful enough to allow testing in complex test cases on exactly the requested code snippets. Junit is an open-source framework for writing and running repeatable tests. JUnit; runs test data with various test cases to test the expected results from the program. TestNG is more functional and easier to use than Junit. TestNG supports running test cases in parallel on test threads. It also has many features, such as flexible test configurations, detailed analysis of error messages, advanced archiving, and plugin support for editors.

The developer manually codes all these unit tests. Automating the tests is recommended to use the time in a quality manner and speed up the business traffic. Therefore, the speed and accuracy of testing tools are important. For example, keyword-driven scripts have significant advantages. In this approach, the size of the software being tested is important, not the number of tests. This greatly reduces the script maintenance cost and speeds up the implementation of automated tests. At the same time, in order to achieve success in test automation, scripts and data must be reusable. This eliminates repeated tasks and speeds up the implementation of new tests. However, automated tests can also help prevent writing errors compared to manual tests [3].

This study aims to automatically produce unit tests that are planned to work on products in run-time. This software has been developed as a desktop application and can perform unit test transformations on the specified java classes. All these transformations are based on java agents and bytecode. All information about the objects, methods and variables of the sample java classes to be worked on has been converted into data in run-time. During this transformation, information is both saved in the database and unit tests are created automatically through the template engine. However, in this developed study, both the byte code side is analyzed with the improved opcode parsing method and tests are produced on the desired units. At the same time, alternative scenarios were produced for each different use of the objects used in the classrooms. These scenarios can run as soon as the test is applied and perform automatic unit test generation. This automatic unit test generation tool, which can generate unit tests in accordance with the usage differences of each object, has been turned into a flexible framework that can encounter minimum errors at run-time.

Various approaches for automated unit test generation have been presented in the literature. The contributions of the study to the literature are as follows;

1. An opcode parsing method has been developed with the help of java string functions to work during byte code conversions. With this method, values such as objects, variables and input-output parameters against each opcode are distinguished and these data are listed in JSON format. The developed opcode parsing method is open to be developed in line with different needs in the future. In the studies in the literature, limited ready functions of the bytecode API are designed similarly to these operations.
2. Each object in a java class is saved in JSON format using the NoSql database collections so that the values of the variables and input-output parameters can be reused in the unit tests to be created or to assign similar random values to these values. These data are also stored in an archive file in the system. In the studies in the literature, different data storage environments such as XML and oracle have been used.
3. While creating test scenarios, some rules should be observed. The first of these is the alternative cases such as condition and loop structures used in the method. The other is the stage where the mock-stub distinction should be made for different objects used in the method and their value transformations. The user is asked to select a test case through the desktop application developed for these. This selection automatically creates unit tests in the framework structure developed according to the desired scenario. At the same time, the opcode

parsing method developed for mock-stub separation leads to automated unit testing generation software.

4. The assertion structure required for each unit test to be created in JUnit standards is prepared using the FTL template engine. While the application automatically prepares unit tests at run-time, annotations are created according to the test case chosen by the user interface. The use of the FTL template engine has not been found in any automated unit testing generation products in the literature. Instead, different methods have been developed.

RELATED WORKS

Among the different studies in the literatüre published in the last 20 years on unit test generation in software testing applications, the main ones are examined.

Csallner et al. developed an automated test generation tool called JCrasher in their work. This tool can work integrated with Eclipse IDE. After examining the information of the sample java class given to the vehicle for test generation, it is tested with random data. They performed the test with JUnit. The tool can also detect errors that occur during the testing phase [4].

Pacheco et al. developed an automated unit test generation tool called Randoop using JUnit. This tool can create feedback-driven unit testing for object-oriented programs. At the same time, it can catch and archive the errors that occur. It is a tool developed to generate random test data and combine test results [5,6].

Simons et al. developed a unit testing tool for java called JWalk. The study consists of two main stages. First of all, advanced features of a sample class are revealed, and then unit tests are applied systematically. As a result, it can provide information about the status of java test classes. It has also been compared to specialist unit testing applications such as JUnit [7].

Sen developed an application named Cute in the java environment in his study. The application was developed by targeting the C programming language for testing the codes written in the c language. This application works by combining automatic and random test logic. It uses executing symbolic code that helps to overcome distinctive input and restrictive solutions. There are cases where this tool needs to be improved, such as the inability to analyze system calls and solve nonlinear integer equations [8].

Charreteur et al. obtained automated test input for Java bytecode programs using a constraint-based reasoning approach. The method has been developed as a constraint model that allows the bytecode program to be searched backward for each bytecode and solve complex constraints on memory shape. This study, which they named JAUT, is a precedent for studies such as Cute, JTEST and PEX [9].

Fraser et al. developed a test generation tool called EvoSuite in their work. Written in Java, this test generation tool has extensive features. All tests performed with this tool are compared against the desired criteria. Analysis and optimization processes are performed as a result of the comparison [10].

Sakti et al. developed an automated test generation tool called JTExpert that can be used in java programming. The JTExpert tool, which is an executable jar file, takes a java file or java project directory as input and automatically generates a test data package in JUnit format for each tested java class [11].

Tanno et al. developed a hybrid unit test tool called CATG in their work. They used a concept called the concholic test, which dynamically performs the symbolic and concrete inputs [12].

Brill et al. developed an open-source tool called TACKLETEST to create test scenarios at the automatic unit level for java applications. It was developed in the context of application modernization at IBM, but is also used as a general-purpose test creation tool. Overall, it implements a new and complementary way of calculating coverage targets for unit testing through a new white-box combinatorial testing application. This tool establishes a new combinatorial test-based approach for computational scope targets that extensively implement different combinations of parameter types of methods tested at configurable interaction levels [13].

Higo et al. developed a dataset creation tool using automated test creation techniques. They predict that there is a large amount of source code with different implementations of the same functions and that these can be compiled into a dataset useful for various research in software engineering. However, they generate a dataset of functionally equivalent java methods from a source code of about 36 million lines [14].

Lukasczyk et al. developed an automated unit test creation software for the Python programming language named Pynguin. While many researchers focus on static software programming languages such as Java, researchers have focused on Python, the dynamic programming language that has become popular in the last decade. They have developed an extensible test generation framework for Python that generates regression tests with high code coverage. This tool can be expanded to allow researchers to tailor it to their needs and enable future research [15].

Bardin et al. developed an integrated framework for automated test generation in their 2021 study. The aim of the study was to adapt DSE (Dynamic Symbolic Execution), an ATG (Automatic Test Input Generation) technique, to effectively cover a wide test target class derived from the source code of the tested program [16]. In addition, in this study, they did not use model-based testing techniques that cover the characteristics of the tested code. Only certain sequential programs are considered here, and no study related to Bytecode is encountered. At the same time, they have developed a useful framework for test automation that can create tag coverage in the relevant programs.

Arcuri proposed a white-box testing approach where the tested code is fully accessible by the developers. Test

cases are automatically generated using an evolutionary algorithm, such as the MIO algorithm. Tests are rewarded based on code coverage and fault detection measures using this optimization algorithm. The developed approach is used in conjunction with the open-source tool EvoMaster. Experiments were conducted on five different web services containing more than 22,000 lines of code. The applied technique was able to generate test cases that detected 80 faults in the web services. However, the coverage obtained was lower compared to the existing test cases in these projects. A manual analysis of the results found that interactions with SQL databases currently prevented achieving higher coverage [17].

MATERIALS AND METHODS

Logic of Automated Unit Testing

Software testing is an essential element for the development of a software product. Software tests, each test progresses with the right results, reaching a whole so that the software can work in a performance way. Examples are the selection of test data, the test's prerequisites, and the intended and expected results.

The logic of software test automation is to automate handwritten tests and turn it into a tool. Automation means; is the constant repetition of test scenarios. In this case, the personnel who prepare the test scenarios need automatically running software that can run the codes, and these software are called software test automation. In software test automation, all kinds of testable algorithms, methods, and classes have these algorithms. Test modules are written for testing. This is called unit testing.

Test scenarios used in unit tests are first developed and the firmware is coded according to the results of these scenarios. Its purpose is to search for truths and faults in test results. All bugs found by test code developers can be fixed at run-time if they can be fixed. If not, they can help the relevant units by directing the reports of the test results. The test of each unit that is accepted as correct is continued by writing the code of the firmware, so that after each module is completed, each module is integrated and the product is completed.

In this study, a framework structure of automatic software test automation, in which unit tests will be produced, is designed. During the design phase of this structure, structures such as java bytecode and javassist were used.

Java Bytecode

Just like C and C++ compilers are represented by the assembler, java programs are represented by byte code. The byte to be generated by a java compiler is actually the program itself. Bytecode is required to be a solution to Java's problems such as portability and security. Since the Java compiler's output is not executable, it has to use bytecode. Bytecodes are interpreted by the JVM. In this way, bytecodes are well-optimized at run-time. Through bytecode, a java program can be run in many different environments.

The bytecode stream of a method is a sequence of instructions for the JVM. Each instruction consists of a one-byte opcode followed by zero or more operands. The opcode indicates the action to be performed. If more information is required before the JVM can take action, this information is encoded into one or more operands that immediately follow the opcode. Each opcode type has a mnemonic.

The bytecode instruction set is designed to be complex. All instructions are aligned with byte boundaries except for two codes for table creation. The total number of opcodes is small enough, so that byte codes only take up one byte. This helps minimize the size of class files before they are run by the JVM. It also helps to keep the size of the JVM application small. All computations in the JVM are performed on the stack. Therefore, bytecode instructions run primarily on the stack [18].

The working logic of Bytecode; Java bytecode is machine code in .class file format. Bytecode in Java is the command set for the JVM and works similarly to a compiler. A close examination of the bytecode reveals that there are certain opcodes. Some opcodes have letters like a or i in front of them. For example, aload_0 and iload_2. These prefixes represent the types that the opcode has worked with. The prefix a means that the opcode modifies an object reference. The prefix i means that the opcode is processing an integer. Other opcodes; They are used as b for byte, c for char, and d for double. These prefixes provide information about what type of data is being processed.

A stack-based machine is used for the execution of the bytecode by the JVM. Each thread has a JVM stack that stores its data in its frame and turns it into a framebuffer. Each time a method is called, a frame stack is created and the operand stack contains data such as a set of local variables and the run-time of the current class. The local variables array, also known as the local variables table, contains the method's parameters and is also used to hold the values of local variables. Parameters are stored in the directory, starting with index 0. If the structure is for a constructor or dynamic (instance) method, the reference is stored at position 0. Then position 1 gets the first formal parameter and position 2 gets the second formal parameter. For a static method, the first formal method parameter is stored at position 0, and the second at position 1. The size of the array of local variables is determined at compile time, and the number and size of local variables depend on a formal procedure parameter. The operand stack uses the LIFO (Last in First out) method stack to push and pull values. Certain opcode instructions values are passed to the operand stack: others take the operands from the stack, manipulate it, and pass the result. The operand stack is also used to get return values from procedures.

A Java agent is a special java library that can manipulate bytecodes by interfering with applications running on the

JVM using the Java Instrumentation API. Generally, it is prepared as a jar file. Classes that represent Java agents are nothing more than other classes available in the Java API library. But what makes them special is that they follow a certain rule that allows the java code to block any other application running in the JVM. The sole purpose here is simply to make agents that probe or modify the bytecode, This powerful feature goes beyond what a java program normally does. In a way, it can be entered into a program and alter the bytecode or cause havoc. Javassist, a library for editing bytecodes in Java, allows Java programs to define a new class at run-time and modify a class file when loaded into the JVM.

The Opcode Parsing Method

In order to create unit tests automatically, a java class is first defined in the framework. The defined java class is first converted to bytecode file format with the help of java agent. Then used in this format; Using many opcode java string functions such as class name, variables, objects, methods and their input and output parameters, each of them is determined line by line before the bytecode file and then word by word. For example, opcodes such as invokevirtual, invokestatic, ifge, iflt, ifeq, iinc are controlled by using java string functions in the bytecode file and the corresponding parameters are determined and passed to variables. At the same time, Mock-Stub distinction can be made by controlling the parameters given by the opcodes showing the object types defined in the class. All parameters transferred to variables are listed instantly in JSON format and saved as data using NoSql database structure. The recorded data is automatically converted into unit test format instantly with the help of the FTL template engine after all the conversion works are completed.

Another feature that is noted here is that in this advanced method, unit tests are created simultaneously with different test scenarios for mock-stub structures, which are distinguished according to the difference of objects, and saved to the system. The test codes saved in the system can be requested as a printout according to the test scenario the tester selected from the test case section of the menus.

Thanks to the opcode parsing method produced as a flexible framework, necessary code changes can be made for the desired opcode type. It contributes to this project as it is completely open source code.

Advantages and Disadvantages of the Opcode Parsing Method

Previous studies in the literature used meta-heuristic optimization search techniques such as genetic algorithms to automate a testing task [19]. However, no optimization search method or genetic algorithm method was used in this study. Directly, the opcodes of a java class in which bytecodes are generated are parsed using java string methods.

In addition to the limited and ready-made functions offered by the Bytecode API, and the developed opcode decomposition method, the input-output parameters corresponding to the desired opcode can be accessed in the bytecode-converted output. At the same time, since this developed method is open-source code, the users can add or remove desired features according to their needs within each function.

As long as other classes belonging to an object produced in a class are in the same location, it can be automatically processed with the base class. Especially if it is considered for mock-stub applications, no extra action is required by the user and the developed framework detects this itself.

As used in the study of Venkatesan et al. in the literature, a flexible data storage system was needed in this study to store all kinds of data obtained from the opcode parsing method [20]. Since the output parameters are recorded in JSON format after the opcode parsing method, these data can be easily examined in any NoSQL-based database management system.

A total of 5 types of test scenarios were used in the developed project. In the future, test users can increase or improve the number of these test scenarios in line with their needs, with the project being open source.

The application developed for now can perform detailed operations on a given java class. With this method developed in subsequent studies, it is planned to create automatic unit tests according to the features selected by the user of the whole project after the location of a project belonging to many classes that are connected with each other.

In the system developed as test data, random data belonging to numeric data types such as int and double are assigned. In the next study, data selection is planned automatically over random data covering all data types or even a sample data set that can be imported.

About Other Technologies Used

Apart from java byte code, javassist and java agent technologies, other technologies and structures have been used to create automatic unit test generation software. These; code coverage, mock and stub, NoSql, Maven and FTL.

- Code Coverage is a software measurement technique used to measure how many lines of code are executed during automated tests. A code coverage reports generator for Java projects, JaCoCo is a free code coverage library for Java.
- In non-relational databases, there is no integrity among the data and therefore data can be repeated in different ways. This leads to data inconsistency. Changing the same data in different places on the whole system is difficult to manage. Since the SQL interface is not used, it is named No Relation, meaning non-relational. Thus, the expression NoSQL came to the fore. MongoDB was used for this system in this study.
- Developed by Apache, Maven is a JDT (Java Development Tool) or java developer tool. While developing Maven java projects, it creates a standard within

the project. In line with these standards, it simplifies the project development process and enables the creation of documentation effectively. It is a tool that helps to eliminate the dependency on the library and user interface in the project, and provides convenience for the developer in processes such as compilation and reporting. In fact, it is itself a storage unit rather than a tool. There is a chance to run all libraries, plugins and all necessary information about the developed software on the servers.

- FreeMarker Java Template Engine-FTL (Java Template Engine) is a template engine produced by Apache. It is a Java library that can generate text output such as HTML web pages, emails, configuration files and source codes based on templates and changing data. Generally, a general purpose programming language such as java is used to prepare the data. Then it displays the data prepared using FTL templates. It focuses on "how" the data will be presented in the templates and "what" data will be presented outside the template [21].
- The concept of dependency emerges when the software is considered as a whole.For example, the software may be running dependent on a database. This dependency should be paid attention to when creating the test scenario and it should be tested without using this dependency. At this point, mock objects are needed. When writing unit tests, pseudo-objects are created to replace them in order to be able to work independently of real objects. This event is called Mock-Mocking.

Automatic Unit Test Generator Software by Collecting Run-Time Data

In the study, a framework was developed that enables the creation of automated unit tests by collecting data in run-time. The tool in question was developed as a console and desktop application using the Java programming language in the Netbeans environment. Since this study is built on the java programming language, JUnit framework was used to test java-based codes and development was made in this context.

The flow chart of the developed study is given in Figure 1. Developed framework; It consists of 4 components. These,

- The first module is the screen where the .java and .class files, which contain the codes of the Java programming language, are read,
- The second module is the screen where all data consisting of variables, object, methods and class names read simultaneously from the codes is saved to the database with NoSQL,
- The third module is the screen where the java codes are converted to byte code,
- And lastly, it is the screen about creating the automatically generated unit test class file with the help of byte code and incoming data.

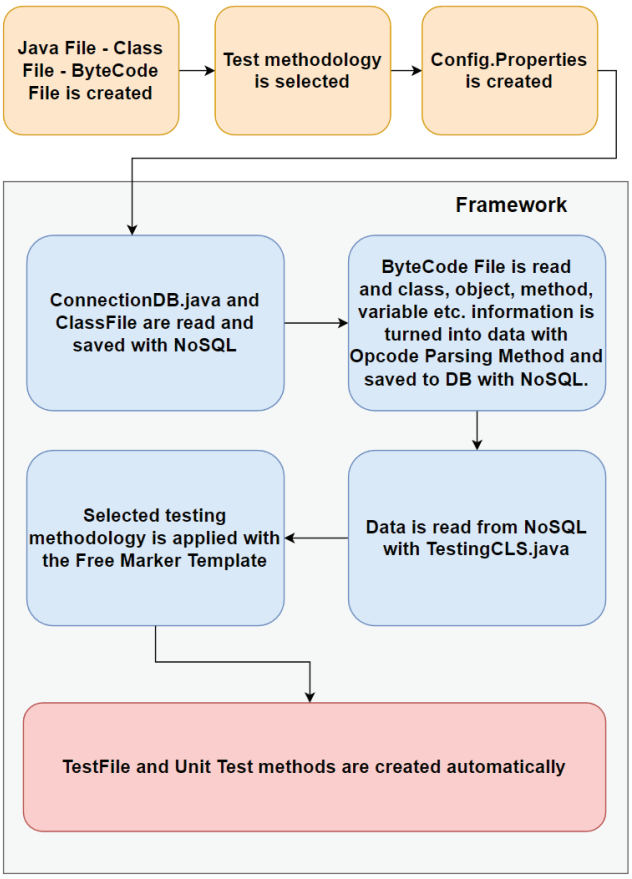


Figure 1. Flowchart of automatic unit test generation software by collecting run-time data.

Identification and Analysis of Scenarios Related to Requirements in Unit Tests

In the study, many scenarios were focused on while creating unit tests. Probability test methods have been created especially for testing methods that contain conditional or loop structures. At the same time, mock and stub structures are also used as needed. A framework has been developed to create automatic unit tests to perform the test stages of all these structures. In this framework, to analyze automatically created test methods and classes, the data used in the software were taken as an example: at the same time, random data were produced and used as a variable. In the sub-headings, each test scenario is examined with code blocks with related simple java class examples.

Alternative States within the Method

When writing a unit test, a single unit test case should be created where a class's method is lean. However, scenarios, where there is more than one situation in the method are very common. In particular, when conditional structures are involved, as in Figure 2, there is a need for a unit test to meet the probability of each condition, and sub-scenarios must be developed for them.

```java
package edu.sdu.sample.project;

public class CalculateCredit {
    public static double CalculateInterest(double principal, double rate){
        if (rate<1.0)
          rate = 1.0;
        return principal/100*rate;
    }
}
```

Figure 2. A method structure with a condition in it.

```java
package edu.sdu.ornek.proje.test;

import edu.sdu.ornek.proje.CalculateCredit;
import org.junit.Test;
import static org.junit.Assert.*;

public class CalculateCreditTest {

    CalculateCredit calCredit = new CalculateCredit();

    @Test
    public void CalculateInterestTest() {
        double calculatedInterest = calCredit.CalculateInterest(100.0,1.20);
        assertEquals(1.20, calculatedInterest, 0.001);
    }

    @Test
    public void CalculateInterestExTest(){
    double calculatedInterest = calCredit.CalculateInterest(100.0, 0.80);
    assertEquals(1.00, calculatedInterest, 0.001);
    }
}
```

Figure 3. Method structures for unit tests of a method with a condition in it.

In this case, as in Figure 3, since there will not be a single unit test that will meet each condition, it is necessary to creating a separate unit test for each condition. Managing such conditional states in a single test method is an anti-pattern.

The reason is that when the test code of the first condition fails, full efficiency cannot be obtained from the test result when the test code of the other conditions is run. In this case, independent tests are expected. As a solution to this situation, different scenarios within the methods are met with separate unit tests.

Different Objects in Method

When writing a unit test, methods can contain different objects. There are sub-methods and objects that these objects in the method depend on. For these sub-methods, the object-generated class is expected to exist ready-made. In summary, the method to be tested depends on the sub-methods of the object. When a normal test code of the code given in Figure 4 is written, a *NullPointerException error* is received.

```java
package edu.sdu.ornek.proje;

public class CalculateCredit {

    public RecordCalculation recCalculation;

    public static double CalculateInterest(double principal, double rate){
        if (rate<1.0)
          rate = 1.0;
        recCalculation.saveCalculation(principal, rate);
        return principal/100*rate;
    }
}
```

Figure 4. A method structure with different objects in it.

```java
package edu.sdu.ornek.proje.test;

import edu.sdu.ornek.proje.RecordCalculation;
import edu.sdu.ornek.proje.CalculateCredit;
import org.junit.Test;
import static org.junit.Assert.*;

public class CalculateCreditTest {

    CalculateCredit calCredit = new CalculateCredit();

    @Test
    public void CalculateInterestTest(){
        calCredit.recCalculation = new RecordCalculation();

        double calculatedInterest = calCredit.saveCalculation(100.0, 1.20);
        assertEquals(1.20, calculatedInterest, 000.1);
    }
}
```

Figure 5. Using the Stub method.

The reason for this NullPointerException error is that the object in the method, namely instance, has not been created. This error can be managed by two situations. The first is the Stub method and the other is the Mock method.

Stub method : It provides ready-made responses to calls made during testing. It usually doesn't respond to anything other than situations written for testing. A class or object structure that implements the methods of the class/object to be called and always returns the desired value. In Figure 5, *calCredit.recCalculation = new RecordCalculation();* line has been added.

This is an option that is not easy to manage. Because there are sub-components of the class that need to be called and an object structure that cannot be easily created. Besides all these, there may be content dependent components such as Spring and EJB. Therefore, due to all these disadvantages, the stub method, which enables to create a counterpart of these objects, is not preferred as much as possible.

Mock method : It is used to ensure everything in the method is correct before returning the correct value. It just tests the behavior and makes sure certain methods are called. Like the stub method, the mock method also acts as a mock. This method has a structure that allows an object to serve as if it were operating normally without its actual existence. That is, empty methods and objects replace real methods and objects. However, the difference is; stub focuses on a testable version of a particular object, while mock focuses on the correctness of everything.

Mockito library is used together with JUnit for mocking. The downside is that there is no integration with the

```java
package edu.sdu.ornek.proje.test;

import edu.sdu.ornek.proje.RecordCalculation;
import edu.sdu.ornek.proje.CalculateCredit;

import org.junit.Test;
import org.mockito.Mockito;
import static org.junit.Assert.*;

public class CalculateCreditTest {

    CalculateCredit calCredit = new CalculateCredit();

    @Test
    public void CalculateInterestTest() {
        calCredit.recCalculation = Mockito.mock(RecordCalculation.class);
        double calculatedInterest = calCredit.CalculateInterest(100.0, 1.20);
        assertEquals(1.20, calculatedInterest,0.001);
    }
}
```

Figure 6. Using the mock method.

respective class. However, unit tests by their nature do not deal with integrations with subunits. Therefore, such tests are expected to be performed within integration tests. In Figure 6, *calCredit.recCalculation = Mockito.mock(RecordCalculation.class);* line is added to mock the object. The existence of a class file belonging to the object is defined as a parameter to the mock method. Accordingly, the content of the save method in that object will no longer be executed.

Annotation-based uses are also available to manage mocking. In Figure 7, the Mockito library shows a method for mocking the object by adding annotations directly to it.

As seen in the figure, if annotations are defined in the test class, the code will become even more readable and concise. Thus, related classes were defined in annotations and an object was produced from them. Afterward, these objects were used in test methods.

Value Conversions on Mocked Objects

Sometimes a real unit test environment cannot be created because methods or objects carry their default values in a mocked object. In this case, AssertionError error is received at the time of test. The meaning of this error indicates that a different result is obtained as a result of comparing the expected and actual values. When the mock method mocks an object, it removes the method bodies inside it. In order to manage this, certain states are assigned to the mocked objects. Given-When-Then standards were used to manage these situations, as seen in Figure 8.

Thus, even if the object is mocked, the desired states can be created and tests can be written. In the Given stage, the variables that are desired to be obtained and to be used are created. The object to be tested is configured. In the When phase, the values and variables are brought together with the configured object in the Given phase, and the code to be tested is processed. In the Then phase, the expected result is passed to the test code.

Design of Automated Unit Test Generation Software

Java-class-bytecode conversions

In the developed framework, first of all, a java class is loaded into the system to create unit tests. This *java file* is converted into a *.class file* with the click of a button, and the .class file is used in almost every part of the developed software. The *JavaCompiler* library is used to compile the class file. If a class that needs more than one java class is read here, class files are read in other classes in the specified location. The developed system stores all the files it reads in the *Iterable* collection list, and then this list is used for conversion operations.

After conversion from Java file to Class file, byte code conversion operations are performed. For this process, *ClassPool, CtClass, CtMethod* and *InstructionPrinter* subclasses in the javassist library are used. The class and method information in the class file being read is kept in the class pool named *classPool*. Afterward, information is drawn from the class pool with *InstructionPrinter* and the results are displayed as a printout.

As seen in Figure 9, in a byte code output, each method in the class is separated between the MethodName lines. Each method starts with line number 0. This method

```
package edu.sdu.ornek.proje.test;

import edu.sdu.ornek.proje.RecordCalculation;
import edu.sdu.ornek.proje.CalculateCredit;

import org.junit.Test;
import org.junit.runner.RunWith;
import org.mockito.InjectMocks;
import org.mockito.Mock;
import org.mockito.runners.MockitoJUnitRunner;
import static org.junit.Assert.assertEquals;

@RunWith(MockitoJUnitRunner.class)
public class CalculateCreditTest {

    @InjectMocks
    CalculateCredit calCredit;
    @Mock
    RecordCalculation recCalculation;

    @Test
    public void CalculateInterestTest() {
        double calculatedInterest = calCredit.CalculateInterest(100.0, 1.20);
        assertEquals(1.20, calculatedInterest,0.001);
    }
}
```

Figure 7. Using the mock method with annotation.

```java
package edu.sdu.ornek.proje.test;

import edu.sdu.ornek.proje.InterestRate;
import edu.sdu.ornek.proje.CalculateCredit;

import org.junit.Test;
import org.junit.runner.RunWith;
import org.mockito.InjectMocks;
import org.mockito.Mock;
import org.mockito.runners.MockitoJUnitRunner;
import static org.junit.Assert.assertEquals;
import static org.mockito.Mockito.*;

@RunWith(MockitoJUnitRunner.class)
public class CalculateCreditTest {

    @InjectMocks
    CalculateCredit calCredit;
    @Mock
    InterestRate intRate;

    @Test
    public void CalculateInterestTest() {
        when(intRate.monthlyRate()).thenReturn(1.20);
        double calculatedInterest = calCredit.CalculateInterest(100.0);
        assertEquals(1.20, calculatedInterest,0.001);
    }
}
```

Figure 8. Using the When-Then Standard when mocking.

```
MethodName : InterestRate
0: dload_0
1: ldc2_w #2 = int 100.0
4: ddiv
5: dload_2
6: dmul
7: dreturn
MethodName : InterestRateEx
0: dload_2
1: dconst_1
2: dcmpg
3: ifge 8
6: dconst_1
7: dstore_2
8: dload_0
9: ldc2_w #2 = int 100.0
12: ddiv
13: dload_2
14: dmul
15: dreturn
```

Figure 9. An example of Bytecode.

returns a value with the *dreturn* keyword, and it also shows that it takes parameters with the keyword *lcd2_w*.

These byte codes are read by the Javassist library and keep the filename and path of the relevant class in parameters such as *ClassPool*, *CtClass* and *CtMethod*. All code lines converted to byte code can be seen on the console screen of the *Netbeans* interface, as well as output as text documents. Another class developed for this framework is *FindObjectsInClasses*. With the help of this class, each byte code is read line by line from the output of the byte code file and these codes are stored in the MongoDB database as NoSQL format. For this, a class called *ConnectionDB* has been created in the framework. In this class, features such as creating a collection and accessing the data in the collection using the methods of the *com.mongodb* library are gathered in a common class. This class also provides archiving of all JSON format data in a text document

The important point here is that the help of java string functions is taken to retrieve the information and save it to the database. This is done by *the opcode parsing method* developed for the study. Variables used in the class, method names, if any, the parameters they have received and the results they have sent, all created objects, Information about used condition and loop blocks are retrieved with the help of string functions, respectively. and saved in MongoDB. In addition to these, information about the opcodes used such as *invokevirtual, getfield, invokestatic, getstatic, ifge, ifle, iflt, ifgt, ifeq, ifne, iinc, if_icmp* and *goto* are also recorded. At the same time, all this information is archived under system files in text document (.txt) format according to the day and time of the transaction. Thus, two collections are used by MongoDB. The first of these is the collection named *byteCoding* seen in Figure 10. Under NoSQL structure, data is listed with key-value structure. The naming of key fields is related to the operations to be performed in some, while in others it reminds of byte code terms. For example, the

```
/* 1 */
{
    "_id" : ObjectId("622e1bb20fb2df14e4b6d6ce"),
    "Execution Id" : "'25'",
    "MethodName" : "'InterestRate'",
    "ClassName" : "'CalculateCredit'",
    "Returned" : "'728.28'"
}

/* 2 */
{
    "_id" : ObjectId("622e1bb20fb2df14e4b6d6cf"),
    "Execution Id" : "'26'",
    "MethodName" : "'InterestRateEx'",
    "ifge12Line" : "< 1.0",
    "ClassName" : "'CalculateCredit'",
    "Returned" : "'1936.69'"
}

/* 3 */
{
    "_id" : ObjectId("622e1bb20fb2df14e4b6d6d0"),
    "Execution Id" : "'27'",
    "MethodName" : "'InterestRateFor'",
    "Loop" : "InterestRateFor",
    "ClassName" : "'CalculateCredit'",
    "Returned" : "'475800'"
}

/* 4 */
{
    "_id" : ObjectId("622e1bb20fb2df14e4b6d6d1"),
    "Execution Id" : "'28'",
    "MethodName" : "'InterestRateIF'",
    "ifle41Line" : "> 7.0",
    "ifge45Line" : "< 9.0",
    "iflt52Line" : ">= 3.0",
    "ifgt56Line" : "<= 6.0",
    "ifle63Line" : "> 1.6",
    "ifgt67Line" : "<= 2.95",
    "ClassName" : "'CalculateCredit'",
    "Returned" : "'4.56'"
}
```


Figure 10. Structure of the collection named *byteCoding*.

key name of *MethodReturnType* stores the value that that method sends as a result, while the key name *ifge45Line* stores the value of the operation to which the if block corresponds.

The second collection name used in MongoDB is *kayitlar*. As seen in Figure 11, all information about classes and methods has been recorded. So much so, that the class and object names of the objects to be mocked are included in this list under the mocking keyword.

Data Migration – Template Extraction

All data archived on MongoDB must be able to communicate with automated unit testing software for operations such as data reading or writing. For this reason, a POJO class has been written to enable communication between the database and the software. All constructor and getter-setter methods suitable for the scenarios are defined together with their variables. The POJO class acts as a data carrier within the framework. The data is stored in two tables in the database. These tables were combined with ids using join and they were communicated with each other. Two different classes are written for the data to be read from the database with these join operations.

The first of these is the *GetItFromMongoDB* class developed for software. The codes of the conditions and loops read in run-time were kept in the *byteCoding* collection. The data read in this collection is directed to *FTL* format to be converted to unit test code with the help of pojo class. Variables for keywords that start with *if* in the *byteCoding* collection; the class name is sent to the pojo class along with the method name and return values. Then, these values are written to the template file named *writeIfTests.ftl* as a unit test method.

The *ReadDataFromDB* class has been created, which pulls data from the MongoDB database. In this class, two collections are connected using join. All relevant variables are both read from the database and checked if there are fields that need to be updated for join operations. All variables are redirected to the framework's main class. At the same time, unit test methods are created by copying the template variables selected for the *FTL* file.

All data pulled from the database is fetched with the bearer class POJO. Together with these data stored in the database, unit tests are automatically prepared in the desired file format using FTL and turned into an output. The content of these outputs are codes belonging to a new class created based on run-time collected data.

The data produced from the parameters used in the java class read to the frame are saved in MongoDB and the data in the database is pulled with FTL


```
/* 1 */
{
    "_id" : ObjectId("622e16570fb2df4654f26bbb"),
    "Execution Id" : "'1:1'",
    "Kaynak" : "'public static double org.brutusin.instrumentation.logging.CalculateCredit3.CalculateInterest(double)'",
    "Baslangic Suresi" : "'Sun Mar 13 19:05:43 EET 2022'",
    "Degiskenler" : "'[484.0]'",
    "ClassName" : "'CalculateCredit3'",
    "MethodName" : "'CalculateInterest'",
    "MethodReturnType" : "'double'",
    "MethodParameters" : "'[double arg0]'",
    "ObjectInClass_Field" : "intRate",
    "ObjectClassName_Method" : "InterestRate",
    "Mocking" : "InterestRate.monthlyRate",
    "ImportMethod" : "org.brutusin.instrumentation.logging.InterestRate",
    "ImportField" : "org.brutusin.instrumentation.logging.CalculateCredit3",
    "Toplam Suresi" : "'277 ms'",
    "Returned" : "'5.808'"
}
```


Figure 11. Structure of the collection named *kayitlar*.

templates. A class that generates random data named *GenerateRandomJavaDataType* has been written in the framework and this random data is used where it is needed in the application. Java string functions are also used in this class, and with the help of these functions, it is determined which data type will be generated randomly for the relevant test class. All generated data is kept in java collection lists and directed for sending to the relevant test class.

RESULTS AND DISCUSSION

Implementation of Automated Unit Test Generation Software

After the back-end codes of the application were completed, the front-end codes were prepared and the unit test code was produced to work in run-time on the scenarios specified in the requirements section.

The application was developed not only as a console environment, but also as a desktop platform. For this, a class called *UnitTestGeneratorGUI* has been designed. The screenshot given in Figure 12 is designed with this class. The class uses the necessary hierarchy for the relevant parts of the framework to work; it contains the functions of all command buttons, lists and text boxes.

Many methods and class types were studied on the developed framework, scenarios created for each alternative situation were tried, and automatic unit test codes were created from all of them, as planned.

Various trial tests and performance analyzes were carried out within the scope of the study. Framework doesn't just work on variables, objects and methods. In addition, additional features such as conditional constructs and loops are also included. With these additional features, studies were carried out with scenarios prepared on the framework, and from all of them a test class file containing them and automated unit test codes that may be suitable as planned.

As seen in the screenshot in Figure 12, the interface consists of two main parts. First, the java code for which the test file is to be prepared is selected with the *Choose* button. Then, in the second part, the data to be sent to the framework is prepared by using the tab object on the screen. The content of the selected java class comes to the *Java Code File* window as a dump. Each code within the class codes will be converted to the bytecode, but any code structure in comments or comment lines are not converted by bytecode. This feature is available in all compilers. In the example in the figure, this class has both conditional and loop constructs.

When the *Get Info* button is clicked, the name of the class and the names of the methods it belongs to will be listed on the right side of the screen. For this, a class named *GetInfoAboutJavaAndByteCodeFile* was designed and run in the background. Here, first of all, the java file, which is

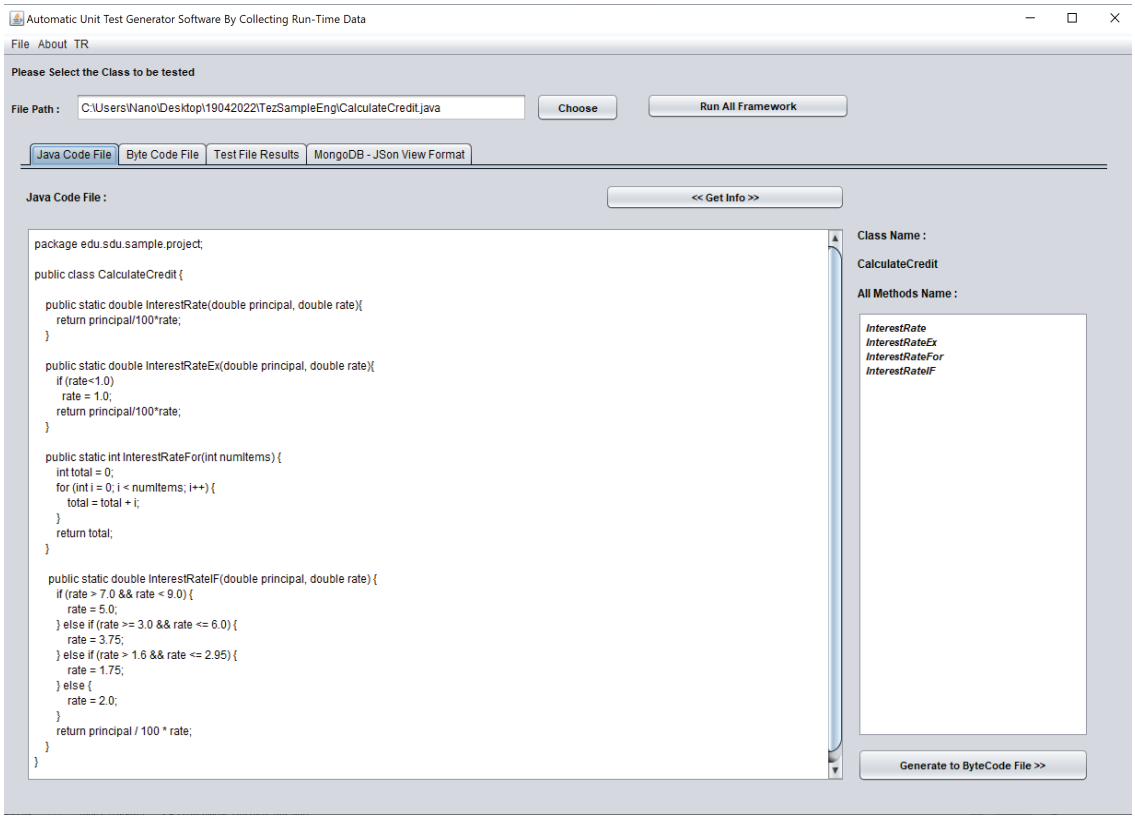


Figure 12. Screenshot of automated unit test generation Framework.

read with the file path, is transformed into a class file by using the *JavaCompiler* library. Along with this conversion, byte code conversion of the relevant class is also performed. After these transformations are completed, the class name of the relevant class and the names of its methods are collected with an array variable and this information is transferred to the list on the right side of the screen. When the *Generate Bytecode File* button is clicked, the *Bytecode File* tab shown in Figure 13 will open.

With the help of the Javassist library, the class file of the relevant example class was created and converted to byte code. Byte codes are created in accordance with the standard determined by the javassist library, separated by method names. The required values from this byte code will be directed into the unit test classes and methods created as a template, and unit tests will be created automatically at run-time. Which test scenario will be used for this, its test case should be selected from the drop-down list at the top of the screen. In addition to normal test methodologies, there are names of test scenarios developed with mock and stub methods. According to the structure of the class opened with the framework, these test scenarios can be selected and the results can be obtained.

By choosing the test case, it is determined with which test scenario the automatic unit test will be generated. Definitions of these scenarios and sample file names are given in the drop-down list with their explanations. After this selection is made, a configuration file is created, such as the lines displayed under the *Contest of the Config File* title in the figure, by clicking the *Generate to ConfigFile* button. In the created configuration file; the names of the java and class files and their path, chosen test method, the name of the java class and the method names it belongs to, along with the file name of the byte code, there is information such as the creation date and time of this configuration file.

Clicking the *Run All Framework* button will pull all the necessary information from this configuration file sequentially. As seen in Figure 14, the command line screen opens and the framework starts working with the information it reads from the *config.properties* configuration file.

Thanks to the Maven structure, it creates three different test class instances that belong to the test codes of the given class according to the data it pulls from the configuration file. At the same time, data is recorded in the database in run-time. The documents belonging to the test classes that come out are both saved in the system as a file and also collectively reflect the results on the command line screen as seen in the figure.

After the command line screen is closed, the *Export Test Results to Lists* button is clicked. The inventories of these unit test files, which are created and then saved in the system, are presented to the user as a screen output in the third

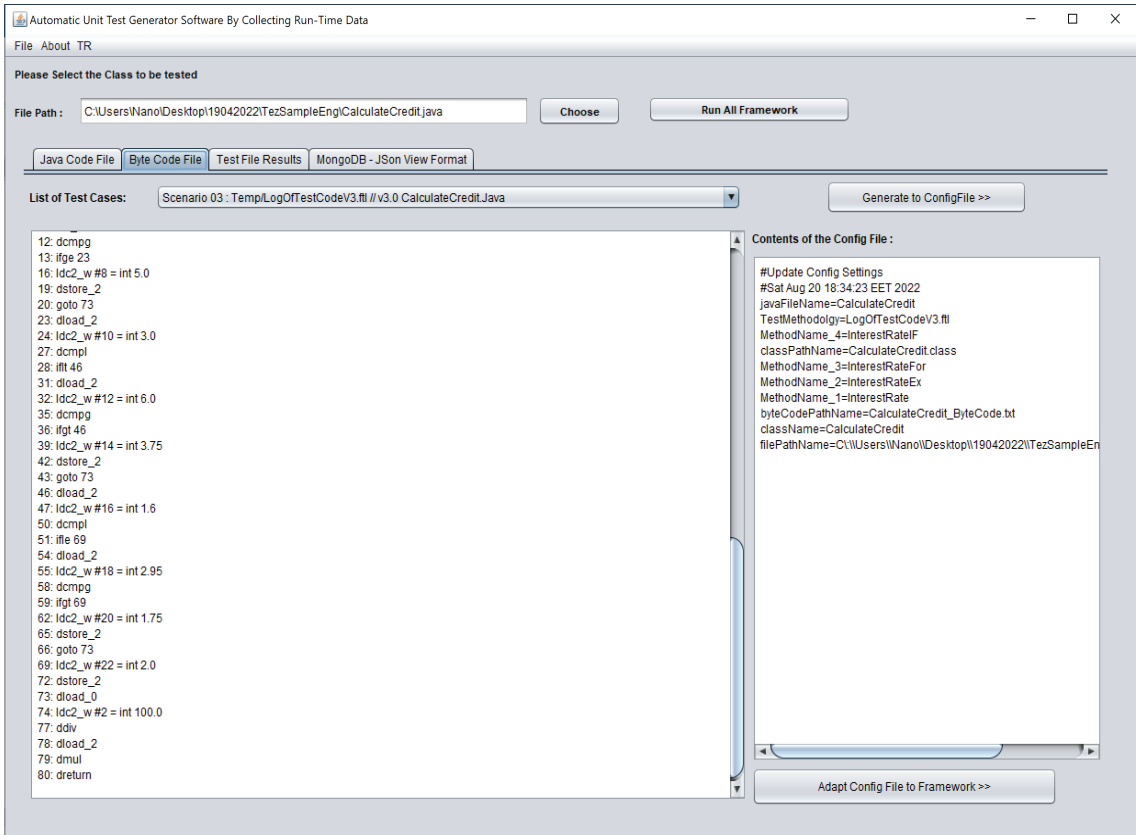


Figure 13. Bytecode conversion of the Java class.

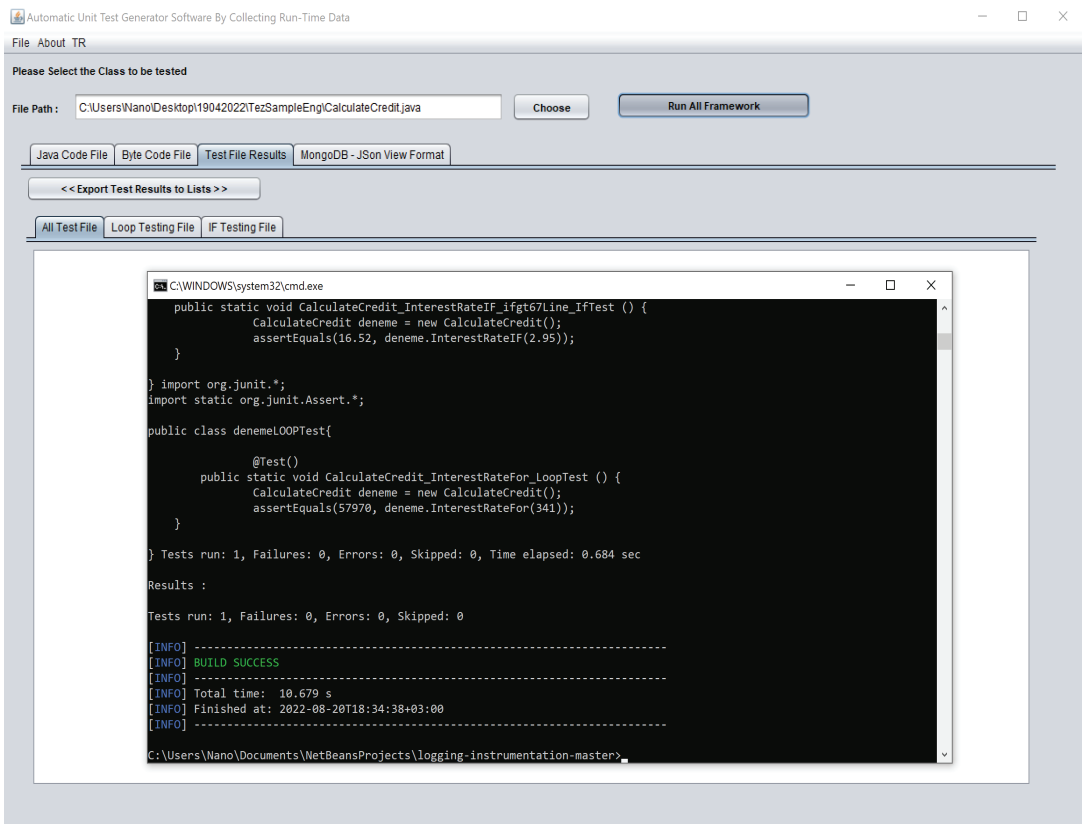


Figure 14. run-time of framework and obtaining test results.

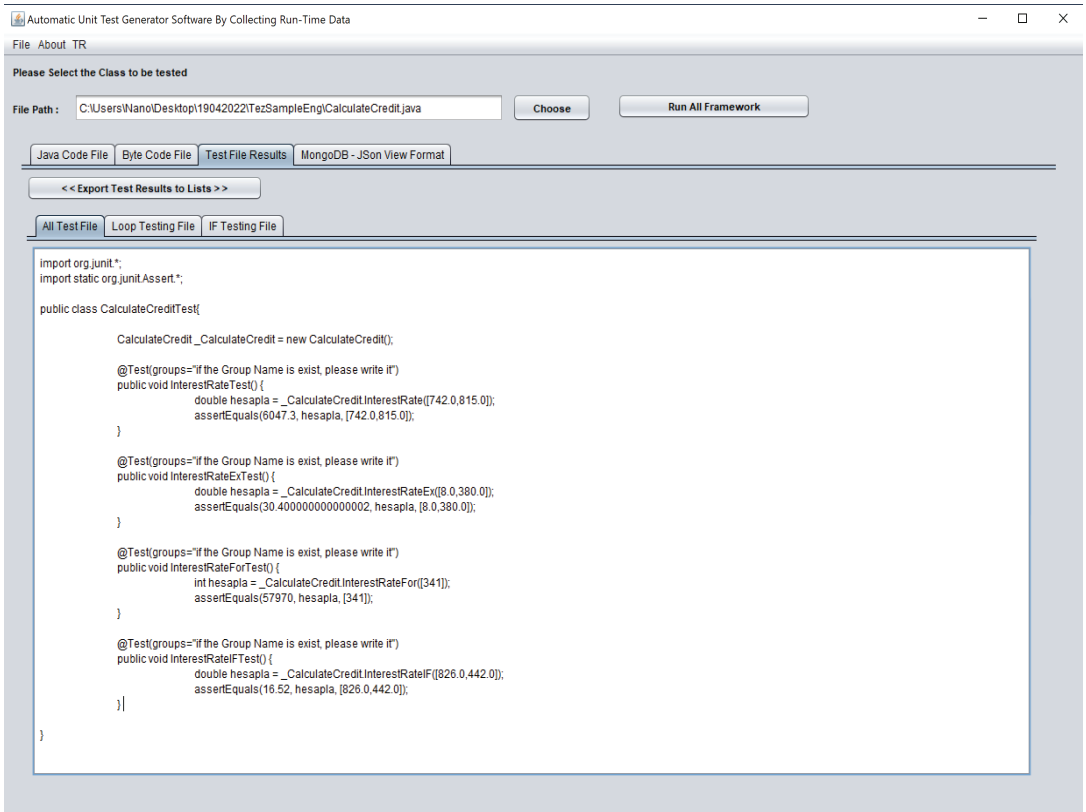


Figure 15. Display of all possible unit test methods.

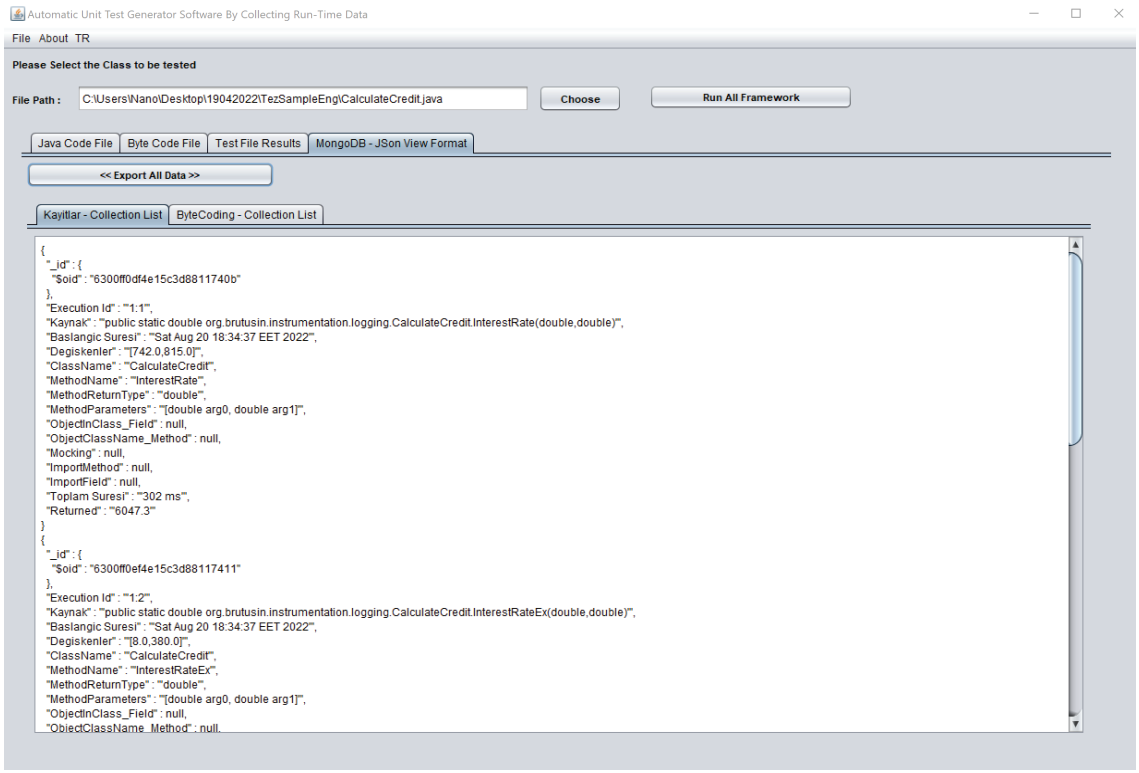


Figure 16. JSon format of *kayitlar* collection in MongoDB.

tab of the application, named Test Results, as seen in Figure 15. There are three sub-tabs on this screen. The first of these tabs is called *All Test File* and this area contains codes for all possible test methods. The second tab contains all possible unit test methods related to loops and the third tab all possible unit test methods related to condition structures.

The fourth tab of the application belongs to the screen named *MongoDB – JSon View Format*. Figure 16 shows two tabs. There are also two NoSQL collection structures in MongoDB software. In the first tab; when the data collected by the framework is logged, class name and method names of the relevant java class, the parameters it takes, values associated with objects, the names belonging to the class names they have, return values from methods, It is the list of the records collection that includes information such as whether there is a mocking or not.

In the second tab; The data of the byteCoding collection can be seen. The analysis of the loop and condition structures of each method in the Java class was made separately, and the summary information of the values related to them was collected and recorded separately.

In Figure 17, a screenshot of the results of the code coverage of each test scenario generated from this software is given.

Realization of Alternative Situations Scenario in Method

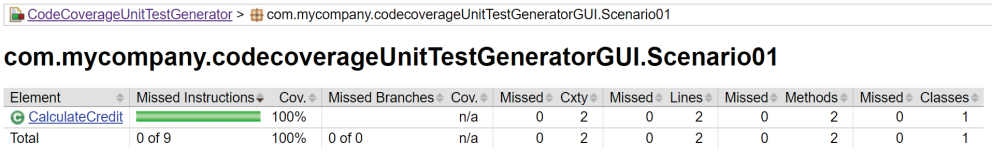
The scenarios given under the title of alternative situations in the method are analyzed under the title of this section. In this case, the code coverage result of the test scenario of a class and a method structure created at a simple level is given in Figure 18. According to this result, it is seen that a successful test is produced with the percentage value of the generated test code.

CodeCoverageUnitTestGenerator — Sessions

CodeCoverageUnitTestGenerator

Element	Missed Instructions	Cov.	Missed Branches	Cov.	Missed	Cxty	Missed	Lines	Missed	Methods	Missed	Classes
com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario06		68%		n/a	2	4	2	4	2	4	1	2
com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario04		88%		50%	2	5	1	8	1	4	0	2
com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario05		88%		50%	2	5	1	8	1	4	0	2
com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario03		100%		50%	1	5	0	8	0	4	0	2
com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario02		100%		50%	1	3	0	4	0	2	0	1
com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario01		100%		n/a	0	2	0	2	0	2	0	1
Total	11 of 121	90%	4 of 8	50%	8	24	4	34	4	20	1	10

Figure 17. Code coverage results of scenarios.


CodeCoverageUnitTestGenerator > com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario01

com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario01

Element	Missed Instructions	Cov.	Missed Branches	Cov.	Missed	Cxty	Missed	Lines	Missed	Methods	Missed	Classes
CalculateCredit		100%		n/a	0	2	0	2	0	2	0	1
Total	0 of 9	100%	0 of 0	n/a	0	2	0	2	0	2	0	1

Figure 18. Code coverage result of Scenario1.

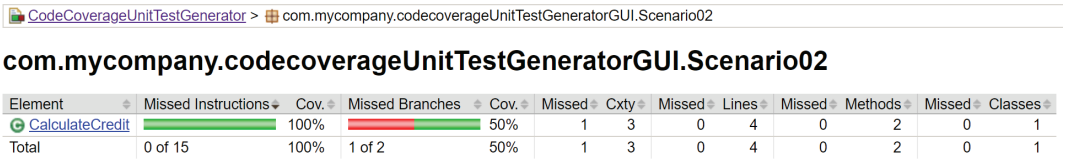
It does not always work with a simple method. There are other alternative method structures in the class. For this reason, thanks to the conditional structure in the method, the test developer resorts to alternative ways while preparing the test scenario. This conditional structure gives the user a kind of guidance on whether the method to be written unit test will work correctly. The result of the code coverage of the scenario of the output given by the automatic unit test creation software is given in Figure 19. According to this result, it is seen that a successful test is produced with the percentage value of the generated test code. Because the condition structure is used in this scenario, the percentage value in the missed branches column has changed.

To summarize this section's analysis results of this section; objects used in a method, loops, and conditional statements are checked and recorded at the time of operation with their relevant data and unit test code generated. However, when there is more than one method in a class, all the values in each method are checked and solutions are developed for each of them.

Realization of Different Objects Scenario in Method

The scenarios given under the title of different objects in the method are analyzed under the title of this section. In this case, different objects are checked within the method as priority. For this, scenarios are realized by creating objects from classes that have interdependencies. For this scenario in the method, a unit test was prepared with two different methods. These methods are realized with the use of mock and stub objects. As can be seen in Figure 20, the result regarding the code coverage of the scenario obtained using the stub method has been obtained.

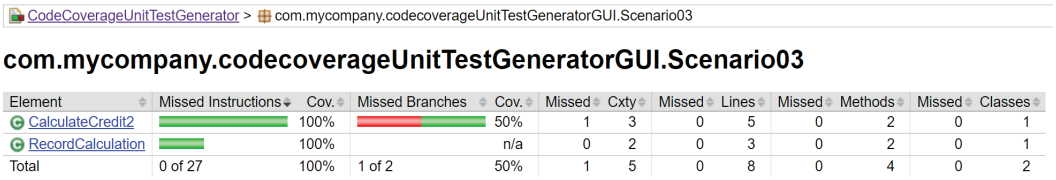
In another scenario, Mockito library is used together with JUnit. With the addition of this library, the object is mocked up. In Figure 21, the code coverage result of the fourth scenario is given. Here, after mocking, both the


CodeCoverageUnitTestGenerator > com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario02

com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario02

Element	Missed Instructions	Cov.	Missed Branches	Cov.	Missed	Cxty	Missed	Lines	Missed	Methods	Missed	Classes
CalculateCredit		100%		50%	1	3	0	4	0	2	0	1
Total	0 of 15	100%	1 of 2	50%	1	3	0	4	0	2	0	1

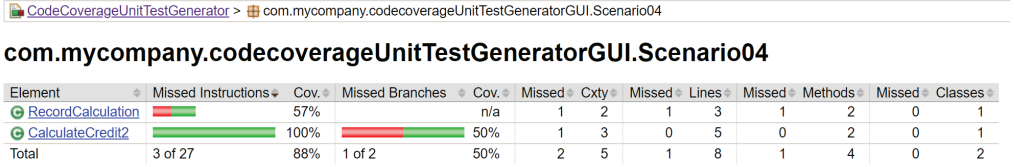
Figure 19. Code coverage result of Scenario2.


CodeCoverageUnitTestGenerator > com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario03

com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario03

Element	Missed Instructions	Cov.	Missed Branches	Cov.	Missed	Cxty	Missed	Lines	Missed	Methods	Missed	Classes
CalculateCredit2		100%		50%	1	3	0	5	0	2	0	1
RecordCalculation		100%		n/a	0	2	0	3	0	2	0	1
Total	0 of 27	100%	1 of 2	50%	1	5	0	8	0	4	0	2

Figure 20. Code coverage result of scenario3.


CodeCoverageUnitTestGenerator > com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario04

com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario04

Element	Missed Instructions	Cov.	Missed Branches	Cov.	Missed	Cxty	Missed	Lines	Missed	Methods	Missed	Classes
RecordCalculation		57%		n/a	1	2	1	3	1	2	0	1
CalculateCredit2		100%		50%	1	3	0	5	0	2	0	1
Total	3 of 27	88%	1 of 2	50%	2	5	1	8	1	4	0	2

Figure 21. Code coverage result of Scenario4.

CodeCoverageUnitTestGenerator > com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario05

com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario05

Element	Missed Instructions	Cov.	Missed Branches	Cov.	Missed	Cxty	Missed	Lines	Missed	Methods	Missed	Classes
RecordCalculation		57%		n/a	1	2	1	3	1	2	0	1
CalculateCredit2		100%		50%	1	3	0	5	0	2	0	1
Total	3 of 27	88%	1 of 2	50%	2	5	1	8	1	4	0	2

Figure 22. Code coverage result of scenario5.

percentage value of *RecordCalculation* and the value of *CalculateCredit2* appear low.

Another use of the mock method is with annotations. Mockito directly marks the object with annotations and performs the mocking process. In Figure 22, the code coverage result of the fifth scenario is given. Here, after mocking, both the percentile value of *RecordCalculation* and the value of *CalculateCredit2* appear low.

Another feature that needs attention in mocked objects is the value transformations of the objects. Methods and objects in the mocked object have their default values. For this reason, baseline values cannot be generated in the unit test environment from time to time. In order to overcome this, *Given-When-Then* standards are followed. Figure 23 shows how this situation was resolved with a test scenario.

In Figure 24, the code coverage result of the sixth scenario is given. Here, the existence of the mocked object and

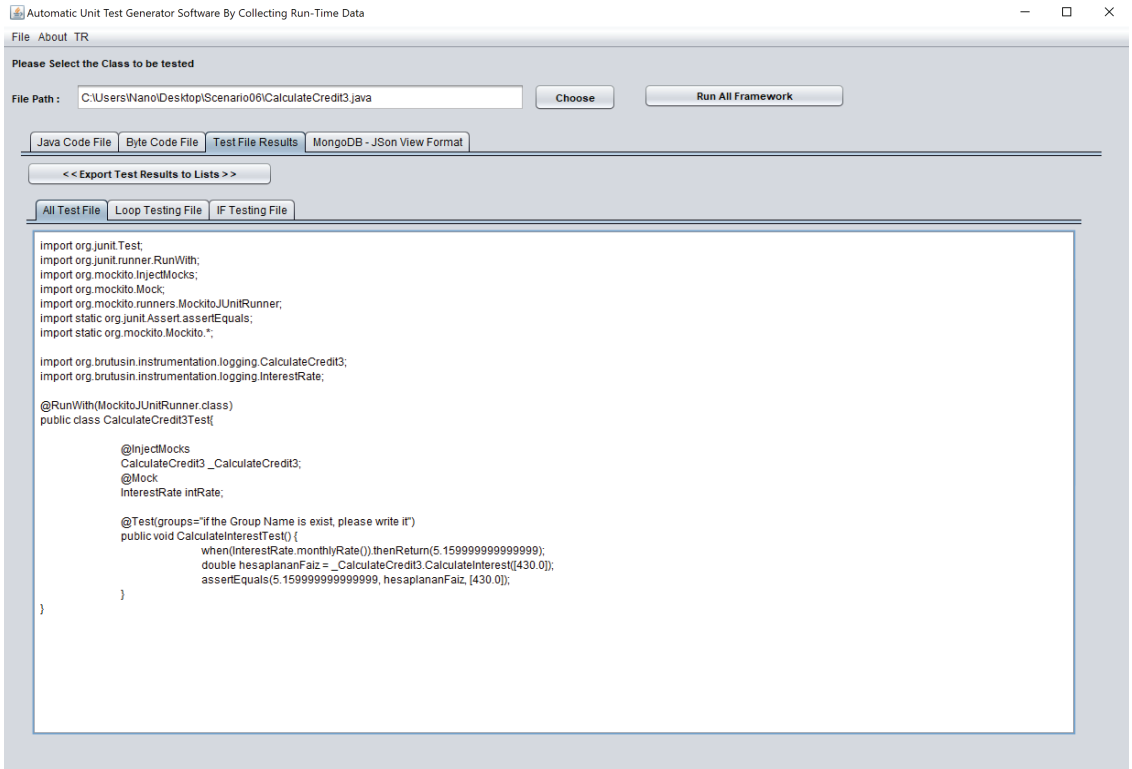


Figure 23. Realization of scenario 6.

CodeCoverageUnitTestGenerator > com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario06

com.mycompany.codecoverageUnitTestGeneratorGUI.Scenario06

Element	Missed Instructions	Cov.	Missed Branches	Cov.	Missed	Cxty	Missed	Lines	Missed	Methods	Missed	Classes
InterestRate		0%		n/a	2	2	2	2	2	2	1	1
CalculateCredit3		100%		n/a	0	2	0	2	0	2	0	1
Total	5 of 16	68%	0 of 0	n/a	2	4	2	4	2	4	1	2

Figure 24. Code coverage result of scenario6.

the *InterestRate* class can be interpreted from the falling percentage rate.

Thus, it is shown how six important scenarios are resolved in automated unit test creation.

Examination of the Developed Framework in Real-World Unit Test Case Studies

Experiments were conducted on a computer with an i7 2.50 GHz CPU and 16.0 GB memory, running on Windows 10 with JDK 11. For the developed framework, real-world unit test case studies are also reviewed. Multiple experiments were conducted for each test scenario given in the study. For example, for each test case, five independent and simple Java classes were selected from the SF110 Corpus of Classes [22].

SF110 classes collection is an open source repository of 100 sample java projects. Within this repository, unit tests of each project are also included. Line spacing of each java class is generally; It is 10 - 100, 100 - 1000 and 1000 - 10000. These experiments included code blocks and class structures that could yield different CPU and memory results, such as a simple test method and a comprehensive test method like mock. In particular, data obtained from bytecode files were transformed using an *opcode parsing method* in this test production framework. Therefore, the number of lines of code (LOC) for this was also added to the table. The results such as CPU, memory, time, number of lines of code in Java class files and number of lines of code in bytecode files, which are approximate value ranges for all examples, are shown in Table 1.

In this study, test scenarios were determined and test cases were classified in a table. Test cases were examined in six different categories: test case of a simple java method (*TC - 01*), test case created for alternative situations in a java method (including complex methods with conditional and/or loop structures) (*TC - 02*), test case created for different objects in a java method using the Stub method (*TC - 03*), test case resolved with the Mockito library (*TC - 04*), test case resolved with the Mock method's annotations (i.e. MockitoJUnitRunner) (*TC - 05*), and test case created for value conversions performed for fake objects using the Given-When-Then standard (*TC - 06*).

According to the results, the conversion time varies with an increase in the number of code lines for bytecode and Java classes. Additionally, there is an increase in difficulty levels of test scenarios from *TC - 01* to *TC - 06*, which also affects the conversion time. Furthermore, the processing of different classes, object creation, mock-stub, loop and conditional structures were considered in the analysis.

The comparison of the automatic unit test creation software and the processes performed with it by collecting the generated run-time data with other widely used automatic unit test creation tools in the literature is discussed in the discussion and conclusion section.

The application of this developed study has been published at *https://github.com/SevdanurGENC/Nano-Automatic-Unit-Test-Generator*.

Table 1. Approximate value ranges obtained from The Opcode Parsing Method when considering Bytecode LOC.

Test Case	Bytecode LOC	Java Class LOC	Time (Sec)	CPU (%)	Memory (Mb)
	5 – 200	*10 – 100*	*10 – 13*		*15 – 18*
TC – 01	*200 – 2100*	*100 – 1000*	*13 – 16*	*15 – 20*	*18 – 21*
	2100 – 22000	*1000 – 10000*	*16 – 20*		*21 – 25*
	5 – 230	*10 – 100*	*12 – 16*		*16 – 21*
TC – 02	*230 – 2250*	*100 – 1000*	*16 – 19*	*17 – 22*	*21 – 26*
	2250 – 23200	*1000 – 10000*	*19 – 22*		*26 – 33*
	5 – 250	*10 – 100*	*13 – 17*		*19 – 25*
TC – 03	*250 – 2340*	*100 – 1000*	*17 – 22*	*17 – 25*	*25 – 33*
	2340 – 24000	*1000 – 10000*	*22 – 25*		*33 – 39*
	5 – 270	*10 – 100*	*15 – 18*		*22 – 31*
TC – 04	*270 – 2400*	*100 – 1000*	*18 – 24*	*20 – 30*	*31 – 39*
	2400 – 27900	*1000 – 10000*	*24 – 29*		*39 – 43*
	5 – 300	*10 – 100*	*16 – 19*		*25 – 35*
TC – 05	*300 – 2800*	*100 – 1000*	*19 – 28*	*23 – 39*	*35 – 49*
	2800 – 30000	*1000 – 10000*	*28 – 37*		*49 – 61*
	5 – 320	*10 – 100*	*17 – 25*		*29 – 41*
TC – 06	*320 – 2950*	*100 – 1000*	*25 – 32*	*23 – 42*	*41 – 55*
	2950 – 31500	*1000 – 10000*	*32 – 41*		*55 - 73*

CONCLUSION

In this study, an application has been developed that collects data at run-time with the help of Java Agent, stores the collected data in the NoSql database and transforms this data into unit test using a JTL template engine. The studies conducted and the developed tool use a different structure compared to previous studies in the literature in various aspects.

A structure has been created to respond to all possible test scenarios mentioned in the study. When the resulting test classes are run in JUnit, results about which class is being tested can be easily obtained. At the same time, it was observed that successful results were obtained with code coverage. Through this framework, which is easy to use for users, unit tests are created automatically in a very short time.

Recent studies show that each test tool prepared for studies is intended to generate test scenarios healthily automatically. While most of the studies were prepared with the Java programming language, programming languages such as C/C++ and C# make up the rest. Many of these are desktop applications, while the rest are developed as web applications or plug-ins. Usually, these test cases are randomly generated to check the running of the programs. For the development of random tests and their derivatives, techniques based on dynamic symbolic execution are mostly preferred.

Pex, one of the important studies in the literature, is a tool developed for unit testing of C# code. It generates test inputs with different parameters for test scenarios by using dynamic symbolic execution techniques. It also produces results based on the return values of the methods. However, it is limited to classes that require complex method arrays. Randoop and EvoSuite can be given as examples of other important studies in the literature made in the Java programming language. Randoop is known for its ease of use, but unlike EvoSuite, it cannot test complex code structures without guidance. It also aims to produce compact test cases with high code coverage. When using code coverage, a common systematic approach is to select a coverage target (for example, a control flow) at a time and generate a test case that implements that specific objective. They developed this technique by working with the bytecode API. In this study, each control flow of the condition and loop blocks is controlled by the opcode parsing method developed on the bytecode. Separate automatic unit test methods were created with solutions suitable for all conditions. At the same time, all objects defined within the Java class are individually determined by the bytecode. During unit test generation, it is decided whether to use the mock or stub method in accordance with the selected test case while transferring these objects. While preparing the output of the code of the unit test, notation operations are created automatically according to the chosen test case.

Both TestFul and eToc tools used a search-based approach aimed at creating JUnit test cases to maximize structural coverage. However, eToc has not been updated for several years. Therefore, it does not include the latest developments for generating test data. On the other hand, TestFul differs from EvoSuite in many critical details and does not have a fully automated feature. For example, TestFul requires manual editing of XML files for each tested class. EvoMaster, an automatic test generation RESTful API, developed using evolutionary algorithm and optimization techniques such as the MIO algorithm. It can export test files in JSON format and can be used in integration with software such as EvoMaster. Within the scope of this study, both the data used in the Java class imported to the application and random data that have a similar approach to these data are produced and used as parameters in the methods required for unit testing. All these operations are recorded at run-time both in a backup file to the system and in a NoSql collection in JSON structure.

Charreteur et al., in their work where they used byte code, they used the limited memory variable method in the java virtual machine. Their application named JAUT, which tests input generation at the bytecode level, performs constraint-based test input generation from Java bytecode. Therefore, it is mainly associated with other works named JPF, Cute and Pex. Unlike these three tools, JAUT performs backward discovery, i.e. it starts from a target bytecode location and discovers a suitable path to step-by-step input. In fact, JPF, Cute, and Pex rely on forward symbolic execution, which involves symbolically evaluating instructions along a path in the same order as execution. Within the scope of this study, a system that converts byte codes with opcode parsing method has been prepared and a different perspective has been brought to these studies. Each opcode line obtained after the conversion of the Java class to java byte code was analyzed in turn, and the object, variable or input-output parameters of the opcode, if any, were determined. These values were then used in unit test generation.

On the subject of Assertion, Randoop allows descriptions of the source code to specify the observer methods to be used to create the annotation. Orstra generates assertions based on observed return values and object states, and adds assertions to check future work against these observations. While such approaches can be used to produce efficient objects, they do not serve to determine which of these assertions are actually useful, and so such techniques can only be checked in regression testing. In contrast, the μTest tool uses mutation testing to select an effective subset of assertions via EvoSuite. Within the scope of this study, the case can be selected from a list of the test scenario optionally from the user through the application. In accordance with this test case, annotations are turned into unit tests while turning them into outputs via FTL.

It is thought that the tool developed within the scope of this study will also take an important place in the national software testing field. The developed tool is

intended to be actively used in unit tests to be carried out by software testers. This tool, which can respond to the most basic test scenarios in its current form, has a structure that can be developed about how it should behave in much more advanced scenarios, since it has a bytecode-based framework. One of the biggest reasons for this is that Java has an open source system. Different modules can also be developed for this framework that will translate other relevant bytecodes in future test scenarios that may be required.

With the development of the automatic unit test creation software, which is targeted as a domestic product, automatic unit tests are created after the opcode conversion processes are carried out. This work currently works for a single java file defined, excluding linked classes of objects generated within the class. In future studies, after adding multiple Java files as project integrity to the framework, generating possible automatic unit tests according to the specified test scenarios is planned. In addition, the latest developments will be compared with all other examples given for SF110 corpus of classes [23]. At the same time, random test data within the framework is currently only performed for numerical data types. Another goal of the study is to both replicate these data types and perform random data assignments from a test data set that can be exported by the user. For this, it is planned to use fuzz testing methods.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Damar M, Özdağoğlu G, Özdağoğlu A. Software quality and standards on a global scale: Trends in the literature from scientific and sectoral perspective. Alphanumeric J 2018;6:325-348. [CrossRef]

[2] Felice S. JUnit Vs TestNG: Differences between JUnit and TestNG. Available at: https://www.browserstack.com/guide/junit-vs-testng. Accessed on Jun 26, 2024.

[3] Graham D, Fewster M. Software test automation: effective use of test execution tools. Boston: Addison-Wesley Professional; 1999.

[4] Csallner C, Smaragdakis Y. JCrasher: An automatic robustness tester for Java. Softw Pract Exp 2004;34:1025-1050. [CrossRef]

[5] Pacheco C, Lahiri SK, Ernst MD, Ball T. Feedback-directed random test generation. In proceedings of the 29th International Conference on Software Engineering; 2007: Minneapolis, MN, USA. IEEE; 2007. pp.75-84. [CrossRef]

[6] Pacheco C, Ernst MD. Randoop: Feedback-directed random testing for Java. Proceedings of the Conference on Object-Oriented Programming Systems, Languages, and Applications, OOPSLA 2007:815-816. [CrossRef]

[7] Simons AJH. JWalk: A tool for lazy, systematic testing of java classes by design introspection and user interaction. Autom Softw Eng 2007;14:369-418. [CrossRef]

[8] Sen K, Marinov D, Agha G. Cute: A concolic unit testing engine for C. Available at: https://www.cs.columbia.edu/~junfeng/08fa-e6998/sched/readings/cute.pdf. Accessed Jun 26, 2024.

[9] Charreteur F, Gotlieb A. Constraint-based test input generation for java bytecode. In proceedings of the 21st IEEE International Symposium on Software Reliability Engineering; 2010 Nov; San Jose, CA, USA. IEEE; 2012. [CrossRef]

[10] Fraser G, Arcuri A. Evosuite: automatic test suite generation for object-oriented software. In proceedings of the 19th ACM SIGSOFT symposium and the 13th European Conference on Foundations of Software Engineering; 2011 Sept; New York, United States. 2011. pp. 416-419. [CrossRef]

[11] Sakti A, Pesant G, Gueheneuc YG. Instance generator and problem representation to improve object oriented code coverage. IEEE Trans Softw Eng 2015;41:294-313. [CrossRef]

[12] Tanno H, Zhang X, Hoshino T, Sen K. TesMa and CATG: automated test generation tools for models of enterprise applications. In proceedings of the 37th IEEE International Conference on Software Engineering; 2015 May 16-24; Florence, Italy. IEEE; 2015. pp. 717-720. [CrossRef]

[13] Tzoref-Brill R, Sinha S, Abu Nassar A, Goldin V, Kermany H. TackleTest: A tool for amplifying test generation via type-based combinatorial coverage. Available at: https://research.ibm.com/publications/tackletest-a-tool-for-amplifying-test-generation-via-type-based-combinatorial-coverage. Accessed on Jun 26, 2024.

[14] Higo Y. Constructing dataset of functionally equivalent Java methods using automated test generation techniques. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4543198. Accessed on Jun 26, 2024. [CrossRef]

[15] Lukasczyk S, Fraser G. Pynguin: Automated unit test generation for python. Available at: https://arxiv.org/abs/2202.05218. Accessed on Jun 26, 2024.

[16] Bardin S, Kosmatov N, Marcozzi M, Delahaye M. Specify and measure, cover and reveal: A unified framework for automated test generation. Sci Comput Program 2021;207:102641. [CrossRef]

[17] Arcuri A. RESTful API automated test case generation with EvoMaster. ACM Trans Softw Eng Methodol 2019;28:1-37. [CrossRef]

[18] Venners B. Bytecode basics : A first look at the bytecodes of the Java virtual machine. Available at: https://www.infoworld.com/article/2077233/bytecode-basics.html?page=2. Accessed on Jun 26, 2024.

[19] McMinn P. Search-based software testing: past, present and future. In proceedings of the IEEE Fourth International Conference on Software Testing, Verification and Validation Workshops; 2011 Mar 21-25; Berlin, Germany. IEEE; 2011. pp. 153-163. [CrossRef]

[20] Venkatesan P, Rozario RG, Fiaidhi J. Junit framework for unit testing. Available at: https://www.techrxiv.org/doi/full/10.36227/techrxiv.12092259.v1. Accessed on Jun 26, 2024.

[21] FreeMarker. What is a Apache FreeMarker? Available at: https://freemarker.apache.org/index.html. Accessed on Jun 26, 2024.

[22] Evosuite. SF110 corpus of classes. Available at: https://www.evosuite.org/experimental-data/sf110/. Accessed on Jun 26, 2024.

[23] Fraser G, Arcuri A. A large-scale evaluation of automated unit test generation using evosuite. Available at: https://www.evosuite.org/wp-content/papercite-data/pdf/tosem_evaluation.pdf. Accessed on Jun 26, 2024.

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2023.00070

Research Article

Characterization of sol-gel derived NiO thin films: Effect of post-heat treatment

Djanette MERIEM BLIZAK[1,*], Drifa BRINIS[2]

[1]*Laboratory of Bioinformatics, Applied Microbiology and Biomolecules, University M'Hamed Bougara of Boumerdes, Boumerdes, 57767, Algeria*
[2]*Faculty of Hydrocarbons and Chemistry, University M'Hamed Bougara of Boumerdes, Boumerdes, 57767 Algeria*

ARTICLE INFO

Article history
Received: 04 November 2022
Revised: 28 December 2022
Accepted: 23 March 2023

Keywords:
Heat Treatment; NiO; Sol-gel; Spray Pyrolysis; Thin Films

ABSTRACT

Nickel Oxide (NiO) thin films have a wide range of applications in the field of photovoltaics and optoelectronics technology, due to their excellent properties. By modifying the composition and state of the structure with heat treatments, most of the properties of the film are improved. Thus, this study aim to show the behavior of NiO thin films with the post-heat treatment. We prepared thin films of nickel oxide (NiO) by the spray pyrolysis method, deposited on glass substrates and heated to different temperatures (200 °C, 250 °C and 300 °C). The obtained samples were characterized by UV-Visible spectroscopy, contact angle measurement and Raman spectroscopy. We found that the transmittance of films increases with the annealing temperature and Urbach energy (disorder) decreases with increasing optics gap and annealing temperature. Also, the contact angle increases with the temperature of the annealing, but the surface remains hydrophilic. Furthermore, the Raman spectrum has shown the presence of the crystalline state and amorphous carbon (graphite).

Cite this article as: Meriem Blizak D, Brinis D. Characterization of sol-gel derived NiO thin films: Effect of post-heat treatment. Sigma J Eng Nat Sci 2024;42(4):1009–1015.

INTRODUCTION

Transparent conductive oxides with a wide band gap (>3eV) such as nickel oxide (NiO) are very attractive and of great interest due to their various applications in the microelectronics and optoelectronics devices including radiation detection diodes, photovoltaic cells, diodes emitters, lasers, gas sensors, solar cells and photocatalytic cells [1]. Nickel oxide has p-type oxide semiconductor character with large band gap energy. It provides an antimagnetic arrangement based on the properties of face-centered cubic (Perovskite) crystal symmetry [2]. The lattice parameter is 0.417 nm, which is 18% larger than that of metallic nickel [3]. Nominally pure stoichiometric NiO is an insulator with a resistivity at room temperature of the order of 1013 Ω-cm 68. NiO is characterized by high chemical and thermodynamic [4] stability and high oxidation resistance. Also, It is non-toxic, inexpensive, easy to produce and very abundant in nature,. Depending on the method of preparation, it comes in the form of a more or less dense and less black

*Corresponding author.
*E-mail address: d.blizak@univ-boumerdes.dz; bdmeriem@yahoo.fr

This paper was recommended for publication in revised form by Editor in-Chief Ahmet Selim Dalkilic

greenish-gray powder [5]. Structural properties are very sensitive to the nature and number of point (interstitial, substitute, vacant), line (dislocation) and plane (grain boundary...) defects that may exist inside the crystalline structure [6]. As a semiconductor thin films, it opened new ways of artificial structures for electronic and optical devices [7].

There are various thin film deposition methods that are under continuous development and improvement. The sol-gel process is the best alternative to partially costly and complex techniques which makes possible to produce thin layers by spin-coating, dip-coating and spraying [8]. Sol-Gel is a process for developing materials from precursors in solution. It allows, within the framework of soft chemistry approaches, to produce thin layers consisting from a pile of metal oxides nanoparticles [9]. The simplest deposition technique used in the field of thin layers is spray pyrolysis. in this method, the layers are obtained by atomizing the precursor into fine droplets and transporting these drops onto a heated substrate [10].

Metal oxide thin films like NiO deposited by sol-gel usually require heat treatment in order to induce crystallization and remove inorganic/organic residues. The annealing temperature mainly depends on the nature of the Ni precursor and the solvent [11]. The particles become highly crystalline, resulting in a reduction in pore and grain boundary density. Therefore, the optical band gap (Eg) of films decreases with increasing annealing temperature [12]. However, Madhavi et al. [13] claimed that the optical band gap of NiO thin films increased with increasing annealing temperature up to 400 °C (from 3.43 eV to 3.68 eV) and then decreased to 3.48 eV at a higher annealing temperature of 500 °C due to better crystallization of NiO films. On the other hand, in the study of Mallick [14] a NiO film annealed at 400 °C showed an optical band gap of 3.52 eV which increased to 3.62 eV with increasing annealing temperature at 600 °C. In another study [15], NiO films were annealed for 2 h between 100 °C and 400 °C in ambient air. As the annealing temperature increased, the film became transparent in the visible and near infrared region. On the other hand, the duration of the heat treatment in the atmospheric environment also affects the structure in two different aspects. As the annealing time increases the particle size of the structure increases, In addition, with the increase in the interaction time of the transparent conductive oxide layers with the heat treatment atmosphere, the quantity of oxide phase increases due to the amount of oxygen in the structure. The chemically absorbed oxygen reduces the carrier concentration by acting as a trap against free carriers. Due to the decrease in carrier concentration, the band gap narrows [16].

In this work, we report the results of characterization of the thin layers of NiO deposited by the spray pyrolysis technique on the glass substrates, while showing the effect of post-heat treatment study the hydrophilic properties.

EXPERIMENTAL METHOD

The experimental part of our study was carried out in two parts: The elaboration of the samples and the characterization of the NiO thin films. The experimental procedure that we adopted in this study is summarized by the concept map below (Figure 1).

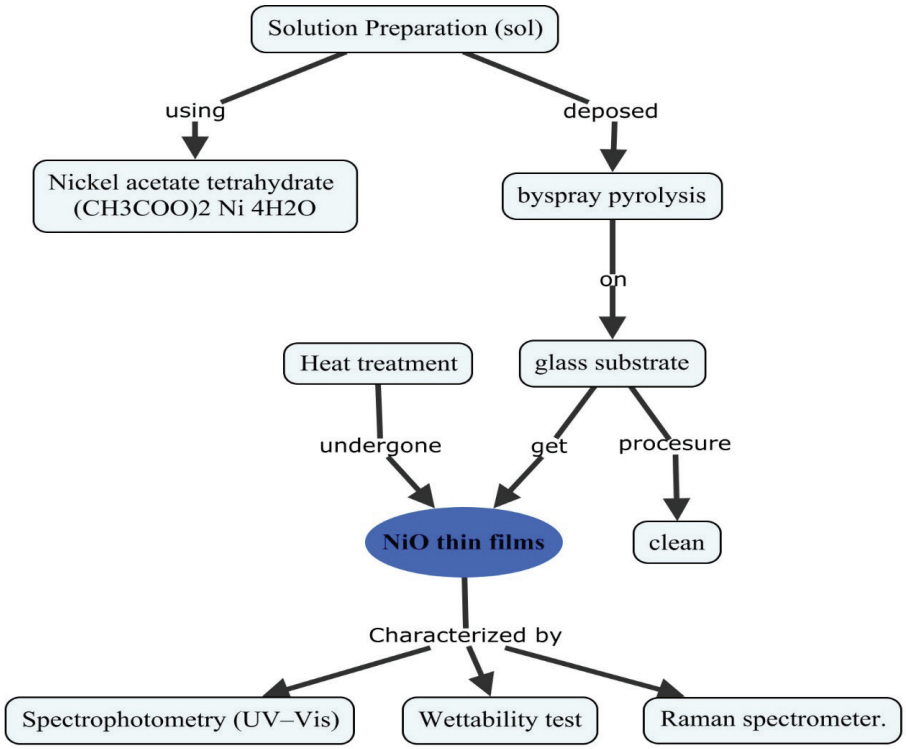


Figure 1. The experimental procedure.

Sample Preparation

The precursor solution (sol) with a concentration of 0.1 M was obtained by dissolving a few grams (m=2.49 g) of nickel acetate tetrahydrate $(CH_3COO)_2$ Ni $4H_2O$ (Biochem, 99.5%) in a volume of 100 ml of distilled water (solvent). After stirring for one hour with a magnetic stirrer at 60°C, "a few drops" of Monoethanolamine C_2H_3NO (MEA) was added to stabilize the solution. The suspension is left stirring for 2 hours at 60° and 4 hours at ambient temperature. We added a few drops of HCL (Sigma Aldrich 35%) to lower the pH and obtain a transparent solution of bluish color and slightly viscous. The sol is left to stand for 72 hours in a beaker, covered to prevent any sort of contamination and evaporation, before being used for the deposition of NiO thin films.The commercial transparent glass slides (1.5 cm x 2.5 cm, 1 mm thickness) used as substrates in our study were cleaned using the following steps:

- Cleaning with warm water and soap
  Immersion in acetone (C_3H_5OH) for 10 minutes, in an ultrasonic bath to eliminate all traces of oil.
- Rinse with distilled water under ultrasound for 5 minutes.
  Soaking in ethanol (C_2H_5OH) for 10 minutes in an ultrasonic bath to remove impurities that remain stuck to the surface.
- After Rinse with distilled water and drying, the substrates are kept free from all kinds of impurities.

In this simplified spray technique, atomization is based on hydraulic pressure. Pulsed solution feed with an interval of 5s of spraying and 60 s of pause was used. A volume of 30 ml of the solution was sprayed at a distance of 20 cm from the substrate heated to 250 °C. The thin layers were deposited on several glass substrates at the same time and under the same conditions.

The samples prepared by technical spray were thermally annealed in air for 2 hours at different temperatures (200, 250 and 300 °C.). We have chosen to use temperatures which are not very high and extend the annealing time. We used a digital muffle furnace . Our samples underwent slow cooling to room temperature after the oven was turned off.

Sample Characterization

After annealing, the NiO films were characterized using UV–Visible Spectrophotometer (JENWAY 6715) to record the transmission spectrum of the films. To study the hydrophilic properties of NiO films, we measured the contact angle of water drop on the layers using a standard goniometer and ImageJ software. A micro-Raman spectrometer (Lab-Ram HR) was used as a qualitative structural and molecular analysis technique. This device uses He-Cd laser (325 nm, 32 mW). The thickness of the films has been measured using Stylus profilometer (MUTITOYO).

RESULTS AND DISCUSSION

The Effect of Post-Heat Treatment on the Thickness of Thin Layers

The values of the thicknesses of the NiO thin films after post-heat treatment are shown in the Table 1. We notice that the values of the thickness of the nickel oxide decrease with the increase of the annealing temperature. According to Godse et al. [17], this decrease in thickness is due to the increase in crystal grain size with increasing annealing temperature. In their study, they found that the thicknesses of nanocrystalline NiO thin films produced by the sol-gel spin coating method using nickel acetate $Ni(CH_3COO)_2$ $4H_2O$ as the Ni source, vary from 906 nm to 642 nm for the annealing temperature of 400 °C and 600 °C respectively.

The Effect of Post-Heat Treatment on Water Contact Angle

The ImageJ software was used to measure the contact angles between the water drop and the sample surface (Figure 2). The angle values are shown in the Table 2. Knowing that the contact angle with surface water is

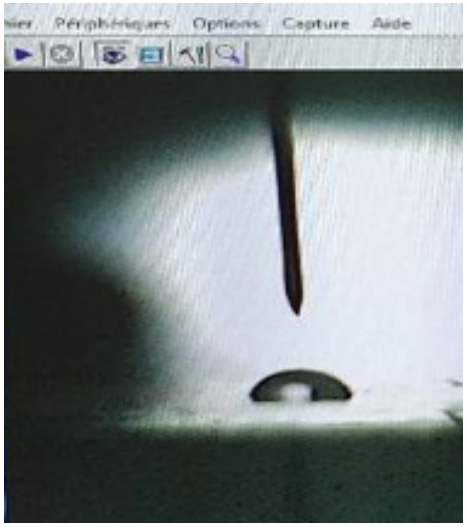


Figure 2. Contact angle measurement.

Table 1. The effect of annealing on thickness

	Annealing temperature	Annealing time	Thickness
Sample 1	200°C	2H	225 nm
Sample 2	250°C	2H	223 nm
Sample 3	300°C	2H	218 nm

Table 2. Water contact angle variation

	Galss	Sample 1	Sample 2	Sample 3
Annealing temperature	NO ANNEALING	200°C	250°C	300°C
Contact angle	35,305	61.2	73	86

inversely proportional to the wettability, we deduce that all the layers of NiO produced are hydrophilic because the contact angles with water are less than 90°. This may be due to the strong cohesive force between the water droplet and the NiO oxide [18].

It is important to emphasize that the contact angle between the water drop and the sample surface is strongly influenced by the topological structure, directly implying the wettability state of the film. the data obtained show that with the increase in the annealing temperature, the wettability decreases, which reveals that the structure of the samples has a more crystalline character [19].

The Effect of Post-Heat Treatment on Optical Properties

The spectra of the transmittance of thin layers of NiO as a function of the wavelength for the three annealing temperatures that we have chosen are presented in Figure 3. In the absorption region (280-380 nm) corresponding to the near UV region, there are sharp drops in the measured transmittances due to the light absorption induced by the electronic excitation [17]. In addition, it can be observed that the transmission level increases with increasing annealing temperature. The variation in transparency can be attributed to the change in structure from amorphous to crystalline caused by the increase in temperature [20].

To determine the energy of the band gap (Eg), we used Tauc Method for a direct transition [14]. Figure 4 shows the variation of $(\alpha h\nu)^2$ (α is absorption coefficient) as a function of incident photon energy ($h\upsilon$) for (NiO) thin films deposited on glass substrates at 200 °C, 250 °C and 300 °C. We observe that the Eg increases with increasing annealing temperature from 3.42 eV to 3.63 eV. These results are in good agreement with the values obtained in previous researches [14,17, 21-25] as shown in Table 3. The blue shift Eg is due to the enhancement of crystal quality [4]. Due to the effect of temperature, the widening of the band gap is produced by the filling of electrons near the bottom of the valence band Ev which have moved to the upper part of the conduction band Ec. The increase in the forbidden bandwidth by the additional electrons in Ec, favors an increase in the concentration of charge carriers [26].

the NiO thin films in our work were worked out by the sputtering technique, the atoms arriving on the substrate after the sputtering are not usually in an ideal position to form a desired compound, which can cause different types of defects in the structure. The existence of this disorder can be expressed by the Urbach energy (Eu). We observe in Figure 5 that the energy value of Urbach Eu decreases with the increase of the annealing temperature from 3.06 eV to 2.1 eV. While Ikhmayies et al. [27], stated that Eu interpreted as the tail width of band gap localized states is weakly temperature dependent.

We believe that the disordered and amorphous structure of our layers producing the extended localized states in the band gap has been reduced which has been reflected by the decrease in the values of Eu.

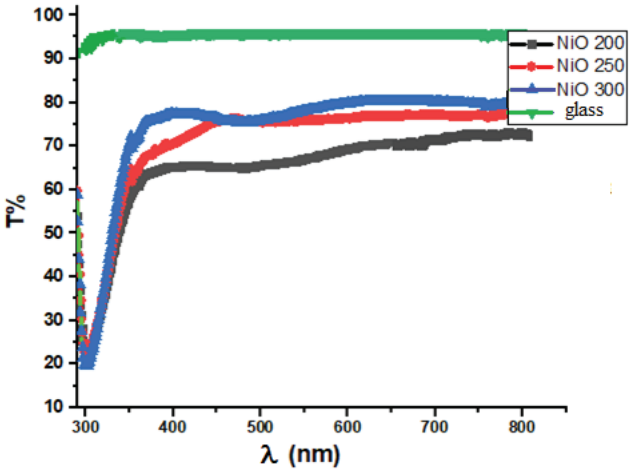


Figure 3. Transmittance spectra as a function of wavelength.

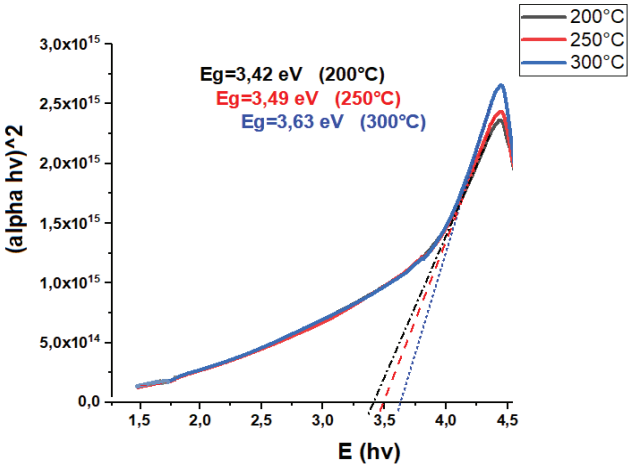


Figure 4. Variation of ($\alpha h\nu$) as a function of incident photon energy ($h\upsilon$).

Table 3. NiO thin film gap value in previous research

Thin film deposition technique	Precursor	T°C	Thickness	Gap (eV)	Reference
Spin coating	$Ni(NO_3)_2$ $6H_2O$	400	/	3.52	[14]
Spin coating	Nickel acetate	400	906 nm	3.86	[17]
Spray pyrolysis	nickel chloride	350	/	3,51	[21]
Spray pyrolysis	Nickel chloride	/	230	3,4	[22]
Spray pyrolysis	Nickel acetate	450	170	3,65	[23]
Spray pyrolysis	Nickel acetate	350	/	3.83	[24]
Spin-coating	Nickel acetate	450	/	3.7	[25]

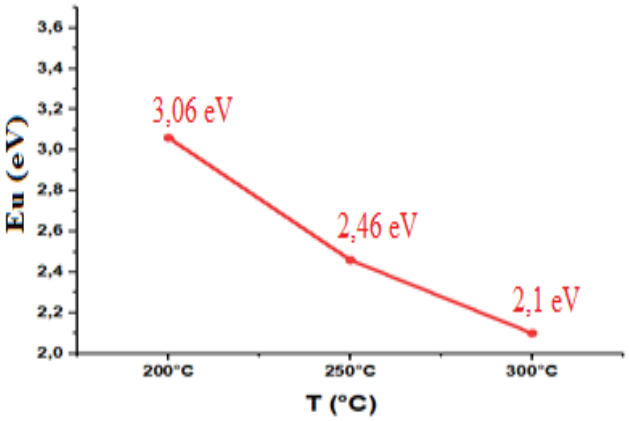


Figure 5. Variations of the optical gap with the Urbach energy as a function of the annealing temperature.

The effect of post-heat treatment on the structural properties of thin NiO layers

The Raman spectrum of the NiO thin film in Figure 6 shows the positions of the peaks corresponding to the characteristic vibrations of the molecular bonds. Ni-O bond peaks were observed at 574 and around 1094 cm^{-1} confirming the presence of NiO in crystalline state for all of the three

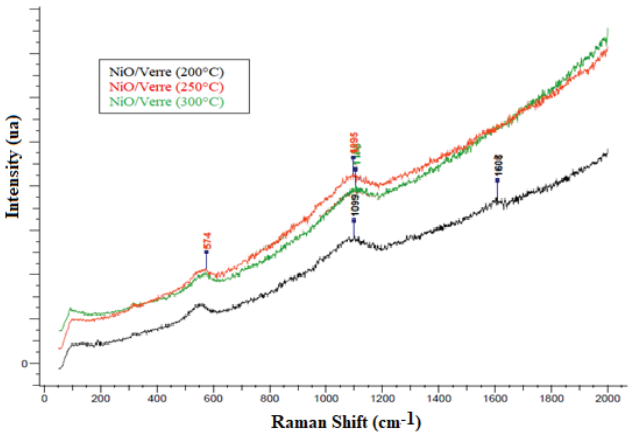


Figure 6. Raman photoluminescence emission spectra of NiO thin films.

thermal anneals. However, in the case of disordered semiconductors, such as amorphous and nanocrystalline, the phonons other than centered phonons on the area can also be observed in their Raman spectra [28]. The 547 cm^{-1} band is generally attributed to first-order one-phonon (1P) excitation related to nickel starvation point defects [29]. While the band at about 1090 cm^{-1} corresponds to the second-order longitudinal two-phonon (2P) vibrational modes.

For the film annealed at 200 °C, a peak at about 1583.2 cm^{-1} due to optical vibrations (mode) in the graphite (presence of amorphous carbon) [30].

CONCLUSION

In order to demonstrate the behavior of thin layers of NiO, with subsequent heat treatment, we developed thin layers of nickel oxides by sputtering technique on glass substrates. We have chosen three annealing temperatures and several characterization technics. The optical properties carried out with spectrophotometry have shown that the transmittance of the NiO films increases with the annealing temperature. the values of the optical gap obtained (3.42 eV, 3.49 eV and 3.63 eV) increase with increasing annealing temperatures, which is in agreement with other studies. It was found that Urbach energy (disorder) decreases with increasing gap and annealing temperature. The measured water contact angles are all less than 90°, which indicates that the NiO films produced are hydrophilic. Raman spectrum confirms the presence of the crystalline state in all samples. It revealed also, the presence of carbon in the NiP thin film annealed at 200 °C showed the presence of amorphous carbon (graphite).

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Herissi L, Hadjeris L, Aida M S, Bougdira. J. Properties of (NiO)1-x (ZnO)x thin films deposited by spray pyrolysis. Thin Solid Films 2016;605:116–120. [CrossRef]

[2] Blizak D, Remli S, Blizak S, Bouchenak O, Yahiaoui K. NiO thin films for environmental photocatalytic applications: A review. Alger J Environ Sci Technol 2021;7:1950–1957.

[3] Bouzoubaa A. Modélisation atomistqiue des interactions entre les ions chlorures et la surface du nockel passivé. Dissertation. Paris: Paris Sciences and Lettres Univ; 2008.

[4] Chavillon B. Synthèse et caractérisation d'oxydes transparents conducteurs de type p pour application en cellules solaires à colorant. Dissertation. Loire-Atlantique: Nantes Univ; 2011.

[5] Gerl M, Issi JP. Physique des matériaux. Lausanne: EPFL Press; 1997.

[6] Ghougali M. Elaboration and characterization of nanostructuring NiO thin films for gas sensing applications. Dissertation. Biskra, Algeria: Univ of Mohamed Khider; 2019.

[7] Gürel HH, Akıncı Ö, Ünlü H. Modeling Of heterostructures for nanoelectronic devices: Tight binding view. Sigma J Eng Nat Sci 2004;1:1–20.

[8] Gözükızıl MF, Temel S, Özbay N. Some physical properties of czo thin films produced by a novel magnetic spin coating technique. Sigma J Eng Nat Sci 2020;38:955–960.

[9] Bokov D, Jalil AT, Chupradit S, Suksatan W, Ansari MJ, Shewael IH, et al. Nanomaterial by sol-gel method: synthesis and application. Adv Mater Sci Eng 2021;2021:5102014. [CrossRef]

[10] Perednis D, Gauckler LJ. Thin film deposition using spray pyrolysis. J Electroceram 2005;14:103–111. [CrossRef]

[11] Yang P, Li L, Yu S, Zheng H, Peng W. The annealing temperature and films thickness effect on the surface morphology, preferential orientation and dielectric property of NiO films. Appl Surf Sci 2019;493:396–403. [CrossRef]

[12] Ho SM. Preparation and characterization of nickel oxide thin films: A review. Int J Appl Chem 2016;12:87–93.

[13] Madhavi A, Harish GS, Reddy PS. Effect of annealing temperature on optical and electrical properties of electron beam evaporated NiO thin films. Int J Sci Technol Eng 2016;2:742–747.

[14] Mallick P. Effect of Thermal Annealing on the Optical Band Gap of NiO Thin Film. Orissa J Phys 2017;24:91–94.

[15] Atak G, Coşkun ÖD. Annealing effects of NiO thin films for all-solid-state electrochromic devices. Solid State Ion 2017;305:43–51. [CrossRef]

[16] Dominguez MA, Flores F, Martinez J, Orduña-Diaz A. Effects of low-temperature annealing on electrical properties of Thin-film Transistors based on Zinc Oxide films deposited by ultrasonic spray pyrolysis: Impact of annealing time. Thin Solid Films 2016;615:243–246. [CrossRef]

[17] Patil V, Pawar S, Chougule M, Godse P, Sakhare R, Sen S, et al. Effect of annealing on structural, morphological, electrical and optical studies of nickel oxide thin films. J Surf Eng Mater Adv Technol 2011;1:35–41. [CrossRef]

[18] Yuan Y, Lee TR. Contact Angle and Wetting Properties. In: Bracco G, Holst B, editors. Surface science techniques. Heidelberg, Berlin: Springer; 2013. p. 3–34. [CrossRef]

[19] Navale ST, Mali VV, Pawar SA, Mane RS, Naushad M, Stadler FJ, Patil V. Electrochemical supercapacitor development based on electrodeposited nickel oxide film. RSC Adv 2015;5:51961–51965. [CrossRef]

[20] Atak G, Coşkun ÖD. Annealing effects of NiO thin films for all-solid-state electrochromic devices. Solid State Ion 2017;305:4351. [CrossRef]

[21] Sharma R, Acharya AD, Shrivastava SB, Shripathi T, Ganesan V. Preparation and characterization of transparent NiO thin films deposited by spray pyrolysis technique. Optik 2014;125:6751–6756. [CrossRef]

[22] Patil PS, Kadam LD. Preparation and characterization of spray pyrolyzed nickel oxide (NiO) thin films. Appl Surf Sci 2002;199:211–221. [CrossRef]

[23] Romero R, Martin F, Ramos-Barrado JR, Leinen D. Synthesis and characterization of nanostructured nickel oxide thin films prepared with chemical spray pyrolysis. Thin Solid Films 2010;518:4499–4502. [CrossRef]

[24] Kamel H, Elmaghraby EK, Ali SL, Abdel-Hady K. J Cryst Growth 2004;262:424–434. [CrossRef]

[25] Haider AJ, Al-Anbari R, Sami HM, Haider M J. Enhance preparation and characterization of nickel-oxide as self-cleaning surfaces. Energy Procedia 2019;157:1328–1342. [CrossRef]

[26] Raj AM, Lalithambika KC, Vidhya VS, Rajagopal G, Thayumanavan A, Jayachandran M, et al. Growth mechanism and optoelectronic properties of nanocrystalline In2O3 films prepared by chemical spray pyrolysis of metal-organic precursor. Phys B Condens Matter 2008;403:544–554. [CrossRef]

[27] Ikhmayies SJ, Ahmad-Bitar RN. A study of the optical bandgap energy and Urbach tail of spray-deposited CdS: In thin films. J Mater Res Technol 2013;2:221–227. [CrossRef]

[28] Turan R, Serincan U, Kartopu G, Finstad TG. Structural and optical properties of si and ge nanocrystals embedded in SiO2 matrix by ion implantation. Sigma J Eng Nat Sci 2004;22:1–12.

[29] Rooksby HP. Structure of nickel oxide. Nature 1943;152:304. [CrossRef]

[30] Roro KT, Mwakikunga B, Tile N, Yalisi B, Forbes A. Effect of accelerated thermal ageing on the selective solar thermal harvesting properties of multiwall carbon nanotube/nickel oxide nanocomposite coatings. Int J Photoenergy 2012;2012:678394. [CrossRef]

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1016–1022, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00056

Research Article

The effect of alternative measurement and evaluation techniques on safety performance in employee training

Orkun DALYAN[1,*], Hatice DALYAN[1], Mehmet PİŞKİN[2,*], Ömer Faruk ÖZTÜRK[3], Erdal CANPOLAT[4]

[1]*Department of Occupational Health and Safety, Intitute of Graduate Studies, Çanakkale Onsekiz Mart University, Çanakkale, 17100, Türkiye*
[2]*Department of Food Processing, Çanakkale Technical Sciences Vocational School, Çanakkale Onsekiz Mart University, Çanakkale, 17020, Türkiye*
[3]*Department of Chemistry, Faculty of Science, Çanakkale Onsekiz Mart University, Çanakkale, 17020, Türkiye*
[4]*Department of Science Education, Faculty of Education, Fırat University, Elazığ, 23119, Türkiye*

ARTICLE INFO

Article history
Received: 09 September 2023
Revised: 16 December 2023
Accepted: 15 April 2024

Keywords:
Alternative Evaluation; Alternative Measurement; Employee Training; Safety Performance

ABSTRACT

Occupational health and safety training constitutes an important implementation step of the proactive employee welfare, health, and safety approach. In occupational health and safety training, measurement and evaluation methods should be applied accurately, efficiently, and effectively to evaluate whether learning and behavioral goals are achieved. In this research, a survey consisting of thirty questions was applied to a sample of 140 employees to determine the effect of alternative assessment tools on safety performance in occupational health and safety training. The obtained data were analyzed using the Statistical Package for Social Sciences 22 software program for reliability and exploratory factor analysis, and the Analysis of Moment Structures and Structural Equation Model 21 software program for confirmatory factor analysis, respectively. It was determined that the average score of the experimental group subjected to hazard detection poster exams using alternative assessment tools exceeded the average score of the control group taking traditional multiple-choice exams and that the level of participation in occupational health and safety regulations had a moderate positive effect on safety performance.
The importance of both compulsory occupational health and safety training for primary school graduates and customized training programs for this group was emphasized. There was a positive and significant correlation between occupational health and safety compliance and safety performance. It was strongly suggested that interventions be tailored to optimum effectiveness according to training levels and that businesses also take proactive approaches to increase compliance with occupational health and safety regulations, such as organizing incentive activities and providing ongoing occupational health and safety training. A safer workplace environment should be promoted by effectively transforming the knowledge and skills acquired in occupational health and safety training into observable behaviors.

Cite this article as: Dalyan O, Dalyan H, Pişkin M, Öztürk ÖF, Canpolat E. The effect of alternative measurement and evaluation techniques on safety performance in employee training. Sigma J Eng Nat Sci 2024;42(4):1016–1022.

***Corresponding author.**
*E-mail address: orkundalyan@outlook.com; mehmetpiskin@comu.edu.tr

This paper was recommended for publication in revised form by Editor in Chief Ahmet Selim Dalkilic

This study was derived from the master's thesis titled "The Effect of Alternative Measurement and Evaluation Techniques in Employee Training on Safety Performance" prepared by Hatice DALYAN under the supervision of Assoc. Prof. Mehmet PİŞKİN.

Published by Yıldız Technical University Press, İstanbul, Turkey

INTRODUCTION

Measurement and evaluation are significant parts of the learning process and are used to understand individuals' levels of learning [1]. Good measurement and evaluation not only help accurately determine individuals' learning levels but also help educators and individuals understand their needs. To create an effective measurement and evaluation system, the steps to follow include selecting appropriate measurement tools, evaluating the suitability of these tools, evaluating individuals' performance, providing feedback, and analyzing data [2].

Occupational health and safety (OHS) is an important issue for the health and safety of everyone in the workplace. Various measures need to be taken to ensure the safety of employees in the workplace. However, for these measures to be effective, employees need to be aware of and trained on this subject [3]. OHS training helps employees identify workplace risks and take measures to protect themselves and other employees [4]. By recognizing workplace hazards and learning safe working methods, employees can prevent accidents and injuries in the workplace.

Assessment tools used in OHS training help evaluate the effectiveness of the training and the individuals' learning levels [5]. These assessment tools include exams and tests, simulations and scenarios, evaluation forms, and feedback [6]. Exams and tests enable individuals to understand what they have learned and remember the information accurately. Simulations and scenarios help employees learn about the hazards and risks they may encounter in real life and how to deal with such situations [7]. Evaluation forms are used to measure employees' awareness and knowledge of OHS issues in the workplace. Feedback can be used to measure individuals' learning levels. These assessment tools help evaluate the effectiveness of OHS training and assist in taking necessary precautions for employee safety [8].

Safety performance is defined as the results that aim for safety, in direct proportion to the safety culture that businesses have [9]. It is also used to refer to OHS performance, defined as the measurable results of a workplace's management of OHS risks [10]. It can motivate employees and use the information to reward them. Various data such as accident statistics, near-miss and hazardous incident reports, management's perspective on safety, and evaluation of training are used as indicators of safety performance. Measuring safety performance allows the evaluation of the compliance of the business or its sub-units with OHS and makes possible the improvement of problematic areas [11]. There are various studies in the literature aiming to identify the components that constitute safety performance. Neal et al. [12] emphasized the significant impact of safety participation and compliance components on safety performance in their study. Similarly, in a study in the healthcare sector that determines the role of safety climate in the correlation between job stress and safety performance, safety performance was examined under two factors employee compliance and participation [13]. Sawacha et al. [14] reported that seven factors affect safety performance in construction sites. Wu et al. [15] emphasized that safety performance was mostly affected by the leadership component of management. The uncertainties in the literature about safety performance have led researchers to identify the factors that affect safety performance.

This study aimed to determine the effects of alternative measurement and evaluation systems in OHS on employee safety performance and to determine the importance of adapting interventions according to training levels. For this purpose, a survey consisting of thirty questions was applied to the personnel working in the construction of the 1915 Çanakkale Bridge. The data obtained was analyzed with SPSS and in light of the results, suggestions were made regarding the criteria that would increase employee safety performance. It was also compared with similar research results in the literature. This research is limited to the opinions of employees working in the bridge construction sector in Çanakkale province. Therefore, research limitations should be considered when generalizing the findings.

MATERIALS AND METHODS

Purpose and Model of the Research

This study aimed to determine the effect of measurement and evaluation systems applied after OHS pieces of training on employee safety performance, and it was designed as a real experimental model with pre-test and post-test control group [16]. Research data were collected between 01.08.2022-31.12.2022.

This study has been approved by the Ethics Committee of Çanakkale Onsekiz Mart University Graduate Education Institute with the approval number E-84026528-050.01.04-2200116793 on June 13, 2022.

Population and Sample Selection

The population of the study consists of all personnel (N=220) working in bridge construction in Çanakkale province, Turkey. In determining the sample group, according to the Krejcie and Morgan [17] population-sample table for a confidence interval of 95% and a significance level of 5%, the sample size of the study with a population of 220 should be 140 individuals. According to the research model, seventy individuals were classified as the control group, and seventy individuals as the experimental group. The convenient sampling method was used in selecting individuals.

Data Collection Tools

The questionnaire expressions applied to the employees are the expressions used in the studies available in the literature [18-20]. It was adapted to this study area in line with expert opinions, considering variables such as sample and sector. The first five statements of the scale are purposed at determining the type of employees (gender, marital status, age, professional experience, and educational status).

The 30-item 5-point Likert scale's purpose is to evaluate employee safety performance. After the pilot application with 70 employees, the reliability of the scale was ensured by the Cronbach alpha internal consistency coefficient, its validity was ensured by factor analysis, and no revision was required. Various opinions exist in the literature regarding the sample size calculation for pilot studies. It has been reported in studies available in the literature that 30-50 participants representing the target group are sufficient for pilot studies [21-25]. The questionnaire was conducted face-to-face with the employees and was filled out a total of 280 times before and after the implementation. The Safety Performance Scale (SPS) statements applied in the study are given in Appendix 1.

Data Processing and Analysis

Reliability and Exploratory Factor Analyses (EFA) were conducted using the Statistical Package for the Social Sciences (SPSS) 22, while confirmatory factor analyses were performed using the AMOS 21 program in this study. Normality assumptions of the data were checked by Kolmogorov-Smirnov and Shapiro-Wilk tests and by checking for skewness and kurtosis values. Parametric analysis methods were used for variables that met the normality assumption [26]. The results of the analysis were evaluated with a confidence level of 95% and a significance level of 5%. Cohen's (d) and eta-squared (η^2) effect size coefficients were calculated to define the level of relationship between variables with significant differences [27]. Correlation analysis was preferred in examining the relationship between scale factors [28]. A significance level of 1% was used as a guide for interpreting the results of the correlation analysis [29-31].

RESULTS AND DISCUSSION

The percentages (%) of participants' personal typical are given below. 6.43% of the total participants are female and 93.57% are male. 54 participants 38.57% are married, and 61.43% are single. When the age variable is examined, the group with the highest rate of 38.57% is the age range of 26-35. There are no participants over 65 years of age in the study sample. The group with the highest rate in the variable of professional experience period is between 1-5 years with 35%. 44 participants (%31.43) had primary education, 53 (%37.86) had high school education, 21 (%15.00) had associate degrees, and 22 (%15.71) had bachelor's degrees.

The reliability coefficient of the SPS is found to be α=0.880, which is highly reliable [32]. Before the reliability and EFA, the Kaiser-Meyer-Olkin (KMO) and Bartlett tests were performed to determine the adequacy of the sample size for analysis. As a result of the analysis, the KMO value was found to be 0.683 and the significance value was 0.000. The obtained KMO value agrees with the literature for good factor analysis [33.34]. The data for SPS factors are given in Table 1.

According to the results of EFA, the SPS was found to have a 5-factor structure. The total variance explanation rate

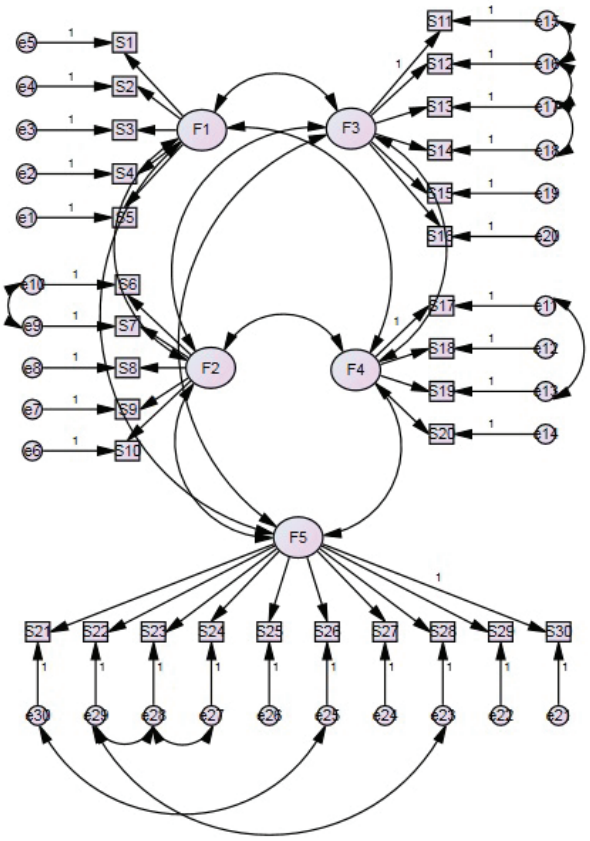


Figure 1. Structural equation model of Safety Performance Scale with plotted covariances.

Table 1. Statistics of Safety Performance Scale Factors

Code	Item	Factor	α	Eigenvalue	Variance	$\bar{X}$	df	S.E.mean
F-1	5	Participation in OHS Rules	0.744	7.203	24.010	3.69	0.76	0.09
F-2	5	Compliance with OHS Rules	0.857	2.991	9.968	3.73	0.96	0.11
F-3	6	OHS Training	0.799	2.820	9.400	3.75	0.69	0.08
F-4	4	OHS Risk Perception	0.713	2.488	8.294	3.15	0.69	0.08
F-5	10	OHS Awareness	0.833	1.630	5.434	3.12	0.65	0.07

Note: α = Cronbach alpha reliability coefficient, $\bar{X}$= Mean, df= Degree of freedom, S.E.mean= Standard Error Mean

of the 5 factors is 57.10%, which is sufficient for explaining the measuring tools [35]. To evaluate the suitability of the factors obtained from EFA, Confirmatory Factor Analysis (CFA) was performed. The Structural Equation Model (SEM) of the SPS, which shows the covariances drawn because of the analysis, is given in Figure 1.

As a result of the SEM created, the results of the CFA are given below.

- Relative Chi Square Index (CMIN/DF): 1.565
- Goodness of Fit Index (GFI): 0.896
- Normed Fit Index (NFI): 0.887
- Comparative Fit Index (CFI): 0.909
- Root Mean Square Error of Approximation (RMSEA): 0.060

The results of the CFA conducted in the SEM are within the accepted values in the literature [36-40]. Because of the analyses, it was defined that SPS is appropriate for measuring with 30 items and five factors (sub-dimensions). The Independent Samples' t-test results of the pre-test score averages and the post-test and pre-test difference mean of the control and experimental groups are given in Table 2.

Accordingly, in Table 2, the pre-test scale mean of the control group participants is $\bar{X}$=3.45 ± 0.49, and the pre-test scale mean of the experimental group participants is $\bar{X}$=3.46 ± 0.52. Accordingly, in the Independent Samples t-test results, there is no statistically significant difference between the pre-test scale means of the control and experimental groups (t=0.132, p>0.05). This finding supports that the pre-test scale means the control and experimental groups are not statistically different from each other. A statistically significant difference was found in the pre-and post-test difference averages between the control and experimental group participants in favor of the experimental group (t=3.231, p<0.05). The effect of the significant difference detected is of moderate level according to the effect size coefficient (d=0.54) [41]. The one-way analysis of variance (ANOVA) results for the OHS Training (F-3) and the education status variable with the pre-test and post-test difference averages are presented in Table 3.

According to Table 3, there is a significant difference between the pre-and post-application mean scores of the OHS Education factor and the educational level variable [$F_{(3,136)}$= 3.740, p<0.05]. A significant difference was found in favor of the primary education level compared to high school, associate degree, and bachelor's degree levels. It was determined that the effect of the OHS Education factor on the education level variable was moderate (η^2=0.07). Pearson Correlation analysis results between the SPS factors are given in Table 4.

According to Table 4, there is a moderate positive correlation between participation in OHS rules and compliance with OHS rules, as well as OHS training factors. There is also a moderate positive correlation between OHS risk perception and compliance with OHS rules and OHS awareness factors.

In a study investigating the effects of alternative evaluation tools on the success scores and knowledge retention of individuals in measuring OHS training, it was reported that exams conducted with a 3D hazard detection poster increased both the success score and knowledge retention of individuals [42]. Similarly, in this study, it was determined that the mean score of the experimental group to

Table 2. Independent Samples t Test results of research groups

	Groups	f	$\bar{X}$	df	S.E.Mean	t	p	d
Pre-test	Control Group	70	3.45	0.49	0.05	0.132	0.895	-
	Experimental Group	70	3.46	0.52	0.06			
Post-test and pre-test difference	Control Group	70	0.71	0.67	0.08	3.231	0.002*	0.54
	Experimental Group	70	1.02	0.44	0.05			

Note: f= Frequency, $\bar{X}$= Mean, df= Degree of Freedom, S.E.mean= Standard Error Mean, *=p<0.05, d= Cohen d Effect Size Coefficient

Table 3. ANOVA results between the OHS Training factor post-test and pre-test difference mean and the educational status variable

Groups	f	$\bar{X}$	df	S.E.	F	p	η^2	Difference
Primary Education	44	0.90	0.87	0.13	3.740	0.013*	0.07	1>2-3-4
High School	53	0.50	0.77	0.10				
Associate's Degree	21	0.48	0.91	0.19				
Bachelor's Degree	22	0.23	0.75	0.15				

Note: f= Frequency, $\bar{X}$= Mean, df= Degree of Freedom, S.E.= Standard Error, *=p<0.05, η^2= Eta-Square Effect Size Coefficient, 1= Primary Education, 2= High School, 3= Associate's Degree, 4= Bachelor's Degree

Table 4. Correlation analysis results between Safety Performance Scale factors

		F-1	F-2	F-3	F-4	F-5
F-1	r	1				
	p	-				
F-2	r	0.367	1			
	p	0.002**	-			
F-3	r	0.391	0.287	1		
	p	0.001**	0.016	-		
F-4	r	0.212	0.306	0.050	1	
	p	0.077	0.010**	0.680	-	
F-5	r	0.203	0.267	0.147	0.470	1
	p	0.091	0.025	0.225	0.000**	-

Note: r= Pearson Correlation Coefficient (two-tailed), *= p<0.05, **= p<0.01, F1= Participation in OHS Rules, F2= Compliance with OHS Rules, F3= OHS Training, F4= OHS Risk Perception, F5= OHS Awareness

which the alternative assessment and evaluation system was applied was higher than the mean score of the control group to which the multiple-choice exam was applied. In studies on safety performance measures available in the literature, it has been reported that no significant difference was found between the demographic characteristics of the participants and safety performance and factors [43]. However, in this study, a significant difference was found between the level of education and the OHS Training factor, in favor of primary school graduates.

In a study examining the effect of organizational factors on safety performance in the oil and gas sector in Nigeria, it was stated that safety training had a significantly positive relationship with both safety compliance and safety participation [44]. In research aimed at developing a model for predicting safety performance in the construction sector, it was reported that feedback on safety-related data, the effectiveness of OHS training, and management support factors would be supported as leading indicators [45]. In a study examining the factors affecting safety performance among employees of 30 construction companies in three major cities in Iran, it was stated that there is a positive relationship between employee competence and OHS training and safety performance [46]. In research conducted with data collected from 128 companies across Turkey, it was reported that the OHS training factor directly and positively affected employee participation and compliance factors [47]. In a study examining the factors affecting safety performance among healthcare providers in Jordan, management commitment and OHS training were reported as primary factors [48]. In this study, it was determined that the compliance with OHS rules factor, compliance with OHS rules and OHS training factors, as well as the factors of OHS risk perception and OHS risk awareness, had a positive effect at a moderate level. It was determined that these findings were different from those of the studies in the construction and steel sectors in the literature [49, 50, 51].

CONCLUSION

In the evaluation of training provided in OHS, the significant positive effects of alternative evaluation tools on safety performance were determined and reported.

The distinctive contribution of this study from others is the necessity of providing mandatory OSH training to primary school graduate employees, as well as the importance of customizing OHS training programs to meet the specific needs of those in this demographic. To optimize the effectiveness of the OSH training planned to be given, it should be adapted to different training levels.

There is a positive relationship between adherence to OSH regulations and safety performance, with increased compliance contributing to a safer working environment. The use of an alternative assessment tool in evaluating OHS training resulted in higher mean scores compared to traditional multiple-choice exams.

It has been reported that effective factors that may affect the evaluation results of OHS training, such as inequalities in question content, individual attitudes towards evaluation tools, and the examination environment, should be taken into consideration.

It is recommended to conduct tests on the Safety Performance Scale devised for this study, as well as alternative measurement tools specifically tailored for Occupational Health and Safety training across diverse sectors and samples. The results obtained from such assessments should be systematically compared with analogous studies, thereby enhancing the overall efficacy of the system.

ACKNOWLEDGEMENTS

This work was supported by the Çanakkale Onsekiz Mart University Scientific Research Projects Coordination Unit under Grant number FYL-2022-4293.

AUTHORSHIP CONTRIBUTIONS

Main Idea/Planning: Mehmet Pişkin, Hatice Dalyan, Orkun Dalyan, Data collection/Processing: Mehmet Pişkin, Hatice Dalyan, Orkun Dalyan, Erdal Canpolat, Ömer Faruk Öztürk, Data analysis and interpretation: Mehmet Pişkin, Orkun Dalyan, Literature review: Mehmet Pişkin, Hatice Dalyan, Orkun Dalyan, Written by: Mehmet Pişkin, Orkun Dalyan, Review and correction: Mehmet Pişkin, Hatice Dalyan, Orkun Dalyan, Erdal Canpolat, Ömer Faruk Öztürk, Consultancy: Mehmet Pişkin, Ömer Faruk Öztürk, Erdal Canpolat.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Baran H. Assessment and evaluation in open and distance education. AUAd 2020;6:28–40.

[2] Endroyo B, Yuwono BE, Mardapi D, Soenarto. Model of learning/training of occupational safety & health (OSH) based on industry in the construction industry. Procedia Eng 2015;125:83–88. [CrossRef]

[3] Dalyan O, Canpolat E, Pişkin M. Effects of occupational health and safety training on the perception of hazard. Int J Adv Eng Pure Sci 2021;33:670–676. [CrossRef]

[4] Kim NK, Abdul Rahim NF, Iranmanesh M, Foroughi B. The role of the safety climate in the successful implementation of safety management systems. Saf Sci 2019;118:48–56. [CrossRef]

[5] Dalyan H, Dalyan O, Öztürk ÖF, Pişkin M. Comparison of face-to-face and distance education systems in occupational health and safety. Karaelmas J Occup Health Saf 2021;5:219–228. [CrossRef]

[6] Xu S, Sun M, Fang W, Chen K, Luo H, Zou P. A Bayesian-based knowledge-tracing model for improving safety-training outcomes in construction: An adaptive learning framework. Dev Built Env 2023;13:100111. [CrossRef]

[7] Rokooei S, Shojaei A, Alvanchi A, Azad R, Didehvar N. Virtual reality application for construction safety training. Saf Sci 2023;157:105925. [CrossRef]

[8] Ceylan H. Occupational health and safety education in Turkey problems and solution suggestions. Ejovoc 2012;2:94–104.

[9] Christian MS, Bradley-Geist JC, Wallace JC, Burke M. Workplace safety: A meta-analysis of the roles of person and situation factors. J Appl Psychol 2009;94:1103–1127. [CrossRef]

[10] Burke M, Signal DeBose S. Workplace safety: A multilevel, interdisciplinary perspective. Res Pers Hum Resour Manag 2010;29:1–47. [CrossRef]

[11] Lingard H, Wakefield R, Cashin P. The development and testing of a hierarchical measure of project OHS performance. Eng Constr Archit Manag 2011;18:30–49. [CrossRef]

[12] Neal A, Griffin MA, Hart PM. The impact of organizational climate on safety climate and individual behavior. Saf Sci 2020;34:99–109. [CrossRef]

[13] Ayyıldız F, Çam Dİ. The moderator role of safety climate in the relationship between job stress and safety performance: An investigation in the health sector. Stud Psychol 2020;40:451–475.

[14] Sawacha E, Naoum S, Fong D. Factors affecting safety performance on construction sites. Int J Proj Manag 1999;17:309–315. [CrossRef]

[15] Wu TC, Chen CH, Li CC. A correlation among safety leadership, safety climate, and safety performance. J Loss Prev Process Ind 2008;21:307–318. [CrossRef]

[16] Karasar N. Bilimsel araştırma yöntemleri. Ankara: Nobel Publishing; 2012. [Turkish]

[17] Krejcie VR, Morgan WD. Determining sample size for research activities. Educ Psychol Meas 1970;30:607–610. [CrossRef]

[18] Üngüren E, Koç TS. Occupational health and safety application performance evaluation scale: Validity and reliability study. Soc Secur J 2015;5:124–144.

[19] Olcay ZF. Occupational health and safety culture scale; Validity and reliability study. EJOSAT 2021;23:678–685.

[20] Ekingen E. Adaptation of occupational safety performances scale into Turkish. BANU J Health Sci Res 2021;2:110–117. [CrossRef]

[21] Yılmaz Ö, Tuncer M. The importance of pilot study in experimental research: The effect of teaching on academic success according to Dale's cone of experience. Electron J Educ Sci 2020;9:89–96.

[22] In J. Introduction of a pilot study. Korean J Anesthesiol 2017;70:601–605. [CrossRef]

[23] Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomized trial to minimize the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res 2016;25:1057–1073. [CrossRef]

[24] Hertzog MA. Considerations in determining sample size for pilot studies. Res Nurs Health 2008;31:180–191. [CrossRef]
[25] Eldridge SM, Costelloe CE, Kahan BC, Lancaster GA, Kerry SM. How big should the pilot study for my cluster randomized trial be?. Stat Methods Med Res 2016;25:1039–1056. [CrossRef]
[26] Sallis JE, Gripsrud G, Olsson UH, Silkoset R. Research methods and data analysis for business decisions. Berlin: Springer; 2022. [CrossRef]
[27] Muijs D. Doing quantitative research in education with IBM SPSS statistics. New York: SAGE Publications Inc.; 2022.
[28] Metsämuuronen J. Artificial systematic attenuation in eta squared and some related consequences: attenuation-corrected eta and eta squared, negative values of eta, and their relation to Pearson correlation. Behaviormetrika 2023;50:27–61. [CrossRef]
[29] Schober P, Boer C, Schwarte LA. Correlation coefficients: Appropriate use and interpretation. Anesth Analg 2018;126:1763–1768. [CrossRef]
[30] Mukaka MM. Statistics Corner: A guide to the appropriate use of correlation coefficient in medical research. Malawi Med J 2012;24:69–71.
[31] Akoğlu H. User's guide to correlation coefficients. Turk J Emerg Med 2018;18:91–93. [CrossRef]
[32] Tavşancıl E. Tutumların ölçülmesi ve SPSS ile veri analizi. Ankara: Nobel Publishing; 2019. [Turkish]
[33] Büyüköztürk Ş. Factor analysis: Basic concepts and using to development scale. Educ Manag Theory Pract 2002;32:470–483.
[34] Kalaycı Ş. Factor Analysis: Applied multivariate statistical techniques with SPSS. Ankara: Asil Publishing; 2005.
[35] Büyüköztürk Ş. Handbook of data analysis for social sciences. Ankara: Pegem Academy Publishing; 2023.
[36] Tabachnick GB, Fidell LS. Using multivariate statistics. London: Pearson; 2013.
[37] Kline RB. Principles and practice of structural equation modeling. New York: The Guilford Press; 1998.
[38] Sümer N. Structural equation modeling: Basic concepts and applications. Turk Psy Artic 2000;3:49–74.
[39] Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Modeling 1999;6:1–55. [CrossRef]
[40] Hair J, Black W, Babin B, Anderson R, Tatham R. Multivariate data analysis. Upper Saddle River: Pearson Prentice Hall; 2006.
[41] Sullivan GM, Feinn R. Using effect size-or why the p value is not enough. J Grad Med Educ 2012;4:279–282. [CrossRef]
[42] Dalyan O, Dalyan H, Ünal E, Pişkin M. Comparison of classic and alternative measurement- evaluation system in occupational health and safety training given to employees. J Labour Relat 2022;1(Suppl):114–129.
[43] Xia N, Griffin MA, Wang X, Liu X, Wang D. Is there agreement between worker self and supervisor assessment of worker safety performance? An examination in the construction industry. J Saf Res 2018;65:29–37. [CrossRef]
[44] Okezie EO, Nwaogazie IL, Chinemerem P. Influence of organizational factors on safety performance in oil and gas industry in the Niger Delta Region. Nigeria. Arch Curr Res Int 2023;23:15–23. [CrossRef]
[45] Awolusi I, Marks E, Hainen A, Alzarrad A. Incident analysis and prediction of safety performance on construction sites. Civil Eng 2022;3:669–686. [CrossRef]
[46] Pourmazaherian M, Musonda I. Worker competence and safety performance the mediation role of safety orientation in the construction industry. Cogent Public Health 2022;9:2145702. [CrossRef]
[47] Bayram M. Safety training and competence, employee participation and involvement, employee satisfaction, and safety performance: An empirical study on occupational health and safety management system implementing manufacturing firms. Alphanumeric J 2019;7:301–318. [CrossRef]
[48] Ashour AM, Hassan Z, Alekam JME. A Conceptual Framework for upgrading safety performance by influence safety training, management commitment to Safety and work environment: Jordanian Hospitals. Int J Bus Soc Res 2018;8:25–35.
[49] Eyiah AK, Kheni NA, Quartey PD. An Assessment of occupational health and safety regulations in Ghana: A study of the construction industry. J Build Cons Plan Res 2019;7:11–31. [CrossRef]
[50] Mayanja LN, Ddamulira C, Katamba P, Mutekanga DR. Determinants of compliance to occupational health and safety measures in construction sites in Kampala Central Division, Kampala city, Uganda. World J Adv Res Rev 2022;14:475–491. [CrossRef]
[51] Mojapelo J, Mafini C, Dhurup M. Employee perceptions of occupational health and safety standards in the steel industry. Int J Soc Sci Hum Stud. 2016;8:106–121.

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1023–1029, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2023.00049

Research Article

New moving frames for the curves lying on a surface

Akın ALKAN[1,*], Hüseyin KOCAYİĞİT[2], Tuba AĞIRMAN AYDIN[3]

[1]*Gördes Vocational School, Manisa Celal Bayar University, Manisa, 45750,Türkiye*
[2]*Department of Mathematics, Manisa Celal Bayar University, Manisa, 45140, Türkiye*
[3]*Faculty of Education, Bayburt University, Bayburt, 69000, Türkiye*

ARTICLE INFO

Article history
Received: 11 January 2023
Revised: 12 March 2023
Accepted: 04 April 2023

Keywords:
Darboux Frame; Moving Frame; Normal Darboux Vector; Osculator Darboux Vector; Rectifying Darboux Vector

ABSTRACT

In this article, three new orthogonal frames are defined for the curves lying on a surface. These moving frames, obtained based on the Darboux frame, are called "Osculator Darboux Frame", "Normal Darboux Frame" and "Rectifying Darboux Frame", respectively. Also, the Osculator Darboux Frame components and curvatures are calculated for a presented example.

Cite this article as: Alkan A, Kocayiğit H, Ağırman Aydın T. New moving frames for the curves lying on a surface. Sigma J Eng Nat Sci 2024;42(4):1023–1029.

INTRODUCTION

The concept of frame is important in the differential geometry of curves. One of the most important tools used to analyze a curve is a moving frame. The relationship of the vector fields forming the frame at the opposite points of two different curves reveals the special curve pairs [1,2]. Curvature functions are defined on the curve using moving frames [3,4]. These curvature functions are called differential invariants of the curve. Curves become special thanks to the relationships between the differential invariants of the curve [5-10]. Many different frames have been defined in different spaces [11-13]. The most commonly used moving frames are the Frenet frame and Bishop frame for the space curves, and the Darboux frame for the surface curves. The Darboux frame is known as the frame of the curve-surface pair [14-17]. Hananoi et al. describe three new vector fields associated with the Darboux frame along the curve on the surface [15]. In addition, Önder defines three new special curves on the surface, taking these three new vectors into account. In this definition, he names these curves as D_i-Darboux slant helices, where the indices $i \in \{o,n,r\}$ represent the osculator, normal, and rectifying planes of the curve on the surface, respectively [18].

In this study, three new moving frames are constructed for the surface curves using these three new vector fields defined in [15]. The curvature functions are calculated for each frame. Relevant theorems are presented with their proofs.

Preliminaries

In this section, some basic concepts related to the subject discussed are presented.

***Corresponding author.**
*E-mail address: akin.alkan@cbu.edu.tr

This paper was recommended for publication in revised form by Editor in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey

Let M be an oriented surface in 3-dimensional Euclidean space E^3, and let $\alpha(s): I \subset R \to M$ be a unit speed surface curve on M and s be the arc-length parameter of α. If we denote the Frenet frame of α by $\{T, N, B\}$, then the Frenet equations of α are given by

$$T' = \kappa N$$
$$N' = -\kappa T + \tau B$$
$$B' = -\tau N$$

where κ(s) is curvature (or first curvature function), τ(s) is torsion (or second curvature function) and, T, N and B are the unit tangent vector, the principal normal vector and the binormal vector of α, respectively.

On the other hand, since the curve α is a surface curve, it has another orthonormal frame called the Darboux frame and this frame is denoted by $\{T, V, U\}$, where T is the unit tangent of the curve, U is the unit normal of the surface M along the curve α, and V is a unit vector defined by $V = UxT$. Using the fact that the unit tangent T is common in both the Frenet frame and the Darboux frame, the relation between these frames can be given as follows

$$\begin{bmatrix} T \\ V \\ U \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} T \\ N \\ B \end{bmatrix},$$

where θ is the angle between the vectors V and N. The Darboux equations of α are given by

$$\begin{bmatrix} T' \\ V' \\ U' \end{bmatrix} = \begin{bmatrix} 0 & k_g & k_n \\ -k_g & 0 & \tau_g \\ -k_n & -\tau_g & 0 \end{bmatrix} \begin{bmatrix} T \\ V \\ U \end{bmatrix}, \quad (1)$$

where k_n, k_g and τ_g are called normal curvature, geodesic curvature, and geodesic torsion of α, respectively [19]. The relations between these curvatures and κ, τ are given as follows

$$k_g = \kappa\cos\theta,\ k_n = \kappa\sin\theta,\ \tau_g = \tau + \frac{d\theta}{ds}.$$

Definition 2.1: The osculator Darboux vector field for a unit speed curve on an oriented surface M with the Darboux frame $\{T, V, U\}$ is defined as $D_o = \tau_g(s)T(s) - k_n(s)V(s)$ [15].

Definition 2.2: The normal Darboux vector field for a unit speed curve on an oriented surface M with the Darboux frame $\{T, V, U\}$ is defined as $D_n = -k_n(s)V(s) + k_g(s)U(s)$ [15].

Definition 2.3: The rectifying Darboux vector field for a unit speed curve on an oriented surface M with the Darboux frame $\{T, V, U\}$ is defined as $D_r = \tau_g(s)T(s) + k_g(s)U(s)$ [15].

Definition 2.4: A unit speed curve is called a generalized helix if its unit tangent vector makes a constant angle with a fixed direction [14].

Definition 2.5: Let α be a unit speed curve on an oriented surface M and $\{T, V, U\}$ be the Darboux frame along α. The curve α is called a relatively normal-slant helix if the vector field V of α makes a constant angle with a fixed direction, i.e. there exists a constant unit vector d and a constant angle φ such that $\langle V, d\rangle = \cos\varphi$ [16].

Definition 2.6: Let α be a unit speed curve on an oriented surface M and $\{T, V, U\}$ be the Darboux frame along α. The curve α is called an isophote curve (or U-strip slant helix) if the vector field V of α makes a constant angle with a fixed direction, i.e. there exists a constant unit vector d and a constant angle φ such that $\langle V, d\rangle = \cos\varphi$ [17].

The Osculator Darboux Frame

In this section, a new frame called the "Osculator Darboux Frame" and related theorem are presented.

Let $\alpha: I \subset IR \to M$ be a unit speed curve with the Darboux frame $\{T, V, U\}$ on surface M in E^3. Let k_g,k_n, τ_g be the curvatures of the curve α and $D_o = \tau_g(s)T(s) - k_n(s)V(s)$ be the osculator Darboux vector. It is clear that;

$$\tilde{D}_o(s) = \frac{D_o(s)}{\|D_o(s)\|} = \frac{\tau_g(s)}{\sqrt{\tau_g^2(s) + k_n^2(s)}}T(s) - \frac{k_n(s)}{\sqrt{\tau_g^2(s) + k_n^2(s)}}V(s).$$

Since the vector $\tilde{D}_o$ is in the plane spanned by T and V, the vectors $\tilde{D}_o$ and V are perpendicular to each other. So, $\tilde{D}_o \perp U$. Hence, the unit vector $Y_o = \tilde{D}_o \times U$ can be defined. It is clear that the vectors $\tilde{D}_o, U, Y_o$ are unit vectors and perpendicular to each other. Therefore, using these vectors $\tilde{D}_o, U, Y_o$, a new orthonormal frame can be constructed along the curve α on the surface.

Definition 3.1. The frame constructed with the vector fields $\{\tilde{D}_o, U, Y_o\}$ defined as $\tilde{D}_o = \|D_o\|$, $\tilde{D}_o \perp U$, $Y_o = \tilde{D}_o \times U$ is called the osculator Darboux frame, or OD-frame briefly.

Theorem 3.1. Let $\alpha: I \subset IR \to M$ be a unit speed curve on surface M in E^3. The osculator Darboux frame of the curve α is defined as

$$\tilde{D}_o' = -\delta_o Y_o$$
$$U' = \mu_o Y_o$$
$$Y_o' = \delta_o \tilde{D}_o - \mu_o U, \quad (2)$$

where $\delta_o = \left(\frac{\tau_g}{k_n}\right)'\left(\frac{k_n^2}{k_n^2 + \tau_g^2}\right) + k_g$ and $\mu_o = \sqrt{k_n^2 + \tau_g^2}$.

Proof: Firstly; $\tilde{D}_o'$ can be expressed as a linear combination of the vectors $\{\tilde{D}_o, U, Y_o\}$ as follows:

$$\tilde{D}_o' = a_1\tilde{D}_o + a_2 U + a_3 Y_o. \quad (3)$$

By the inner product both sides of the equality (3) with D_o, the equality

$$\left\langle \tilde{D}_o{}', \tilde{D}_o \right\rangle = a_1$$

is obtained. Since $\left\| \tilde{D}_o \right\| = 1$, $\left\langle \tilde{D}_o, \tilde{D}_o \right\rangle = 1$ and thus

$$a_1 = \left\langle \tilde{D}_o{}', \tilde{D}_o \right\rangle = 0.$$

By the inner product of both sides of the equality (3) with U, the equality

$$\left\langle \tilde{D}_o{}', U \right\rangle = a_2$$

is obtained. On the other hand, the vector $\tilde{D}_o$ can be written as

$$\tilde{D}_o = \sin \Phi T - \cos \Phi V, \tag{4}$$

with the help of the equations $\frac{\tau_g}{\sqrt{\tau_g^2 + k_n^2}} = \sin \Phi$ and $\frac{k_n}{\sqrt{\tau_g^2 + k_n^2}} = \cos \Phi$. If the derivatives of the Darboux frame vectors (1) are used in the derivative of the equality (4),

$$\tilde{D}_o{}' = (\Phi' + k_g) \cos \Phi T + (\Phi' + k_g) \sin \Phi V. \tag{5}$$

By the inner product of both sides of the equality (5) with U, the equality

$$a_2 = \left\langle \tilde{D}_o{}', U \right\rangle = 0$$

is obtained. By the inner product of both sides of the equality (3) with Y_o, the equality

$$\left\langle \tilde{D}_o{}', Y_o \right\rangle = a_3$$

is obtained. On the other hand, it is clear that

$$Y_o = \tilde{D}_o \times U = -\cos \Phi T - \sin \Phi V. \tag{6}$$

By the inner product of both sides of the equality (5) with this Y_o, the equality

$$a_3 = \left\langle \tilde{D}_o{}', Y_o \right\rangle = -(\Phi' + k_g)$$

is obtained. In here, by arranging the derivative of the expression $\tan \Phi = \frac{\tau_g(s)}{k_n(s)}$, the equality

$$\Phi'(s) = \left(\frac{\tau_g(s)}{k_n(s)} \right)' \left(\frac{k_n{}^2(s)}{k_n{}^2(s) + \tau_g{}^2(s)} \right) \tag{7}$$

is reached. If this expression is written in the equality $a_3 = -(\Phi' + k_g)$,

$$a_3 = -\left[\left(\frac{\tau_g}{k_n} \right)' \left(\frac{k_n{}^2}{k_n{}^2 + \tau_g{}^2} \right) + k_g \right].$$

Thus, the equation (3) is written as

$$\tilde{D}_o{}' = -\left[\left(\frac{\tau_g}{k_n} \right)' \left(\frac{k_n{}^2}{k_n{}^2 + \tau_g{}^2} \right) + k_g \right] Y_o. \tag{8}$$

Secondly; U' can be expressed as a linear combination of the vectors $\{\tilde{D}_o, U, Y_o\}$ as follows:

$$U' = b_1 \tilde{D}_o + b_2 U + b_3 Y_o. \tag{9}$$

The equality (9) is inner multiplied with $\tilde{D}_o$ and so

$$\left\langle U', \tilde{D}_o \right\rangle = b_1.$$

By the inner product of both sides of the equality (4) with the third equality $U' = -k_n T - \tau_g V$ of the equation system (1), the equality

$$b_1 = \left\langle U', \tilde{D}_o \right\rangle = 0$$

is found. The equality (9) is inner multiplied with U and so

$$\left\langle U', U \right\rangle = b_2.$$

Since $\|U\| = 1$, $\langle U, U \rangle = 1$ and thus

$$b_2 = \left\langle U', U \right\rangle = 0.$$

The equality (9) is inner multiplied with Y_o and so

$$\left\langle U', Y_o \right\rangle = b_3.$$

By the inner product of both sides of the equality (6) with the third equality $U' = -k_n T - \tau_g V$ of the equation system (1), the equality

$$b_3 = \left\langle U', Y_o \right\rangle = \sqrt{\tau_g^2 + k_n^2}$$

is obtained. Thus, the equation (9) is written as

$$U' = \sqrt{\tau_g^2 + k_n^2}\, Y_o. \tag{10}$$

Thirdly; Y_o can be expressed as a linear combination of the vectors $\{\tilde{D}_o, U, Y_o\}$ as follows:

$$Y_o' = c_1 \tilde{D}_o + c_2 U + c_3 Y_o. \tag{11}$$

By the inner product of both sides of the equality (11) with $\tilde{D}_o$, the equality

$$\left\langle Y_o', \tilde{D}_o \right\rangle = c_1$$

is obtained. If the derivatives of the Darboux frame vectors (1) are used in the derivative of the equality (6),

$$Y_o' = \left(\Phi' + k_g\right)\sin\Phi T - \left(\Phi' + k_g\right)\cos\Phi V - \left(k_n \cos\Phi + \tau_g \sin\Phi\right)U. \tag{12}$$

The equality (12) is inner multiplied with the equality (4) and so

$$c_1 = \left\langle Y_o', \tilde{D}_o \right\rangle = \Phi' + k_g.$$

In this here, the equality (7) is used and

$$c_1 = \left(\frac{\tau_g}{k_n}\right)' \left(\frac{k_n^2}{k_n^2 + \tau_g^2}\right) + k_g.$$

By the inner product of both sides of the equality (11) with U, the equality

$$\left\langle Y_o', U \right\rangle = c_2$$

is obtained. Using the equation (12),

$$c_2 = \left\langle Y_o', U \right\rangle = -\sqrt{\tau_g^2 + k_n^2}\,.$$

By the inner product of both sides of the equality (11) with Y_o, the equality

$$c_3 = \left\langle Y_o', Y_o \right\rangle$$

is obtained. Since $\|Y_o\| = 1$, $\langle Y_o, Y_o \rangle = 1$ and thus

$$c_3 = \left\langle Y_o', Y_o \right\rangle = 0.$$

Thus, the equation (11) is written as

$$Y_o' = \left[\left(\frac{\tau_g}{k_n}\right)' \left(\frac{k_n^2}{k_n^2 + \tau_g^2}\right) + k_g\right] \tilde{D}_o - \sqrt{\tau_g^2 + k_n^2}\, U. \tag{13}$$

Finally, the expressions (8), (10) and (13) give the matrix equality

$$\begin{pmatrix} \tilde{D}_o' \\ U' \\ Y_o' \end{pmatrix} = \begin{pmatrix} 0 & 0 & -\delta_o \\ 0 & 0 & \mu_o \\ \delta_o & -\mu_o & 0 \end{pmatrix} \begin{pmatrix} \tilde{D}_o \\ U \\ Y_o \end{pmatrix},$$

for $\delta_o = \left(\frac{\tau_g}{k_n}\right)' \left(\frac{k_n^2}{k_n^2 + \tau_g^2}\right) + k_g$, $\mu_o = \sqrt{k_n^2 + \tau_g^2}$.

Theorem 3.2: In E^3, let $\{D_o, U, Y_o\}$ be the OD-frame and α be a unit speed curve. The curve α is the isophote curve relative to the OD-frame if and only if the expression $\frac{\mu_o}{\delta_o}$ is constant, for $\mu_o \neq 0$ and $\delta_o \neq 0$.

Proof: ($\Rightarrow$) Let α be the isophote curve relative to the OD-frame. Let d be the unit, constant direction and $\langle U, d\rangle = \cos\theta = c \neq 0$. So, the vector d can be expressed as

$$d = a_1 \tilde{D}_o + cU + a_2 Y_o,$$

and the derivative of this equality gives the system.

$$\begin{aligned} a_1' + a_2 \delta_o &= 0, \\ a_2 \mu_o &= 0, \\ a_2' - a_1 \delta_o + c\mu_o &= 0. \end{aligned}$$

Since $\mu_o \neq 0$ and $\delta_o \neq 0$, $a_2 = 0$ and $a_1 =$ constant become. As a result, $\frac{\mu_o}{\delta_o} =$ constant.

($\Leftarrow$) Let $\frac{\mu_o}{\delta_o}$ be constant. This constant value can be selected as $\frac{\mu_o}{\delta_o} = \frac{\cos\theta}{\sin\theta}$. Also the vector d can be taken as $d = \cos\theta \tilde{D}_o + \sin\theta U$. Since $\mu_o \sin\theta = \delta_o \cos\theta$,

$$d' = 0$$

is written with the help of the derivatives of the OD-frame vectors. So, the vector d is constant. In addition, the equality $d = \cos\theta \tilde{D}_o + \sin\theta U$ is inner multiplied with U and so $\langle d, U\rangle = \sin\theta$ Consequently, the constant vector d and U make a constant angle, and the curve α is the isophote curve.

The Normal Darboux Frame

In this section, a new frame called the "Normal Darboux Frame" and related theorems are presented.

Let $\alpha : I \subset IR \to M$ be a unit speed curve with the Darboux frame $\{T, V, U\}$ on surface M in E^3. Let k_g, k_n, τ_g be the curvatures of the curve α and $D_n = -k_n(s)\,V(s) + k_g(s)\,U(s)$ be the normal Darboux vector. It is clear that;

$$\tilde{D}_n(s) = \frac{D_n(s)}{\|D_n(s)\|} = -\frac{k_n(s)}{\sqrt{k_n^2(s)+k_g^2(s)}}V(s) + \frac{k_g(s)}{\sqrt{k_n^2(s)+k_g^2(s)}}U(s).$$

Since the vector $\tilde{D}_n$ is in the plane spanned by U and V, the vectors $\tilde{D}_n$ and T are perpendicular to each other. So, $\tilde{D}_n \perp T$. Hence, the unit vector $Y_n = \tilde{D}_n \times T$ can be defined. It is clear that the vectors $\tilde{D}_n, T, Y_n$ are unit vectors and perpendicular to each other. Therefore, using these vectors $\tilde{D}_n, T, Y_n$ a new orthonormal frame can be constructed along the curve α on the surface.

Definition 4.1. The frame constructed with the vector fields $\left\{\tilde{D}_n, T, Y_n\right\}$ defined as $\tilde{D}_n = \frac{D_n(s)}{\|D_n(s)\|}$, $\tilde{D}_n \perp T$, $Y_n = \tilde{D}_n \times T$ is called the normal Darboux frame or ND-frame briefly.

Theorem 4.1. Let $\alpha : I \subset IR \to M$ be a unit speed curve on surface M in E^3. The normal Darboux frame of the curve α is defined as

$$\begin{aligned}\tilde{D}_n' &= -\delta_n Y_n\\ T' &= \mu_n Y_n\\ Y_n' &= \delta_n \tilde{D}_n - \mu_n T,\end{aligned}$$

where $\delta_n = \left(\frac{k_n}{k_g}\right)'\left(\frac{k_g^{\,2}}{k_n^{\,2}+k_g^{\,2}}\right) + \tau_g$ and $\mu_n = \sqrt{k_n^{\,2}+k_g^{\,2}}$.

Proof: It can be proved in a similar way to Theorem 3.1.

Theorem 4.2: In E^3, let $\left\{\tilde{D}_n, T, Y_n\right\}$ be the ND-frame and α be a unit speed curve. The curve α is the helix relative to the ND-frame if and only if the expression $\frac{\mu_n}{\delta_n}$ is constant, for $\mu_n \neq 0$ and $\delta_n \neq 0$.

Proof: ($\Rightarrow$) Let α be the helix relative to the ND-frame. Let d be the unit, constant direction and $\langle T, d\rangle = \cos\theta = c \neq 0$. So, the vector d can be expressed as

$$d = a_1\tilde{D}_n + cT + a_2 Y_n,$$

and the derivative of this equality gives the system.

$$\begin{aligned}a_1' + a_2\delta_n &= 0,\\ -a_2\mu_n &= 0,\\ a_2' - a_1\delta_n + c\mu_n &= 0.\end{aligned}$$

Since $\mu_n \neq 0$ and $\delta_n \neq 0$, $a_2 = 0$ and $a_1 = \text{constant}$ become. As a result, $\frac{\mu_n}{\delta_n} = \text{constant}$.

($\Leftarrow$) Let $\frac{\mu_n}{\delta_n}$ be constant. This constant value can be selected as $\frac{\mu_n}{\delta_n} = \frac{\cos\theta}{\sin\theta}$. Also the vector d can be taken as $d = \cos\theta\tilde{D}_n + \sin\theta T$. Since $\mu_n \sin\theta = \delta_n \cos\theta$, $d' = 0$ is written with the help of the derivatives of the ND-frame vectors. So, the vector d is constant. In addition, the equality $d = \cos\theta\tilde{D}_n + \sin\theta T$ is inner multiplied with T and so $\langle d, T\rangle = \sin\theta$. Consequently, the constant vector d and T make a constant angle, and the curve α is the helix.

The Rectifying Darboux Frame

In this section, a new frame called the "Rectifying Darboux Frame" and related theorems are presented. Since these theorems given in this section are proven similar to these theorems in section 3, they will be given without proof.

Let $\alpha : I \subset IR \to M$ be a unit speed curve with the Darboux frame $\{T, V, U\}$ on surface M in E^3. Let k_g, k_n, τ_g be the curvatures of the curve α and $D_r = \tau_g(s)T(s) + k_g(s)U(s)$ be the rectifying Darboux vector. It is clear that;

$$\tilde{D}_r(s) = \frac{D_r(s)}{\|D_r(s)\|} = \frac{\tau_g(s)}{\sqrt{k_g^2(s)+\tau_g^2(s)}}T(s) + \frac{k_g(s)}{\sqrt{k_g^2(s)+\tau_g^2(s)}}U(s).$$

Since the vector $\tilde{D}_r$ is in the plane spanned by T and U, the vectors $\tilde{D}_r$ and V are perpendicular to each other. So, $\tilde{D}_r \perp V$. Hence, the unit vector $Y_r = \tilde{D}_r \times V$ can be defined. It is clear that the vectors $\tilde{D}_r, V, Y_r$ are unit vectors and perpendicular to each other. Therefore, using these vectors $\left\{\tilde{D}_r, V, Y_r\right\}$ a new orthonormal frame can be constructed along the curve α on the surface.

Definition 5.1. The frame constructed with the vector fields $\left\{\tilde{D}_r, V, Y_r\right\}$ defined as $\tilde{D}_r = \frac{D_r(s)}{\|D_r(s)\|}$, $\tilde{D}_r \perp V$, $Y_r = \tilde{D}_r \times V$ is called the Rectifying Darboux frame or RD-frame briefly.

Theorem 5.1. Let $\alpha : I \subset IR \to M$ be a unit speed curve on surface M in E^3. The Rectifying Darboux frame of the curve α is defined as

$$\begin{aligned}\tilde{D}_r' &= -\delta_r Y_r\\ T' &= \mu_r Y_r\\ Y_r' &= \delta_r \tilde{D}_r - \mu_r V,\end{aligned}$$

where $\delta_r = \left(\frac{\tau_g}{k_g}\right)'\left(\frac{k_g^{\,2}}{k_g^{\,2}+\tau_g^{\,2}}\right) - k_n$ and $\mu_r = \sqrt{k_g^{\,2}+\tau_g^{\,2}}$.

Theorem 5.2: In E^3, let $\left\{\tilde{D}_r, V, Y_r\right\}$ be the RD-frame and α be a unit speed curve. The curve α is the relatively slant

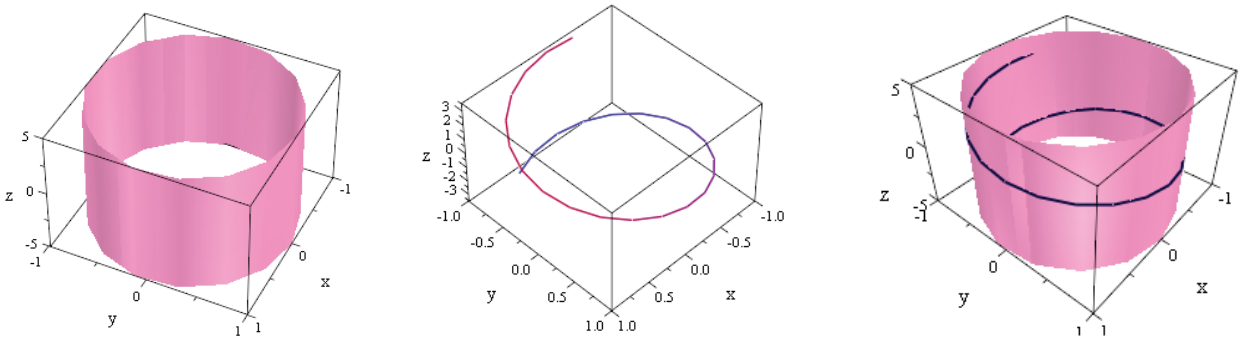


Figure 1. The surface *M*, The curve α(s), The curve α(s) on the surface *M*.

helix with respect to the RD-frame if and only if the expression $\frac{\mu_r}{\delta_r}$ is constant, for $\mu_r \neq 0$ and $\delta_r \neq 0$.

Example: Let's consider the cylinder surface *M* given by parameterization $\varphi(u,v) = (\sin u, \cos u, v)$. The curve $\alpha : I \to M$ given by the parametric form $\alpha(s) = \left(\sin\frac{s}{\sqrt{2}}, \cos\frac{s}{\sqrt{2}}, \frac{s}{\sqrt{2}}\right)$ is a helix on the surface *M*.

The vector fields and curvatures of the Darboux frame for this curve α(s) are calculated as:

$$T(s) = \frac{1}{\sqrt{2}}\left(\cos\frac{s}{\sqrt{2}}, -\sin\frac{s}{\sqrt{2}}, 1\right),$$

$$V(s) = \frac{1}{\sqrt{2}}\left(-\cos\frac{s}{\sqrt{2}}, \sin\frac{s}{\sqrt{2}}, 1\right),$$

$$U(s) = \left(-\sin\frac{s}{\sqrt{2}}, -\cos\frac{s}{\sqrt{2}}, 0\right),$$

$$k_g = 0,\ k_n = 1/2,\ \tau_g = -1/2.$$

The osculator Darboux vector is calculated as

$$\tilde{D}_o = \frac{\tau_g(s)}{\sqrt{\tau_g^2(s) + k_n^2(s)}}T(s) - \frac{k_n(s)}{\sqrt{\tau_g^2(s) + k_n^2(s)}}V(s) = (0,0,-1).$$

Also, it's clear that

$$Y_o = \tilde{D}_o \times U = \left(-\cos\frac{s}{\sqrt{2}}, \sin\frac{s}{\sqrt{2}}, 0\right).$$

Finally, the curvatures of the OD-frame for this curve α(s) are found as $\delta_0 = 0$ and $\mu_o = \frac{1}{\sqrt{2}}$ where

$$\left\langle \tilde{D}_o, \tilde{D}_o \right\rangle = 1, \left\langle U, U \right\rangle = 1, \left\langle Y_o, Y_o \right\rangle = 1,$$
$$\left\langle \tilde{D}_o, U \right\rangle = 0, \left\langle \tilde{D}_o, Y_o \right\rangle = 0, \left\langle Y_o, U \right\rangle = 0.$$

CONCLUSION

In this article, three new orthogonal frames were defined along the curve α lying on the surface, with the help of the Darboux frame.

First, the vector $Y_o = \tilde{D}_o \times U$ was defined because the osculator Darboux vector $\tilde{D}_o$ is perpendicular to the vector *U*. Therefore, using these vectors $\left\{\tilde{D}_o, U, Y_o\right\}$ the osculator Darboux frame was constructed.

After, the vector $Y_n = \tilde{D}_n \times T$ was defined because the normal Darboux vector $\tilde{D}_n$ is perpendicular to the vector *T*. Therefore, using these vectors $\left\{\tilde{D}_n, T, Y_n\right\}$ the normal Darboux frame was constructed.

Finally, the vector $Y_r = \tilde{D}_r \times V$ was defined because the rectifying Darboux vector $\tilde{D}_r$ is perpendicular to the vector *V*. Therefore, using these vectors $\left\{\tilde{D}_r, V, Y_r\right\}$ the normal Darboux frame was constructed.

Also, the Osculator Darboux Frame components and curvatures are calculated for a presented example.

The frame concept is very important in differential geometry. This study has a unique value in terms of putting forward three new frames and it may be a basis for many new studies.

Annotation: This article is prepared from Akın Alkan's doctoral thesis.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw

data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Liu H, Wang F. Mannheim partner curves in 3-space. J Geom 2008;88:120–126. [CrossRef]

[2] J Burke. Bertrand curves associated with a pair of curves. Math Mag 1960;34:60-62. [CrossRef]

[3] Gluck H. Higher curvatures of curves in euclidean space. Am Math Mon 1966;73:699–704. [CrossRef]

[4] Sabuncuoğlu A, Hacısalihoğlu HH. On higher curvatures of a curve. Commun Fac Sci Univ Ankara Ser A1 Math Stat 1975;24:33–46.

[5] Özdamar E, Hacısalihoğlu HH. A Characterization of Inclined Curves in Euclidean n-Space, Communications, Commun Fac Sci Univ Ankara Ser A1 Math 1975;24:15–23. [CrossRef]

[6] Özdamar E, Hacısalihoğlu HH. Characterizations of spherical curves in euclidean n-space. Commun Fac Sci Univ Ankara Ser A1 Math Stat 1974:109–125. [CrossRef]

[7] Izumiya S, Takeuchi N. New special curves and developable surfaces. Turk J Math 2004;28:153–163.

[8] Kula L, Ekmekçi N, Yaylı Y, İlarslan K. Characterizations of slant helices in euclidean 3-space. Turk J Math 2010;34:261–273. [CrossRef]

[9] Öztürk U, Hacısalihoğlu HH. Helices on a surface in Euclidean 3-space.Celal Bayar Univ J Sci 2017;13:113–123. [CrossRef]

[10] Zıplar E, Şenol A, Yaylı Y. On darboux helices in euclidean 3-space. Glob J Science Front Res Math Decis Sci 2012;12:73–80. [CrossRef]

[11] Uzunoğlu B, Gök ·I, Yayli Y. A new approach on curves of constant precession. Appl Math Comput 2016;275:317–323. [CrossRef]

[12] Yılmaz S, Turgut M. 2010. A New version of bishop frame and application to spherical images. J Mathe Anal Appl 2010;371:764–776. [CrossRef]

[13] Bishop LR. There is more than one way to frame a curve. Am Math Mon 1975;82:246–251. [CrossRef]

[14] Düldül M, Uyar Düldül B. Characterizations of helices by using their Darboux vectors. Sigma J Eng Nat Sci 2020;38:1299–1306.

[15] Hananoi S, Ito N, Izumiya S. Spherical Darboux images of curves on surfaces. Beitr Algebra Geom 2015;56:575–585. [CrossRef]

[16] Macit N, Düldül M. Relatively normal-slant helices lying on a surface and their characterizations. Hacettepe J Math Stat 2017;46:397–408.

[17] Doğan F, Yaylı Y. On isophote curves and their characterizations. Turk J Math 2015;39:650–664.

[18] Önder M. Helices associated to helical curves, relatively normal-slant helices and isophote curves. Available at: https://arxiv.org/abs/2201.09684. Accessed on Jul 2, 2024.

[19] O'Neill B. Elementary differential geometry. Cambridge: Academic Press; 1966. [CrossRef]

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1030–1038, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2023.00096

Research Article

Two-stage clustering and routing problem by using FCM and K-means with genetic algorithm

Ebru PEKEL ÖZMEN[1,*], Tarık KÜÇÜKDENİZ[2]

[1]*Department of Industrial Engineering, Samsun University, Samsun, 55139, Türkiye*
[2]*Department of Industrial Engineering, İstanbul University-Cerrahpaşa, İstanbul, 34098, Türkiye*

ARTICLE INFO

Article history
Received: 27 January 2023
Revised: 12 March 2023
Accepted: 04 May 2023

Keywords:
Fuzzy Clustering; Genetic Algorithms; K-means; LAR; SBRP

ABSTRACT

The School Bus Routing Problem (SBRP) is a challenging optimization problem that has received increasing attention in recent years. The problem is composed of three sub-problems: facility location selection, assignment problem, and vehicle routing problem, which can be solved in a single stage or across multiple stages. In this study, we propose a novel two-stage approach to solve the SBRP that combines Fuzzy C Means (FCM) and K-means clustering algorithms with a Genetic Algorithm (GA). In the first stage, we used FCM and K-means to identify the optimal bus stop locations and assigned students to the nearest stop based on the distance metric. This two-stage approach reduces the search space and improves the efficiency of the GA in the second stage. In the second stage, we employed the GA to generate the optimal vehicle route that minimizes the total distance traveled by all vehicles. We compared our results with those in the literature and found that the K Means-GA approach outperformed the previous results. However, the FCM-GA approach yielded significantly inferior results, indicating that the choice of clustering algorithm plays a crucial role in the performance of the overall system. Our study provides insights into the importance of selecting appropriate clustering algorithms for solving the SBRP and proposes a two-stage solution that can be easily implemented in real-world scenarios. Our approach reduces the computational time and provides an effective solution for reducing the total distance traveled by school buses.

Cite this article as: Pekel Özmen E, Küçükdeniz T. Two-stage clustering and routing problem by using FCM and K-means with genetic algorithm. Sigma J Eng Nat Sci 2024;42(4):1030–1038.

INTRODUCTION

With urbanization, urban transportation has become a major concern for many large cities. However, public transit systems are often perceived as inefficient, leading people to rely on their own vehicles or taxis instead. In the context of school transportation, the inefficiency of school-provided transportation systems can result in an increase in private vehicle usage, which contributes to traffic congestion, air pollution, and environmental harm. Therefore, it is essential to develop efficient and cost-effective transportation systems for students that prioritize their comfort and safety while reducing the overall environmental impact.

***Corresponding author.**
*E-mail address: ebru.pekel@samsun.edu.tr

This paper was recommended for publication in revised form by Editor in Chief Ahmet Selim Dalkilic

Several studies have addressed the school bus routing problem (SBRP) in recent years. These studies have proposed various mathematical models and metaheuristic algorithms to improve the efficiency of school transportation systems while balancing the service provider's costs and students' convenience. Some studies have considered factors such as students' travel patterns, class periods, and school bell times to optimize school bus routing. However, there is still room for improvement in terms of minimizing the walking distance of individuals and the total distance traveled by vehicles, especially in the context of the first cluster and second route approach. Miranda, de Camargo [1] have discussed SBRP by extending it with mixed loads. Oluwadare, Oguntuyi [2] proposed a Genetic Algorithm to solve the school bus route problem. Sun, Duan [3] addressed the school bus routing problem in a stochastic and time-dependent road environment.

Mahmoudzadeh and Wang [4] tried to identify a feasible technique for scheduling the university shuttle fleet, considering students' travel patterns and class periods. The annual ridership data for 2018–2019 of the Texas A&M University (TAMU) shuttle system were utilized to detect unique weekday and weekend travel patterns. A clustering strategy was presented to create uniform departure times within the current schedules for on-campus and off-campus routes, considering the high student loads and graduate/undergraduate habits. [5] studied a case for the capacitated employee bus routing problem using data from a retail company. They constructed a mathematical model to minimize the total distance of bus routes. The number and location of bus stops was identified using k-means and fuzzy c-means clustering techniques. Then, due to the NP-Hard nature of the bus routing problem, simulated annealing (SA) and genetic algorithm (GA)-based solutions were presented. A hybrid iterative local search (ILS) metaheuristic algorithm for SBRP with various planning scenarios, such as homogeneous or heterogeneous fleets, single load or mixed load modes of operation, is an attempt made in [6]. Pérez, Ansola [7] created a mathematical model based on literature and modified it such that metaheuristics could be used to solve it.

Incorporating school bell times as decision variables rather than input data allowed them to extend the traditional school bus routing problem [8]. The authors used the Memetic algorithm to solve the proposed model and argue that changing school working hours (bell ringing times) is especially helpful in reducing transportation costs. Calvete, Galé [9] consider the SBRP if students are free to select the bus stop of their choice. It is created and presented to convert a two-stage optimization model to a single-level mixed integer linear programming (MILP) model. To address the issue, a straightforward and efficient metaheuristic algorithm has been created.

Pérez, Sánchez-Ansola [10] introduced a novel method for reducing execution time, which is one of the factors considered while solving a problem using SBRP. Hou, Liu [11] proposed a local search-based metaheuristic algorithm framework for SBRP based on the analysis of SBRP problem models. The proposed framework provides metaheuristic development for the Capacitated Vehicle Routing Problem by reducing or modifying the constraints of SBRP. Experimental results of several examples show that the metaheuristic algorithm based on the proposed approach can be implemented quickly and applied to different SBRP applications.

Feng, Zhang [12] proposed two strategies (mixed-load and single-load) for the mixed-load school bus routing problem (SBRP) to improve student commuting efficiency while reducing operational costs. A bi-level programming model was constructed to maximize school accessibility and minimize the number of school buses. The entropy-based weight method was used to evaluate the strategies, and case studies were conducted using school bus data in Dalian, China. [13] addressed SBRP with student input in bus stop selection. A bilevel optimization model with multiple followers was proposed and transformed into a single-level mixed integer linear programming model. A metaheuristic algorithm was developed to solve the problem, which involved solving four MILP problems to obtain upper bounds. Computational experiments using benchmark instances showed the algorithm was effective in terms of solution quality and required computing time.

In this study, the School Bus Routing Problem (SBRP) was addressed in a two-stage approach. Initially, the Fuzzy C Means (FCM) method was utilized to identify the optimal locations for bus stops. Subsequently, students were assigned to the nearest bus stop. Alongside FCM, the K-means algorithm was also employed to pinpoint the optimal bus stop locations. In the second stage, the Genetic Algorithm (GA) was employed to generate the most efficient vehicle route with the goal of minimizing the total distance traveled by all vehicles.

The remainder of this paper is organized as follows: In the second section, the literature review is presented. In the third section, the proposed approaches are introduced. The fourth part shows experimental results to compare the performances of the proposed approaches. Finally, the discussion and conclusions are presented.

MATERIALS AND METHODS

This section presents the materials and methods used in this study to solve the SBRP problem. The first part introduces the K Means algorithm used to identify the bus stop locations, while the second part explains the Fuzzy C-means (FCM) algorithm used as an alternative method. Additionally, the Genetic Algorithm (GA) is described to be utilized in the second stage.

K-Means Algorithm

The K-means method is an iterative technique that attempts to divide the dataset into K unique, non-overlapping

subgroups (clusters), where each data point belongs to only one group. It attempts to make intra-cluster data points as comparable as feasible while maintaining clusters as distinct as possible. It allocates data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (the arithmetic mean of all data points belonging to that cluster) is the smallest possible value. Data points within the same cluster are more homogenous (similar) the less variance there is within clusters. The implementation of the K-means algorithm is as follows ([14]:

Step 1: Indicate the number of cluster K.

Step 2: Initialize centroids by first shuffling the dataset and then picking K data points at random without replacement for use as centroids.

Step 3: Repeat iterating until the centroids do not change. Thus, the assignment of data points to clusters remains unchanged.

Fuzzy C Means (FCM)

The fuzzy C-Means method is a variant of the K-Means algorithm that employs fuzzy logic and seeks to find the least value of the objective function. It is possible for a data sample to belong to more than one cluster at the same time in FCM. Similarity is indicated by the membership value. In FCM, a membership value and a data sample are assigned based on cluster center and similarity ([15]. Membership values range from 0 to 1, and the higher the similarity, the higher the membership value. Steps of the Fuzzy C-Means algorithm are given in Equations 1-3 ([15].

$$O_\mu(M,y) = \sum_{i=1}^{c}\sum_{k=1}^{N}(m_{ik}^{\mu})\|x_i - y_k\|^2 \qquad (1)$$

where y_j is the mean for that points over cluster i, $\|x_i - y_k\|^2$ is a squared inner product distance norm, and m_{ik}^{μ} represents the fuzzy partitions, where m_{ik} denotes the membership degree that the ith data point belongs to the kth cluster and μ is a scalar for controlling the fuzziness of the resulting clusters. Besides, x represents the geographical X and Y coordinates.

Step 1: Assign membership matrix $M^{(0)} = [m_{ik}]$.

Step 2: Compute cluster centers using Equation 2.

$$y_k = \frac{\sum_{i=1}^{n}(m_{ik})^{\mu}x_i}{\sum_{i=1}^{n}(m_{ik})^{\mu}} \qquad (2)$$

Step 3: Arrange membership values using Equation 3.

$$m_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\left|\frac{x_i - y_k}{x_i - c_j}\right|\right)^{\frac{2}{\mu-1}}} \qquad (3)$$

Step 4: If the differences in total costs estimated by consecutive membership matrices are less than the stopping criteria, the algorithm terminates.

In those equations, k and N represent the number of clusters and students, respectively. x_i means the coordinates of the student *i*, and y_k means the coordinates of the cluster *k*.

Genetic Algorithm (GA)

Genetic algorithms are stochastic algorithms whose search techniques are based on biological genetics and natural evolution concepts. Holland first presented the fundamental ideas of genetic algorithms in 1975. Individuals in a population are represented as chromosomes in this methodology, and a sequence of genetic operations is performed. It is widely acknowledged that the possible solution to each problem is an individual who may be represented using a set of parameters. Every cell in nature contains chromosomes that make up the DNA of that cell. Chromosomes contain the genes that regulate a living organism's character differentiation. An individual or chromosome is the potential solution to a problem in a genetic algorithm. The population is made up of individuals. Population, on the other hand, refers to the sum of the genetic algorithm's solutions. Mating, mutation, and reproduction are all genetic processes that are employed in genetic algorithms. Genetic algorithms offer an ideal setting for tackling large-scale optimization challenges. Although it belongs to the class of probabilistic algorithms, it differs from random number generation methods. Genetic operators, such as mutation, can readily escape local optimal points in genetic algorithms ([16]. The details of the method are not provided in this paper because it is a well-known universal algorithm.

A typical GA focuses on two things: fitness function and solutions. Constraints cannot be defined as in linear programming ([17]. As a result, the solution space and fitness function must be suitably adjusted. The pseudocode of GA to solve the proposed SBRP is presented in Table 1.

Table 1. Pseudo code of the applied metaheuristic algorithm for proposed SBRP.

Stage 0:

Obtain student's coordinates P_i

Cap: capacity of the vehicle

Stage 1:

Find the cluster center with FCM

Assign to C_x and C_y

Stage 2:

$x_{values} \leftarrow random\ (0,10) * n$

Generate x_{values} by GA

Stage 3:

Calculate distance between the $distance_k = \{P_i, C_i\}$

$$\sum_{student} walking\ distance\ for\ each\ student + \sum_{arc} total\ travelled\ distance$$

Cluster-First Route-Second Approach for School Bus Routing Problem

Mathematical model for multi-vehicle SBRP

The aim of the mathematical model is to design a system that will minimize the walking distance of each student to the assigned stops and minimize the total distance traveled by the vehicles. The mathematical model of the proposed SBRP problem is a bi-objective mixed integer location routing model. Even the single-objective location routing problem is an NP-hard problem because it combines the two problems, which are the facility location and vehicle routing problem and the location routing problem ([10]. Because the bi-objective position routing issue is more complicated, it is also NP-hard. The representation of the model, its parameters, and the decision variables are presented below.

Model A:

Table 2. Symbols used in the mathematical model

Indices	
m	indices of the student *(m= 1,2,…M)*
i, j	*indices of the* stops *(i=1,2,3…C_i)*
v	*vehicle*
Parameters	
C_i	*the number of the stop*
P_{mx}	*coordinate value of the m^{th} student on the x-axis*
P_{my}	*coordinate value of the m^{th} student on the y-axis*
W	*the capacity of the vehicle*
Models Variables	
Z_{iv}	*{if the vehicle v visits to i^{th} stop, 1 otherwise, 0*
Y_{im}	*{if m^{th} student assigns to i^{th} stop, 1 otherwise, 0*
d'_{im}	*the distance between i^{th} stop and m^{th} student*
d''_{ij}	*the distance between i^{th} stop and j^{th} stop*
Decision Variables	
C_{ix}	*Coordinate value of the i^{th} stop on the x - axis*
C_{iy}	*Coordinate value of the i^{th} stop on the y - axis*
x_{ijv}	*{if the vehicle v travels from i^{th} stop to j^{th} stop, 1 otherwise, 0*

Model:

$$z = \sum_m \sum_i d'_{im} * Y_{im} + \sum_i \sum_j \sum_v d''_{ij} * x_{ijv} \quad (4)$$

$$d'_{im} = |P_{mx} - C_{ix}| + |P_{my} - C_{iy}| \qquad \forall m, i \quad (5)$$

$$\sum_i Y_{im} = 1 \qquad \forall m \quad (6)$$

$$d''_{ij} = \sqrt{(C_{ix} - C_{jx})^2 + (C_{iy} - C_{jy})^2} \qquad \forall i, j \quad (7)$$

$$\sum_i \sum_v x_{ijv} = 1 \qquad \forall j \quad (8)$$

$$\sum_j \sum_v x_{ijv} = 1 \qquad \forall i \quad (9)$$

$$\sum_{i=0,j=0}^{Ck} x_{ijv} \left(\sum_m Y_{im} \right) = W \qquad \forall v \quad (10)$$

$$\sum_i x_{ijv} = Z_{jv} \qquad \forall j, v \quad (11)$$

$$\sum_j x_{ijv} = Z_{iv} \qquad \forall i, v \quad (12)$$

$$\sum_i x_{ijv} = \sum_h x_{jhv} \qquad \forall i, v \quad (13)$$

$$\sum_i x_{0iv} = \sum_j x_{j0v} = 1 \qquad \forall v \quad (14)$$

Equation 4 expresses the objective function. It minimizes the walking distance of students and the total distance traveled by vehicles. Equation 5 and Equation 7 calculate the walking distance of students and the distance between cluster centers, respectively. Equation 6 checks that each student is assigned to only one cluster. Equations 8–9 state that only one vehicle should be assigned between any two points. Equation 10 verifies that the number of students served by the vehicle along the entire route of each vehicle does not exceed the vehicle's capacity. Equations 11–12 state that a bus arriving at a bus stop also leaves the station and stops only on its route. Equation 11 ensures that the flow of each vehicle is maintained. Equation 14 states that each bus departs from school and returns to school. Equations 15–16 are MTZ equations.

Miller-Tucker-Zemlin (MTZ) formulation uses an extra positive variable called as p_i. p_i takes a value for each node except the school. If a vehicle is traveling from node i to node J, p_j must be greater than p_i. The value of p_i grows with each subsequent call point.

$$p_i - p_j - n * \sum_{v}^{V} x_{ijv} \leq n-1 \qquad 2 \leq i \neq j \leq n \quad (15)$$

$$0 \leq u_i \leq n-1 \qquad 2 \leq i \leq n \quad (16)$$

The process of selecting bus stops and generating bus routes is frequently referred to as a location-routing problem (LRP), which is an example of an NP-hard problem. An NP-hard problem is one for which there are no known polynomial-time algorithms that can provide an optimal solution for every instance of the problem. In such situations, dividing the problem into parts helps facilitate its resolution.

The foundational "divide and conquer" approach served as the primary inspiration for the creation of this heuristic, which was built on that method. An assignment model is first solved to establish a cluster of pickup locations for each vehicle using the cluster-first and route-second approaches. Subsequently, the traveling salesman problem (TSP) is addressed for all clusters. Finally, a route is generated using the TSP solutions.

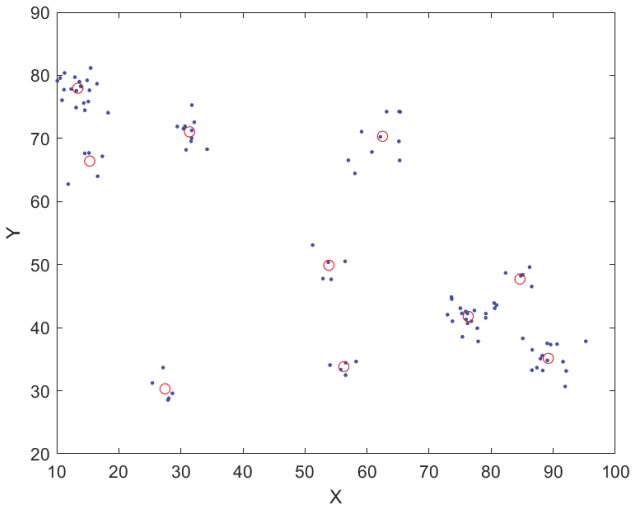


Figure 1. Sample cluster centers for n=10.

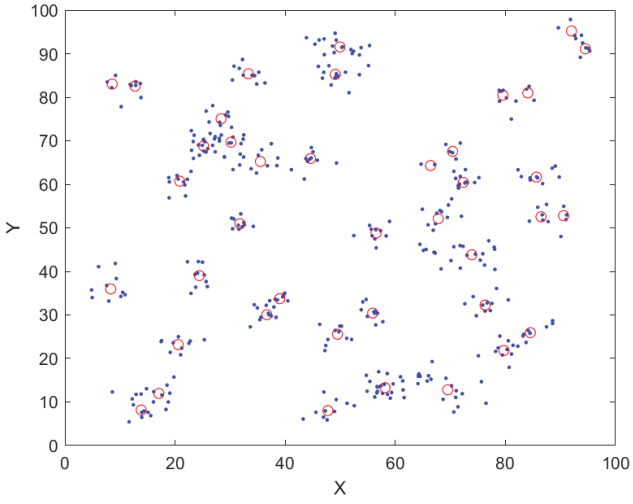


Figure 2. Sample cluster centers for n=40.

The results obtained from the fuzzy C-means algorithm under different conditions are presented in Figures 1 and 2. Figure 1 depicts the cluster centers obtained for a scenario involving 10 bus stops, whereas Figure 2 presents the cluster centers for a scenario with n = 40 bus stops. It is common practice to locate bus stops in parts of a community that have a high concentration of student inhabitants. The decision variables C_{ix} and C_{iy} in Model A may be derived from the outputs that are provided by the FCM. The last variable that must be determined is the one that is referred to as X_{ijv}.

Route Generation

The most expedient route is determined with the help of GA. An initial solution is chosen at random by the genetic algorithm. The resolution that was randomly given is called an initial chromosome. If we assume that the value of each gene (r_i) falls somewhere between 0 and 1, then the composition of a chromosome with 5 stops is as follows:

$$X_{ijv} = (r_1, r_2, \ldots, r_5)$$

Each gene value is sorted from lowest to highest, and the final order of the rankings is kept. This rank indicates the order in which the bus stops are located along the route. When the number of students on the bus that is making the stops exceeds the capacity of the vehicle, another bus is brought into service. There is a limit of one stop per bus when overbooking is permitted.

The calculations are done based on the premise that the school serves as both the beginning and the finish point for the vehicle routes. The fundamental procedures that GA employs to determine the most efficient route are summarized in Table 3.

Table 3. Route *genereation* by GA

1.	Generate the initial solution
2.	Assign the vehicles to each bus stop
3.	Compute the objective function
	3.1. Modify the solution by genetic operators
	3.2. Compute the objective function
4.	Kept the best solution

RESULTS AND DISCUSSION

The dataset originally consisted of 112 cases, all of which were taken from [18]. Sample sizes vary anywhere from 5 stops and 25 students to 80 stops and 800 students. In addition, the dataset took into consideration a maximum

Table 4. Parameters of GA

Parameters	Values
Max_num_iteration	1500
Population_size	100
Mutation_rate	0.1
Elit_ratio	0.01
Crossover_rate	0.5
Parents_portion	0.3
Crossover_type	uniform
Variable_bounds	(0.1)

GA: Genetic algorithm.

of four different maximum walking distances. Since the maximum walking distance was not included as a parameter in this study, there were no experiments conducted to investigate the effects of varying the maximum walking distance. As a result, the experiments were only conducted in a total of 28 instances. The instances have a standard fleet of buses that can accommodate 25 or 50 students.

The parameter values of the employed metaheuristic algorithm are shown in Table 4. The maximum value of the decision variable is set to 1 in GA, while the minimum value is set to 0.

Table 5 contains the results for the problem instances. The number of stops is indicated in the stop column. The *Student* column specifies the number of students served, while the *CAP* column indicates the bus capacity. The students are transported by a homogeneous fleet of vehicles of identical capacity. Total walking distance is shown by the total walking column, while total driving distance is indicated by the total driving distance.

The *Total driving** columns are vital to compare *former_best* and *new_best* value. As an example, the best value is found as 241.68 for Instance 09, which is found in this work; also, this value is obtained as 286.68 in [18] as seen in Table 5. It suggests that the proposed approach improves performance by 15.7%.

Table 5. Results for FCM-GA

				FCM & GA		
ID	**#Stop**	**#Student**	**CAP**	**Total walking**	**Total driving***	**Former best**
1	5	25	25	159.26	266.26	141.01
2	5	25	50	151.88	250.60	161.62
9	5	50	25	254.00	241.68	286.68
10	5	50	50	254.00	229.09	197.20
17	5	100	25	724.46	253.40	360.35
18	5	100	50	600.96	256.95	304.23
25	10	50	25	318.48	251.60	282.12
26	10	50	50	181.51	300.15	244.54
33	10	100	25	464.19	482.80	403.18
34	10	100	50	319.94	286.28	296.53
41	10	200	25	970.05	468.94	735.27
42	10	200	50	983.11	367.40	512.16
49	20	100	25	354.82	478.61	520.24
50	20	100	50	338.29	409.17	420.64
57	20	200	25	409.17	647.29	903.84
58	20	200	50	734.19	489.53	485.65
65	20	400	25	1937.84	738.64	1323.35
66	20	400	50	1937.84	594.74	733.54
73	40	200	25	538.39	794.48	831.94
74	40	200	50	631.57	771.73	593.35
81	40	400	25	1292.30	1185.69	1407.05
82	40	400	50	1333.54	829.79	858.80
89	40	800	25	2607.60	1584.66	2900.14
90	40	800	50	2736.71	1054.22	1345.70
97	80	400	25	1196.71	2571.42	1546.23
98	80	400	50	1038.53	2052.97	1048.56
105	80	800	25	2139.61	2734.03	2527.96
106	80	800	50	2405.05	2376.25	1530.58

CAP: Capacity of the vehicle; FCM: Fuzzy C means; GA: Genetic algorithm.

The purpose of this research was not only to lower the total distance traveled by vehicles but also the total distance traveled by pedestrians. When the first phase of the clustering procedure is complete, it is verified that the students' stopping locations are placed at the geographically closest points to them. This reduces the amount of walking distance these students must travel. In addition, the predefined bus stop locations were used to find the optimal routes

Table 5 displays the results of the suggested approach, indicating that favorable performance was achieved in 17 out of the 28 analyses. It is noteworthy that the "driving" column is of particular significance for the purpose of comparison, while the "walking" column represents an additional accomplishment that we aim to achieve. This is because students are required to walk less concurrently, which has had a notable impact on the identification of bus stop locations. Consequently, the search for the shortest driving distance is complicated by the need to consider the walking distance factor. Despite the multiplicity of focus points, the suggested approach still yielded better results in more than half of the instances.

Table 6 demonstrates the results of the K-means and GA methodology, indicating that this strategy outperformed the suggested approach in 19 of the 28 evaluations. However, it is important to note that as the sample size grows, the two-stage strategy's performance has shown signs of deterioration.

This study aimed to compare the FCM and K-means clustering algorithms, as depicted in Figure 3. The graph illustrates the results obtained from each algorithm, and a detailed analysis of the graph revealed that the K-means

Table 6. Results for K means-GA

K-means & GA						
ID	**#STOP**	**#STUDENT**	**CAP**	**Total walking**	**Total driving***	**Former_Best**
1	5	25	25	137.57	146.05	141.01
2	5	25	50	276.78	157.07	161.62
9	5	50	25	542.70	227.65	286.68
10	5	50	50	387.76	182.66	197.20
17	5	100	25	1096.30	262.52	360.35
18	5	100	50	534.27	253.06	304.23
25	10	50	25	421.21	225.63	282.12
26	10	50	50	301.48	281.13	244.54
33	10	100	25	429.43	280.73	403.18
34	10	100	50	433.37	271.95	296.53
41	10	200	25	1278.26	422.71	735.27
42	10	200	50	1570.33	375.02	512.16
49	20	100	25	415.70	463.32	520.24
50	20	100	50	434.92	417.95	420.64
57	20	200	25	1266.32	625.01	903.84
58	20	200	50	740.33	473.40	485.65
65	20	400	25	1829.96	740.13	1323.35
66	20	400	50	1723.89	499.52	733.54
73	40	200	25	853.92	906.26	831.94
74	40	200	50	600.75	702.89	593.35
81	40	400	25	1198.97	1084.13	1407.05
82	40	400	50	1463.16	926.50	858.80
89	40	800	25	3065.63	1612.96	2900.14
90	40	800	50	2810.92	1183.00	1345.70
97	80	400	25	1197.36	2485.06	1546.23
98	80	400	50	1059.77	2440.06	1048.56
105	80	800	25	2244.93	2876.94	2527.96
106	80	800	50	2488.89	2360.54	1530.58

CAP: Capacity of the vehicle; GA: Genetic algorithm.

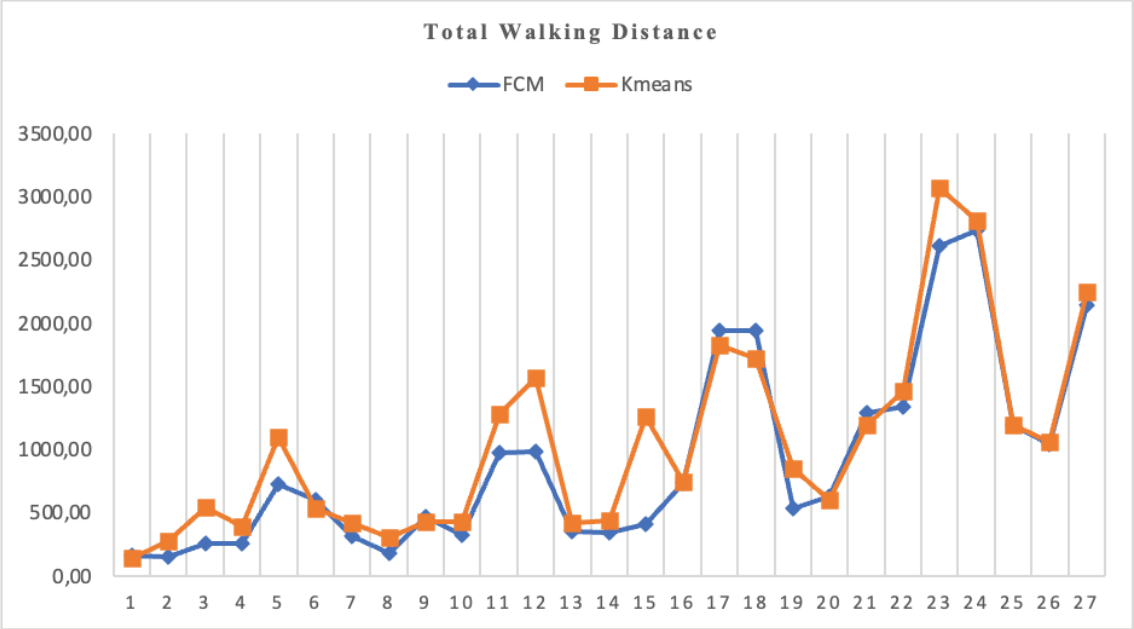


Figure 3. Comparison of the total walking distances.

Table 7. Statistical summary of the proposed approaches

	FCM-GA	K-Means-GA	Former_best
Mean	820.2983	817.2809	817.945
Median	486.1672	468.3602	556.795
Min	229.0923	146.0465	141.01
Max	2734.026	2876.941	2900.14

FCM: Fuzzy C means; GA: Genetic algorithm.

approach demonstrated superior efficiency in reducing the walking distance for students.

Furthermore, Table 7 provides a comprehensive statistical summary of the proposed approaches. The K-Means-GA approach yielded the best mean value of 817.28, surpassing the previous results reported in the literature which stood at 817.945. However, given that the mean alone may not suffice for comparison purposes, the median value was also considered. Remarkably, the K-Means-GA approach also resulted in the smallest median value.

In summary, the findings of this study provide evidence that the K-means clustering algorithm, particularly the K-Means-GA approach, is a promising method for optimizing bus stop locations and reducing walking distances for students.

CONCLUSION

This paper presents a two-stage approach to the school bus routing problem, with the aim of reducing both the total distance traveled by vehicles and students. The first stage involves selecting suitable bus stops using the FCM and K-Means algorithms, while the second stage uses GA to plan bus routes. The dataset utilized in this study was obtained from an easily accessible source. The application process was conducted in a total of 28 distinct instances, wherein both the Fuzzy C Means and the GA implementations were employed to attain optimal outcomes. Analysis of the experimental results revealed that the proposed approach yielded significant advantages overall.

The proposed approach offers decision makers possible solutions and trade-offs between the objectives. The study shows that the proposed approach offers significant benefits, including a considerable reduction in solution time. However, there are some limitations, such as the lack of coordination across phases' objective functions.

A brief summary of the fundamental points outlined in this study is presented below.:

- The proposed two-stage approach reduces solution time and enables modifications at any stage.

- Overbooking is a common occurrence in real-life scenarios, but it is rarely discussed in the SBRP literature.
- The unchangeable rule that prohibits exceeding the maximum capacity of a vehicle's seating capacity has been loosened in this study.
- The first bus stop locations are uncertain, and there is no available sequence of potential bus stops.

In future studies, it may be beneficial to investigate a method that integrates both objective functions into a single phase, potentially enhancing the efficiency and effectiveness of the solution. However, it is important to note that the utilization of a goal programming approach in a one-step process may result in an increased level of complexity in the problem formulation and solution procedure. Thus, the aim of any such approach should be to provide reasonable and convincing results in a practical and feasible manner.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Miranda DM, de Camargo RS, Conceição SV, Porto MF, Nunes NTR. A multi-loading school bus routing problem. Exp Syst Appl 2018;101:228–242. [CrossRef]

[2] Oluwadare SA, Oguntuyi IP, Nwaiwu JC. Solving school bus routing problem using genetic algorithm-based model. Int J Intell Syst Appl 2018;12:50. [CrossRef]

[3] Sun S, Duan Z, Xu Q. School bus routing problem in the stochastic and time-dependent transportation network. PloS One 2018;13:e0202618. [CrossRef]

[4] Mahmoudzadeh A, Wang XB. *Cluster Based Methodology for Scheduling a University Shuttle System*. Trans Res Record 2020;2674:236–248. [CrossRef]

[5] Kiriş SB, Özcan T. Metaheuristics approaches to solve the employee bus routing problem with clustering-based bus stop selection, in Artificial Intelligence and Machine Learning Applications in Civil, Mechanical, and Industrial Engineering, 2020, p. 217–239, IGI Global. [CrossRef]

[6] Xie Y, Kong Y, Xiang H, Hou Y, Han D. A metaheuristic with learning mechanism for solving the multi-school heterogeneous school bus routing problem. IAENG Int J Comput Sci 2021;48:1–9.

[7] Pérez ACP, Ansola ES, Suárez AR. A metaheuristic solution for the school bus routing problem with homogeneous fleet and bus stop selection. Ingeniería 2021;2:233–253. [CrossRef]

[8] Miranda DM, de Camargo RS, Conceição SV, Porto MF, Nunes NTR. A metaheuristic for the rural school bus routing problem with bell adjustment. Exp Syst Appl 2021;180:115086. [CrossRef]

[9] Calvete HI, Gale C, Iranzo JA, Toth P. The school bus routing problem with student choice: a bilevel approach and a simple and effective metaheuristic. Int Trans Oper Res 2021;30:1092–1119. [CrossRef]

[10] Pérez AC, Sánchez-Ansola E, Rosete A, Rojas O, Sosa-Gómez G. A partial evaluation approach for the school bus routing problem. Heliyon, 2022;8:e09291. [CrossRef]

[11] Hou Y, Liu B, Dang L He W, Gu W. A local search-based metaheuristic algorithm framework for the school bus routing problem. Eng Lett 2022;30:1–12.

[12] Feng R, Zhang J, Wu Y, Wu R, Yao B. School accessibility evaluation under mixed-load school bus routing problem strategies. Trans Policy, 2023;131:75–86. [CrossRef]

[13] Calvete HI, Carmen G, Iranzo JA, Toth P. The school bus routing problem with student choice: a bilevel approach and a simple and effective metaheuristic. Int Trans Oper Res 2023;30:1092–1119. [CrossRef]

[14] Sinaga KP, Yang M-S. Unsupervised K-means clustering algorithm. IEEE Access 2020;8:80716–80727. [CrossRef]

[15] Küçükdeniz T, Erkal Sönmez Ö. Integrated Warehouse Layout Planning with Fuzzy C-Means Clustering. in International Conference on Intelligent and Fuzzy Systems. 2022. Springer. [CrossRef]

[16] Pekel Özmen E, Özcan T. Diagnosis of diabetes mellitus using artificial neural network and classification and regression tree optimized with genetic algorithm. J Forecast 2020;39:661–670. [CrossRef]

[17] Abubakar AM, Francis OC, Sarkinbaka ZM, Sabo Yahaya M. Simplex C++ syntax for solving chemical engineering cost optimization problems. Res Invent Int J of Eng Sci 2021;11:39–47. [CrossRef]

[18] Schittekat P, Kinable J, Sörensen K, Sevaux M, Spieksma F, Springael J. A metaheuristic for the school bus routing problem with bus stop selection. Eur J Oper Res 2013;229:518–528. [CrossRef]

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1039–1049, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2023.00069

Research Article

Analysis of grover's quantum search algorithm on a classical computer: Identifying opportunities for improvement

Necati ÇELİK[1,*], Özkan BİNGÖL[2]

[1]*Department of Physics Engineering, Gümüşhane University, Gümüşhane, Türkiye*
[2]*Department of Software Engineering, Gümüşhane University, Gümüşhane, Türkiye*

ARTICLE INFO

Article history
Received: 04 February 2023
Revised: 25 April 2023
Accepted: 30 May 2023

Keywords:
Grover's Search Algorithm; Quantum Information; Quantum Information Processing; Qubit, State Vector

ABSTRACT

In this paper, Grover's quantum search algorithm is analyzed using a classical computer by calculating the amplitudes and the probabilities of finding a single marked state for n=5, 10, 15, 20, 25, and 27 qubit states. The calculations show that the marked state can be found in $\mathcal{O}(\sqrt{N})$ iterations, where $N = 2^n$ is the number of items. The possibility of improving Grover's search algorithm to find a single item in N search elements is discussed by calculating the amplitudes and hence the probabilities of finding a single marked state for n=5, 10, 15, 20, 25, 30, 35, 40, 45, and 50 qubit states. The calculations showed that the marked state could be found with sufficiently high probability in $\mathcal{O}(\ln(N))$ iterations. This is quite a remarkable speed-up that can be achieved to find a single marked element in an unsorted N search element.

Cite this article as: Çelik N, Bingöl Ö. Analysis of grover's quantum search algorithm on a classical computer: Identifying opportunities for improvement. Sigma J Eng Nat Sci 2024;42(4):1039–1049.

INTRODUCTION

Grover's search algorithm [1] provides a relatively fast way of finding a marked element in an unsorted database of size N in the computational complexity $\mathcal{O}(\sqrt{N})$ offered by quantum computers with respect to its classical counterpart, which can only solve the problem in complexity $\mathcal{O}(N)$. This means less memory requirement to represent the input data. The speed-up is a good example of realizing the computational power of quantum computers [2]. This power comes from the laws of quantum mechanics, which are fundamentally different from classical ones. Grover's search algorithm can perform multiple computations simultaneously due to the quantum parallelism. Amplitude of the marked state is amplified at each iteration therefore when measured, the probability of finding the marked element is sufficiently high to be identified. Grover's algorithm is a black-box algorithm, which means that it does not rely on any specific properties of the input data or the function being evaluated. It only requires access to the function that evaluates the input data.

There are numerous useful classical algorithms solving different problems such as generic algorithm to solve second-order boundary value problems [3], for prediction of linear dynamical systems [4], for optimization solution of Troesch's and Bratu's problems [5], and solving singular

*Corresponding author.
*E-mail address: necati.celik@gumushane.edu.tr

This paper was recommended for publication in revised form by Editor in Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey

two-point boundary value problems [6], control of axially moving viscoelastic strip [7], adaptive stabilization of a Kirchoff moving strip [8], and existence and stabilization of a Kirchoff moving strip [9]. Apart from classical ones, there are other quantum algorithms, which solve a given problem more efficiently than corresponding classical algorithms: Shor's algorithm [10,11] solves the problem of factorizing large numbers, which is infeasible on a classical computer. The other one, which should be mentioned here, is the Deusch-Jozsa algorithm [12-14] which finds the character of the boolean function exponentially faster than a classical computer.

After being theoretically proposed, Grover's search algorithm has been implemented in many physical systems. These physical systems include nuclear magnetic resonance (NMR) [15, 16], trapped ions [17-20], superconducting qubits in the quantum electrodynamics QED [21-23], and cavity QED [24-26].

Since it is proposed, the mathematical structure of Grover's search algorithm has been investigated [27, 28] and the complexity analysis was studied [29] and applied in many areas such as [30] in which a new protocol for quantum private query is proposed, which achieves efficiency by utilizing two oracle operations to encode both encryption keys and encrypted data items into corresponding quantum superposition states. The protocol also introduces the use of Grover iteration to extract the target from the superposition state. As well known, quantum computers operate upon a register of qubits, which are the quantum mechanical version of classical bits. A qubit can be described as the superposition of 0 and 1, while a classical bit is confined to a definite value of either one of these numbers. That is, describing the state of an n-bit classical information scale as the number of bits. Thus, the calculation time and the memory resources, which are necessary for classical computation, are generally a polynomial function. However, describing the state of n-qubit information requires 2n complex numbers. Therefore, simulating a quantum computer using a classical computer is exponentially complex with respect to the number of qubits. Although classical simulation of quantum computation is exponentially costly, it is vital for the study of new quantum algorithms and architectures. Classical simulation is also important to investigate the efficiencies and robustness of quantum algorithms as compared to their classical counterparts.

To our best knowledge, there is no quantitative analysis of Grover's search algorithm to show how the amplitudes of marked and unmarked states change at every iteration. This is important to see the evolution of the state vector at every iteration and to be able to discuss the possible improvement of the algorithm to obtain a better speed-up. Therefore, the main objective of the current study is to analyze Grover's search algorithm on a classical computer for various numbers of qubits and try to find possible ways to improve it for a better speed-up in finding marked elements in a certain search space.

MATERIALS AND METHODS

Grover's search algorithm proposes to find a particular element through a search space of N elements. Rather than searching for the element directly, one can concentrate on the index to those elements which is a number ranging from 0 to N-1. Defining N=2n, one can store the index in n bits. A particular instance of the search problem can be represented by a function f taking an integer x as an input in the range 0 to N-1. By definition, f(x)=1 if x is a solution to the search problem which is denoted by x*, and if x is not the solution to the search problem, then f(x)=0.

Grover's search algorithm consists of two steps. The first step is preparing the initial qubits in an equal superposition state by applying the Hadamard gates [31]. The second step is applying the Grover iteration $\mathcal{O}(\sqrt{N})$ times. The Grover iteration itself consists of two steps; running the quantum oracle followed by the diffusion operator. After any iteration, while the amplitude of the marked state grows, all other amplitudes shrink [32].

The recognition of the searched element is achieved by making use of an oracle, which is a unitary operator. The action of the oracle is defined in Equation (1).

$$|x\rangle|q\rangle \xrightarrow{Oracle} |x\rangle|q \oplus f(x)\rangle \tag{1}$$

where $|x\rangle$ is index register, $\oplus$ denotes addition module 2, and $|q\rangle$ is called oracle qubit, which is a single qubit and is flipped if f(x)=1, and is unchanged if f(x)=0. It is useful to put the oracle qubit initially in the state $(|0\rangle - |1\rangle)/\sqrt{2}$. If x is not a solution to the search problem, applying the oracle to $|x\rangle(|0\rangle - |1\rangle)/\sqrt{2}$ does not do anything. But if x is a solution, then $|0\rangle$ and $|1\rangle$ are interchanged after the action of the oracle, giving a final state $-|x\rangle(|0\rangle - |1\rangle)/\sqrt{2}$. Thus, the action of the oracle can be summarized in Equation (2).

$$|x\rangle\left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right] \xrightarrow{Oracle} (-1)^{f(x)}|x\rangle\left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right] \tag{2}$$

Schematic circuit for the quantum search algorithm is depicted in Figure 1.

The schematic circuit shown in Figure 1 consists of n qubits prepared in the $|0\rangle$ state and one oracle qubit prepared in the $|1\rangle$ state. Both of these states pass through the Hadamard gate before a series of Grover iteration operators are applied. Then by measuring the first n-qubits the desired result can be obtained. In the following, we give a detailed description of the circuit by writing the state vectors at every step through the circuit. In the circuit, the first input is called the index register which consists of the state of n qubits all prepared in the $|0\rangle$ state. The second input is called the oracle qubit prepared in the $|1\rangle$ state. Therefore $|\psi_0\rangle$ can be written as in Equation (3).

$$|\psi_0\rangle = |0^{\otimes n}\rangle|1\rangle \tag{3}$$

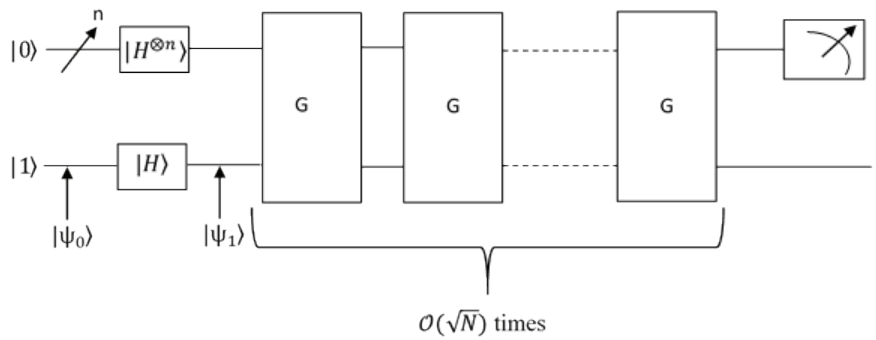


Figure 1. Schematic circuit for the quantum search algorithm.

After the Hadamard transform on the index register and the Hadamard transform on the oracle qubit, we have $|\psi_1\rangle$ as given in Equation (4).

$$|\psi_1\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} |x\rangle \left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right] \tag{4}$$

In $|\psi_1\rangle$, the Hadamard transform puts the index register in a superposition of all values which is also called equal superposition state and the oracle qubit is in an evenly weighted superposition of 0 and 1. After that, the circuit consists of repeated application of the Grover iteration which is a quantum subroutine, and applied $\mathcal{O}(\sqrt{2^n})$ times. Figure 2 shows the schematic diagram for the Grover iteration circuit. We can now follow what happens to the state vector from Figure 2.

After the action of the oracle, the state vector takes the form given in Equation (5).

$$|\psi_2\rangle = \frac{1}{\sqrt{N}} \sum_{x\in\{0,1\}^n} (-1)^{f(x)} |x\rangle \left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right] \tag{5}$$

The action of the oracle is to change the phase of the solution state, which we can also call the marked state $|x^*\rangle$, and separate it from all other states. Thus, a unitary operator as given in Equation (6) can represent the oracle.

$$U_f = I - 2|x^*\rangle\langle x^*| \tag{6}$$

Thus, $|\psi_2\rangle$ can also be written as in Equation (7).

$$|\psi_2\rangle = (I - 2|x^*\rangle\langle x^*|)|S\rangle \left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right] \tag{7}$$

where $|S\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} |x\rangle$ called standard state.

Since the oracle qubit remains $(|0\rangle - |1\rangle)/\sqrt{2}$ throughout the calculations, it can be omitted from further calculations, therefore $|\psi_2\rangle$ can simply be written as in Equation (8).

$$|\psi_2\rangle = (I - 2|x^*\rangle\langle x^*|)|S\rangle \tag{8}$$

After the action of the oracle, the marked state is separated but still, all states have equal amplitudes. Now as the amplitude of the marked state must be increased, while all other amplitudes have to be shrunk so that when measured, the probability of finding the marked state should be the highest. The action of amplitude change can be achieved through the three steps: Hadamard transform

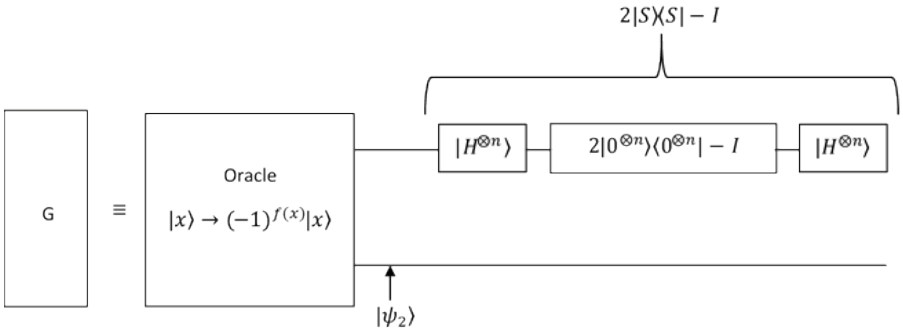


Figure 2. Schematic circuit for the Grover iteration, G.

$H^{\otimes n}$, performing a conditional phase shift with every computational basis except $|0\rangle$ receiving a phase shift of -1, and applying the Hadamard transform $H^{\otimes n}$ again which can be represented by the operator as given in Equation (9) and shown in Figure 2.

$$H^{\otimes n}(2|0^{\otimes n}\rangle\langle 0^{\otimes n}| - I)H^{\otimes n} = 2|S\rangle\langle S| - I = U_s \quad (9)$$

Thus, the Grover iteration may be written as in Equation (10).

$$G = (2|S\rangle\langle S| - I)(I - 2|x^*\rangle\langle x^*|) \quad (10)$$

RESULTS AND DISCUSSION

Analysis of Grover's Search Algorithm

The equal superposition state, $|S\rangle = \frac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}^n}|x\rangle$, can be written as the sum of two orthogonal states as given in Equation (11) and shown geometrically in Figure 3.

$$|S\rangle = \frac{\sqrt{N-1}}{\sqrt{N}}|X\rangle + \frac{1}{\sqrt{N}}|x^*\rangle \quad (11)$$

where N is the number of items which is 2n, and n is the number of qubits, $|X\rangle$ and $|x^*\rangle$ are computational basis and $|x^*\rangle$ is marked state and $|X\rangle$ represent all other unmarked states. The evolution of the state vector $|S\rangle$ is calculated by applying Grover's operator G given in Equation (10). Here U_f is called the phase operator which changes the phase of the marked state performed by the oracle and U_s is called the diffuser which increases the amplitude of the marked state while shrinking the amplitudes of the unmarked states. Probabilities of finding marked and unmarked states are calculated by taking the mod-square of the amplitudes of $|x^*\rangle$ and $|X\rangle$ in Eq. (11), respectively.

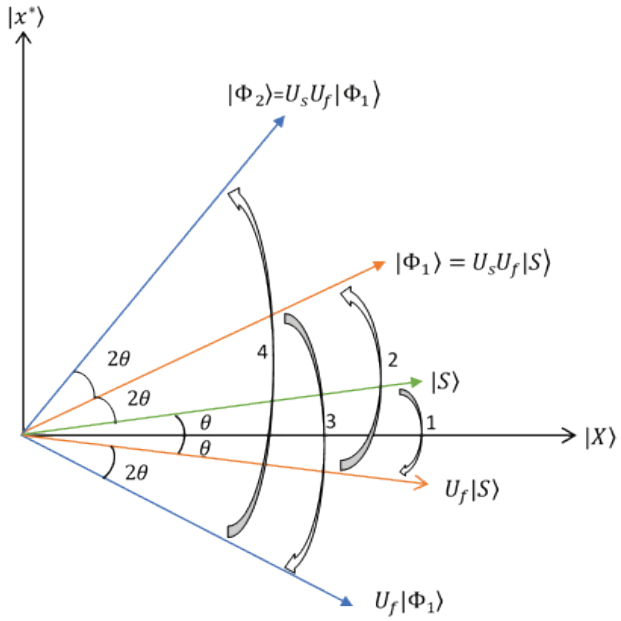


Figure 3. Evolution of equal superposition state $|S\rangle$ in Grover's method.

The number of iterations has to be known in advance so that when measured, the amplitude of the marked state is sufficiently big and can be distinguished from the unmarked states. As seen from Figure 3, at every iteration, the state vector $|S\rangle$ moves away from the state $|X\rangle$ and approaches $|x^*\rangle$ according to $(2m+1)\theta$, where m is the number of iterations. Eventually, the angle between $|S\rangle$ and $|X\rangle$ approaches 90 degrees as given in Equation (12).

$$(2m+1)\theta = \frac{\pi}{2} \quad (12)$$

$m \approx \frac{\pi}{4}\sqrt{N} - \frac{1}{2}$, which is $\mathcal{O}(\sqrt{N})$.

We calculated the evolution of the state vector $|S\rangle$ as given in Equation (13).

$$|\Phi_m\rangle = G|\Phi_{m-1}\rangle \quad ()$$

where $1 \le m \le \frac{\pi}{4}\sqrt{N}$ and $|\Phi_0\rangle = |S\rangle = \frac{b}{a}|X\rangle + \frac{1}{a}|x^*\rangle$, $a = \sqrt{N}$ and $b = \sqrt{N-1}$

Some of the calculated state vectors are given in Equations (14-16):

$$|\Phi_1\rangle = \frac{b}{a^3}(2b^2 - 2 - a^2)|X\rangle + \frac{1}{a^3}(2b^2 - 2 + a^2)|x^*\rangle \quad (14)$$

Letting $(2b^2 - 2 - a^2) = c$ and $(2b^2 - 2 + a^2) = d$,

$$|\Phi_2\rangle = \frac{b}{a^5}(2b^2c - 2d - a^2c)|X\rangle + \frac{1}{a^5}(2b^2c - 2d + a^2c)|x^*\rangle \quad (15)$$

Again letting $(2b^2c - 2d - a^2c) = e$ and $(2b^2c - 2d + a^2c) = f$

$$|\Phi_3\rangle = \frac{b}{a^7}(2b^2e - 2f - a^2e)|X\rangle + \frac{1}{a^7}(2b^2e - 2f + a^2f)|x^*\rangle \quad (16)$$

Considering all these steps, the general formula that calculates each $|\Phi_{i+1}\rangle$ vector iteratively is given in Equation (17).

$$|\Phi_{i+1}\rangle = \frac{b}{a^{(2i+1)}}c_i|X\rangle + \frac{1}{a^{(2i+1)}}d_i|x^*\rangle \quad (17)$$

Here, c_i and d_i are the generalizations of all c and d values.

$$c_{i+1} = 2b^2c_i - 2d_i - a^2c_i \quad (18)$$

$$d_{i+1} \leftarrow 2b^2c_i - 2d_i - a^2d_i \quad (19)$$

To calculate the amplitudes, a function suitable for the pseudo-code given in Algorithm-1 was created.

Algorithm 1. Pseudo Code of Grover's Search Algorithm

Inputs : n – number of qubits
n_step – number of iterations
Outputs : A[] – amplitudes of unmarked states
B[] – amplitudes of marked state
Begin
Set $N \leftarrow 2^n$, % number of search space
Set $a \leftarrow \sqrt{N}$
Set $b \leftarrow \sqrt{N-1}$
Set $c_0 \leftarrow 1$
Set $d_0 \leftarrow 1$
Set $i \leftarrow 0$
while $i < n\_step$ **do**
$A_i \leftarrow \frac{b.c_i}{a^{(2i+1)}}$
$B_i \leftarrow \frac{d_i}{a^{(2i+1)}}$
$|\Phi_{i+1}\rangle \leftarrow A_i|X\rangle + B_i|x^*\rangle$
$c_{i+1} \leftarrow 2b^2c_i - 2d_i - a^2c_i$
$d_{i+1} \leftarrow 2b^2c_i - 2d_i - a^2d_i$
$i \leftarrow i + 1$
end while
End

Depending on the number of qubits, a maximum number of 10000 iterations were analyzed and the amplitudes (and probabilities) were calculated and shown in Figure (a-f) 4 for n=5, 10, 15, 20, 25, and 27 qubits. For n=2 qubit state (N=22=4 search space), only one iteration is required to find the marked state with probability one (not shown in Figure 4).

As the number of qubits increases, the number of iterations also increases smoothly according to $\sim\frac{\pi}{4}\sqrt{N}$. Figure 4 shows that as the number of iterations increases, while the probability of finding the marked state increases, the probability of finding the unmarked states decreases. The number of iterations required to find the marked state with sufficiently high probability for n=5, 10, 15, 20, 15, and 27 qubits is 4, 25, 142, 804, 4549, and 9099 respectively as shown in Table 1.

The Analysis of Proposed Method

The equal superposition state can be given in Equation (20).

$$|S\rangle = \frac{b}{a}|X\rangle + \frac{1}{a}|x^*\rangle \tag{20}$$

where, $a = \sqrt{N}$ and $b = \sqrt{N-1}$

The phase-changing operator is the same as in the case of Grover's method and given as $U_f = I - 2|x^*\rangle\langle x^*|$. In the proposed method, as different from Grover's method, the diffuser operator changes in every iteration, which is given in Equation (21).

$$U_{s_m} = 2|\Phi_m\rangle\langle\Phi_m| - I \tag{21}$$

where $|\Phi_m\rangle$ is the new state vector after $G_m = U_{s_m}U_s$ operator acts, where $1 \le m \le \mathcal{O}(\ln(N))$. After each Grover operator, the change in the state vectors is shown in Figure 5.

Note that the main difference between Grover's method and the proposed method is that the diffuser operator changes at every iteration. In Grover's method, the diffuser operator is $2|S\rangle\langle S| - I$ and stays unchanged throughout the search process. In the proposed method, however it consists of the final state vector at the end of each iteration and the unit operator. Some of the state vectors after the action of Grover's operator are given in Equations (22-24).

$$|\Phi_1\rangle = \frac{b}{a^3}(2b^2 - 2 - a^2)|X\rangle + \frac{1}{a^3}(2b^2 - 2 + a^2)|x^*\rangle \tag{22}$$

Letting $(2b^2 - 2 - a^2) = c_1$ and $(2b^2 - 2 + a^2) = d_1$,

$$|\Phi_2\rangle = \frac{bc}{a^9}(2b^2c_1^2 - 2d_1^2 - a^6)|X\rangle + \frac{1}{a^9}(2b^2c_1^2 - 2d_1^2 + a^6)|x^*\rangle \tag{23}$$

Again letting $(2b^2c_1^2 - 2d_1^2 - a^6) = c_2$ and $(2b^2c_1^2 - 2d_1^2 + a^6) = d_2$

$$|\Phi_3\rangle = \frac{bc_1c_2}{a^{27}}(2b^2c_1^2c_2^2 - 2d_1^2d_2^2 - a^{18})|X\rangle + \frac{1}{a^{27}}(2b^2c_1^2c_2^2 - 2d_1^2d_2^2 + a^{18})|x^*\rangle \tag{24}$$

The general formula that calculates each $|\Phi_{i+1}\rangle$ vector iteratively is given in Equation (25).

$$|\Phi_{i+1}\rangle = \frac{b}{a^{3^{i+1}}}\left(\prod_{j=1}^{i} c_j\right)\left(T_{i+1} - a^{2.3^i}\right)|X\rangle + \frac{1}{a^{3^{i+1}}}\left(T_{i+1} + a^{2.3^i}\right)|x^*\rangle \tag{25}$$

Here T_{i+1} represents a common operation and is specified in Equation (26).

$$|\mathrm{T}_{i+1}\rangle = 2b^2\prod_{j=1}^{i} c_j^2 - 2\prod_{j=1}^{i} d_j^2 \tag{26}$$

Similarly, ci and di are generalized values of all c and d and can be formulated as

$$c_{i+1} \leftarrow 2b^2c_i^2 - 2d_i^2 - a^{2.3^i} \tag{27}$$

$$d_{i+1} \leftarrow 2b^2c_i^2 - 2d_i^2 + a^{2.3^i} \tag{28}$$

To calculate the amplitude in the standard state of the proposed method, a piece of code is written as in Algorithm 2.

As a characteristic future of Grover's search algorithm, the number of iterations has to be known prior to the

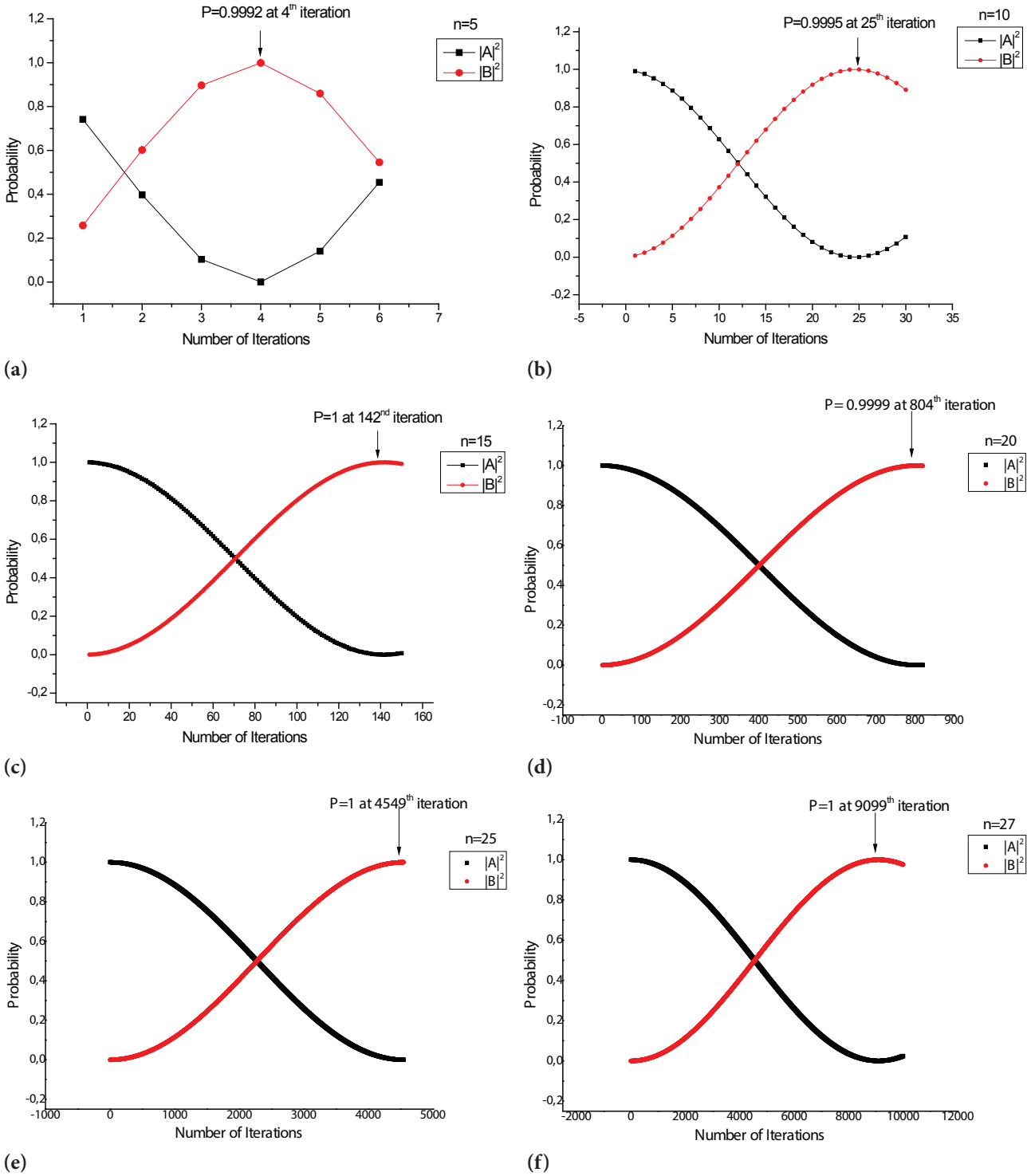


Figure 4. Probabilities finding the marked state and unmarked states.

calculations. The complexity of the proposed method is analyzed as follows. As seen in Figure 5, in every iteration, the state vector approaches the marked state according to Equation (29).

$$3^m\theta \approx \frac{\pi}{2} \tag{29}$$

where m is the estimated number of iterations and given in Equations (30, 31).

$$m \approx \frac{\ln\left(\frac{\pi}{2}\right) + \frac{1}{2}\ln(N)}{ln3} \tag{30}$$

$$m = \mathcal{O}(lnN) \tag{31}$$

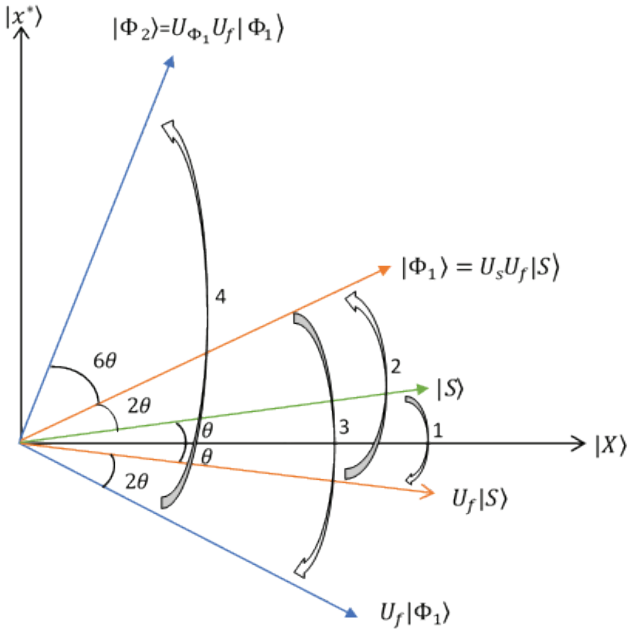


Figure 5. Evolution of equal superposition state $|S\rangle$ in proposed method.

As seen, an exponential speed-up is achieved in the proposed method. At every iteration, the amplitudes (and hence the probabilities) of unmarked and marked states were calculated. The calculations were performed for n=5, 10, 15, 20, 25, 30, 35, 40, 45 and 50 qubits. The results are shown in Figure 6 (a-j).

As for the case of n=2, only one iteration is required to find the marked state with probability one (not shown in Figure 6). The calculations showed that for n= 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50 qubits, the number of iterations required to find the marked state with relatively higher probability as compared to the unmarked states are 2, 5, 5, 7, 8, 10, 11, 13, 16, and 16, respectively as shown in Table 1.

When compared to Grover's method, in some cases, the probability of finding the marked state is relatively low in the proposed method. For example, for the cases n=20, 35 and 45 qubits, the probabilities of finding the marked states are about 0.7. For all other cases, the probabilities are about 0.9, which is sufficiently high.

The implementation of the methods mentioned in this study was carried out on a PC with an i5-8400 processor and 8 GB memory with Windows 10 operating system. MATLAB [33] was preferred as the software development environment. HPF Toolbox [34] is used to solve the large number of problems encountered due to the high number of exponential and multiplication operations during coding.

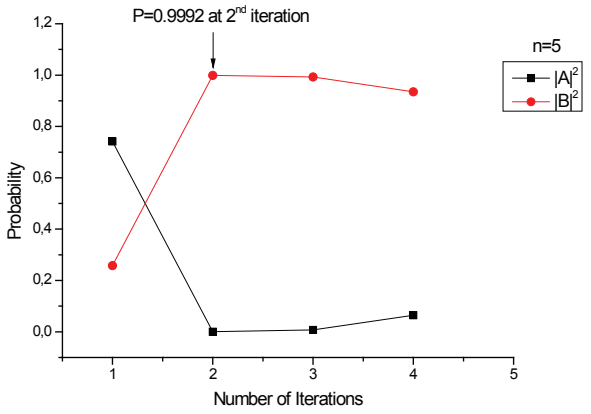
As it is clearly seen in Eq. (27-28), c and d values are calculated iteratively and each value takes the value calculated in the previous iteration as a multiplier. However, the a^{3^i} value is calculated at each stage. Here i denotes the iteration

Algorithm 2. Pseudo Code of Proposed Method

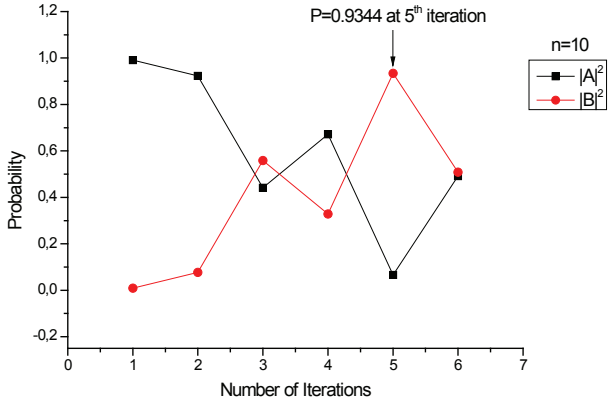
Inputs : n – number of qubits
n_step – number of iterations
Outputs : A[] – amplitudes of unmarked states
B[] – amplitudes of marked state

Begin
$N \leftarrow 2^n$, % number of search space
Set $a \leftarrow \sqrt{N}$
Set $b \leftarrow \sqrt{N-1}$
Set $c_0 \leftarrow 1$
Set $d_0 \leftarrow 1$
Set $i \leftarrow 0$
while $i < n\_step$ **do**
Set $prod\_c_j \leftarrow 1$, % $prod\_c_j$ *is* $\prod_{j=1}^{i} c_j$
Set $prod\_c_j^2 \leftarrow 1$, % $prod\_c_j^2$ *is* $\prod_{j=1}^{i} c_j^2$
Set $prod\_d_j^2 \leftarrow 1$, % $prod\_d_j^2$ *is* $\prod_{j=1}^{i} d_j^2$
Set $j \leftarrow 0$
while $j \leq i$ **do**
$prod\_c_j \leftarrow prod\_c_j * c_j$
$prod\_c_j^2 \leftarrow prod\_c_j^2 * c_j$
$prod\_d_j^2 \leftarrow prod\_d_j^2 * d_j * d_j$
$j \leftarrow j + 1$
end while
$\mathrm{T}_{i+1} \leftarrow 2b^2 prod\_c_j^2 - 2prod\_d_j^2$
$\mathrm{A}_i \leftarrow \frac{b.prod\_c_j}{a^{3^{i+1}}}\left(\mathrm{T}_{i+1} - a^{2.3^i}\right)$
$\mathrm{B}_i \leftarrow \frac{1}{a^{3^{i+1}}}\left(\mathrm{T}_{i+1} + a^{2.3^i}\right)$
$|\Phi_{i+1}\rangle \leftarrow \mathrm{A}_i|X\rangle + \mathrm{B}_i|x^*\rangle$
$c_{i+1} \leftarrow 2b^2c_i^2 - 2d_i^2 - a^{2.3^i}$
$d_{i+1} \leftarrow 2b^2c_i^2 - 2d_i^2 + a^{2.3^i}$
$i \leftarrow i + 1$
end while
End

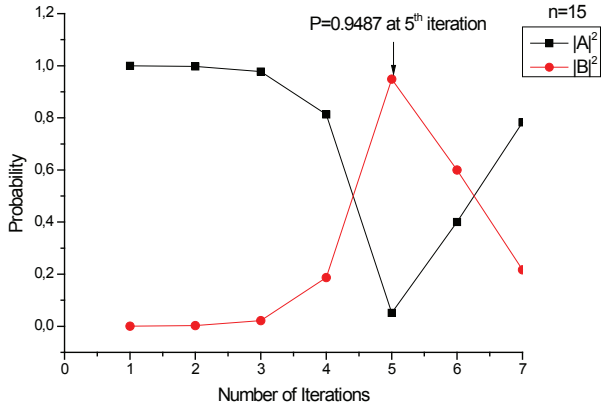
number. Therefore, as the iteration number increases, the value of a^{3^i} grows very quickly. Since the classical computer architecture does not allow the calculation of such large numbers, the HPF [34] library was used. With this library, instead of allocating a standard 16-digit memory for a variable, higher numbers of digits can be used. We preferred variable lengths of 10-200-500-1000 and 2000 digits for the very large numbers that emerged in this study. Although the use of more digits allows much larger numbers to be calculated, we had to terminate it at some point because the calculation time was prolonged. The average of the times required to calculate all other values in an iteration is given in the last column of the table. It is clearly seen in the table that the calculated values and times grow exponentially. However, especially after the 32nd iteration (usually around the 40th iteration), a buffer overflow occurs in the calculated values for all qubits. When all these situations are


P=0.9992 at 2nd iteration
n=5
|A|2
|B|2
Probability
Number of Iterations

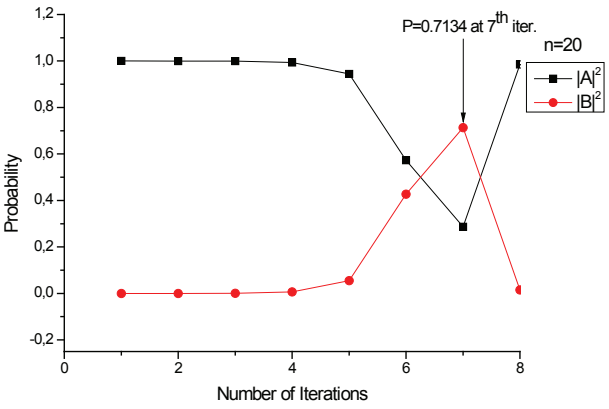
(a)


P=0.9344 at 5th iteration
n=10
|A|2
|B|2
Probability
Number of Iterations

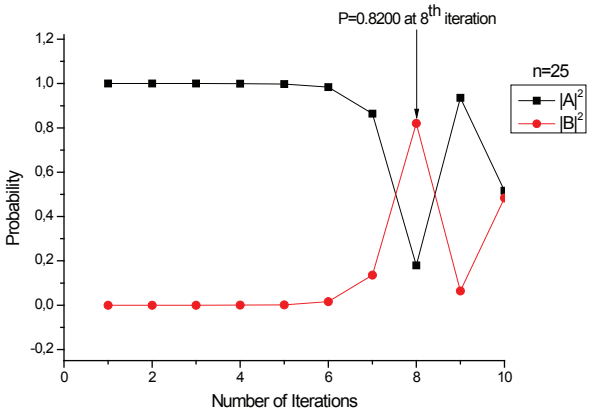
(b)


P=0.9487 at 5th iteration
n=15
|A|2
|B|2
Probability
Number of Iterations

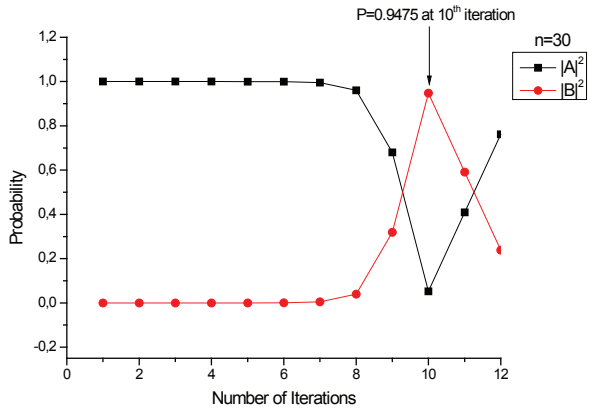
(c)


P=0.7134 at 7th iter.
n=20
|A|2
|B|2
Probability
Number of Iterations

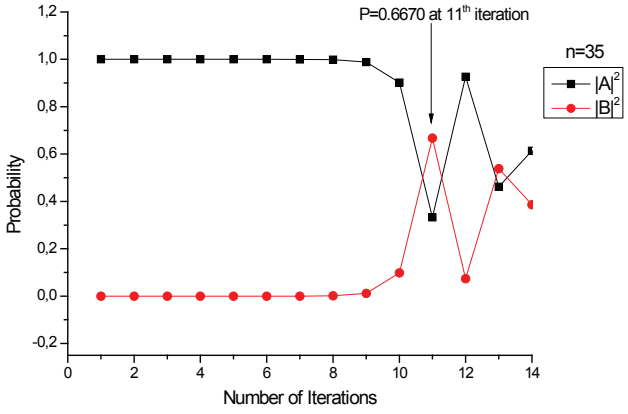
(d)


P=0.8200 at 8th iteration
n=25
|A|2
|B|2
Probability
Number of Iterations

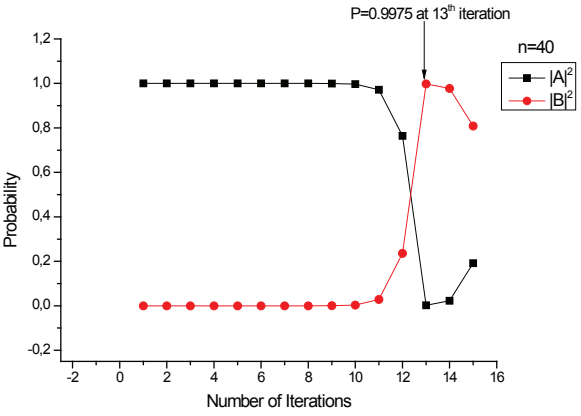
(e)


P=0.9475 at 10th iteration
n=30
|A|2
|B|2
Probability
Number of Iterations

(f)


P=0.6670 at 11th iteration
n=35
|A|2
|B|2
Probability
Number of Iterations

(g)


P=0.9975 at 13th iteration
n=40
|A|2
|B|2
Probability
Number of Iterations

(h)

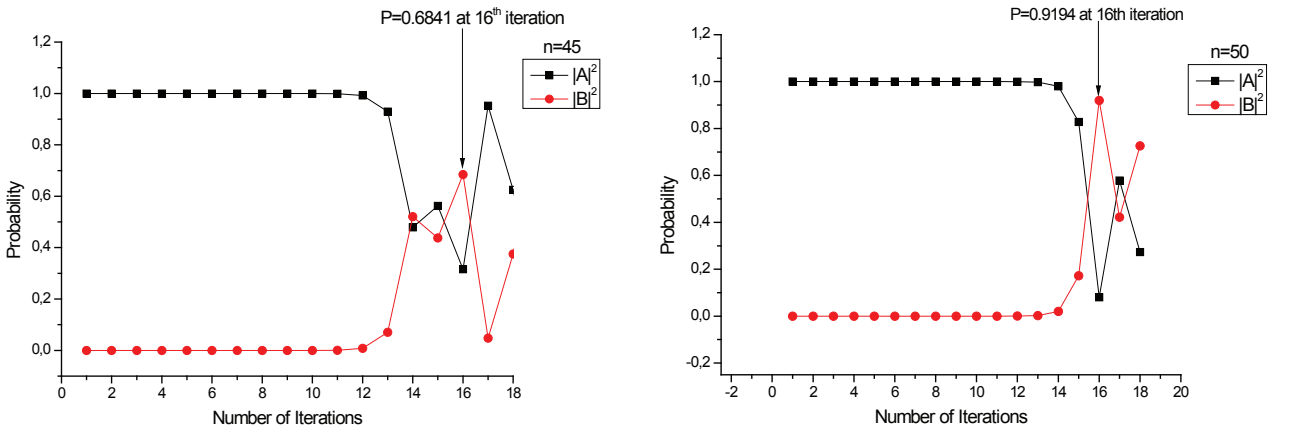


Figure 6. Probabilities of finding the marked and unmarked states.

Table 1. Comparison of the number of iterations and elapsed times for Grover's and proposed method

		Number of Iterations and Elapsed Time			
Number of qubits (n)	**Number of search space (N)**	**Iteration Number (Grover's Method)**	**Elapsed Time (Grover's Method)**	**Iteration Number (Proposed Method)**	**Elapsed Time (Proposed Method)**
1	2	1	0.2 s	1	0.04 s
5	32	4	0.8 s	2	0.07 s
10	1024	25	5 s	5	0.18 s
15	32768	142	28.4 s	5	0.18 s
20	1048576	804	160.8 s	7	0.24 s
25	33554432	4549	909.8 s	8	0.29 s
27	134217728	9099	1819.8 s	9	0.32 s
30	1073741824	25736	5547.2 s	10	0.35 s
35	34359738368	-		11	0.40 s
40	1.09951E+12	-		13	0.46 s
45	3.51844E+13	-		16	0.56 s
50	1.1259E+15	-		16	0.56 s

evaluated, libraries are needed in which much larger numbers can be used in order to reach 1000 iterations in total. The proposed approach was coded with the MATLAB program, and a more efficient library than HPF was not found. In this regard, it is planned to examine the libraries used for finding large prime numbers and to develop the developed approach with these libraries.

CONCLUSION

To our knowledge, there is no quantitative analysis of Grover's iterative search algorithm. In the current paper, we showed how the amplitudes and hence the probabilities of finding a single marked state and unmarked states evolve at each iteration for various qubit states. We also discussed the potential improvement of Grover's quantum search algorithm. The study performed in the current study can be summarized in two steps: First, we analyzed the quantitative analysis of Grover's search algorithm by calculating the amplitudes and hence the probabilities of finding the marked and unmarked states for various qubit states from n=5 to n=27 qubits. The calculations showed that while the probabilities of the marked states increase smoothly with increasing the number of iterations, the probabilities of unmarked states decrease smoothly with increasing the number of iterations. Second, we proposed an improved version of Grover's search algorithm. The calculations showed that a great speed-up is achieved after

improvement is made to Grover's search algorithm. For example, suppose a marked element is desired to be found in N=106 elements. Grover search algorithm can find that element in $\sqrt{10^6} = 10^3 = 1000$ iterations. However, after the improvement one can find the same element in only $\ln(10^6) \approx 14$ iterations which is a great speed up in the search algorithm. However, in the proposed method, for some cases, the probability of finding the marked state does not smoothly increase, and for some cases, the probability of finding the marked state is not high enough. That would limit the use of our proposed method in real applications.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Grover LK. Quantum mechanics helps in searching for a needle in a haystack. Phys Rev Lett 1997;79:325. [CrosssRef]

[2] Biham E, Biham O, Biron D, Grassl M, Lidar DA, Shapira D. Analysis of generalized Grover quantum search algorithms using recursion equations. Phys Rev A 2000;63:012310. [CrosssRef]

[3] Arqub OA, Abo-Hammour Z. Numerical solution of systems of second-order boundary value problems using continuous genetic algorithm. Inf Sci 2014;279:396–415. [CrosssRef]

[4] Abo-Hammour ZE, Alsmadi O, Momani S, Abu Arqub O. A genetic algorithm approach for prediction of linear dynamical systems. Math Probl Eng 2013;2013:831657. [CrosssRef]

[5] Abo-Hammour Z, Abu Arqub O, Momani S, Shawagfeh N. Optimization solution of Troesch's and Bratu's problems of ordinary type using novel continuous genetic algorithm. Discrete Dyn Nat Soc 2014;2014:401696. [CrosssRef]

[6] Abu Arqub O, Abo-Hammour Z, Momani S, Shawagfeh N. Solving singular two-point boundary value problems using continuous genetic algorithm. Abstr Appl Anal 2012;2012:205391. [CrosssRef]

[7] Kelleche A, Tatar NE. Control and exponential stabilization for the equation of an axially moving viscoelastic strip. Math Meth Appl Sci 2017;40:6239–6253. [CrosssRef]

[8] Kelleche A, Tatar NE. Adaptive Stabilization of a Kirchhoff moving string. J Dyn Control Syst 2020;26:255–263. [CrosssRef]

[9] Kelleche A, Tatar NE. Existence and stabilization of a Kirchhoff moving string with a distributed delay in the boundary feedback. Math Model Nat Phenom 2017;12:106–117. [CrosssRef]

[10] Kaye P, Laflamme R, Mosca M. An Introduction to Quantum Computing. Oxford, London: OUP; 2006. [CrosssRef]

[11] Shor PW. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM Rev 1999;41:303–332. [CrosssRef]

[12] Qiu D, Zheng S. Revisiting Deutsch-Jozsa algorithm. Inf Comput. 2020;275:104605. [CrosssRef]

[13] Deutsch D, Jozsa R. Rapid solution of problems by quantum computation. Proc R Soc Lond A Math Phys Sci 1992;439:553–558. [CrosssRef]

[14] Kiktenko EO, Fedorov AK, Strakhov AA, Man'Ko VI. Single qudit realization of the Deutsch algorithm using superconducting many-level quantum circuits. Phys Lett A 2015;379:1409–1413. [CrosssRef]

[15] Chuang IL, Gershenfeld N, Kubinec M. Experimental implementation of fast quantum searching. Phys Rev Lett 1998;80:3408. [CrosssRef]

[16] Chuang IL, Vandersypen LM, Zhou X, Leung DW, Lloyd S. Experimental realization of a quantum algorithm. Nature 1998;393:143–146. [CrosssRef]

[17] Feng M. Grover search with pairs of trapped ions. Phys Rev A 2001;63:052308. [CrosssRef]

[18] Brickman KA, Haljan PC, Lee PJ, Acton M, Deslauriers L, Monroe C. Implementation of Grover's quantum search algorithm in a scalable system. Phys Rev A 2005;72:050306. [CrosssRef]

[19] Xu ZY, Feng M. Addendum to "Grover search with pairs of trapped ions". Phys Rev A 2008;78:014301. [CrosssRef]

[20] Yang WL, Wei H, Zhou F, Chang WL, Feng M. Solution to the satisfiability problem using a complete Grover search with trapped ions. J Phys B At Mol Opt Phys 2009;42:145503. [CrosssRef]

[21] Dewes A, Lauro R, Ong FR, Schmitt V, Milman P, Bertet P, et al. Quantum speeding-up of computation demonstrated in a superconducting two-qubit processor. Phys Rev B 2012;85:140503. [CrosssRef]

[22] Dewes A, Ong FR, Schmitt V, Lauro R, Boulant N, Bertet P, et al. Characterization of a two-transmon processor with individual single-shot qubit readout. Phys Rev Lett 2012;108:057002. [CrosssRef]

[23] Filipp S, Maurer P, Leek PJ, Baur M, Bianchetti R, Fink JM, et al. Two-qubit state tomography using a joint dispersive readout. Phys Rev Lett 2009;102:200402. [CrosssRef]

[24] Yamaguchi F, Milman P, Brune M, Raimond JM, Haroche S. Quantum search with two-atom collisions in cavity QED. Phys Rev A 2002;66:010302. [CrosssRef]

[25] Wang H, Zhang S, Yeon KH. Implementation of Grover quantum search via cavity quantum electrodynamics. J Korean Phys Soc 2008;53:3144–150. [CrosssRef]

[26] Yang WL, Chen CY, Feng M. Implementation of three-qubit Grover search in cavity quantum electrodynamics. Phys Rev A 2007;76:054301. [CrosssRef]

[27] Hua M, Tao MJ, Deng FG. Fast universal quantum gates on microwave photons with all-resonance operations in circuit QED. Sci Rep 2015;5:1–8. [CrosssRef]

[28] Szabłowski PJ. Understanding mathematics of Grover's algorithm. Quantum Inf Process 2021;20:1–21. [CrosssRef]

[29] Han Q. Several remarks on Grover's quantum search algorithm with a single marked element. Sch J Phys Math Stat 2021;3:62–67. [CrosssRef]

[30] Cohn I, De Oliveira ALF, Buksman E, De Lacalle JG. Grover's search with local and total depolarizing channel errors: Complexity analysis. Int J Quantum Inf 2016;14:1650009. [CrosssRef]

[31] Xiao H, Huang WH, Zhou M. An efficient quantum private query protocol based on Oracle and Grover iteration. Int J Theor Phys 2019;58:3025–3035. [CrosssRef]

[32] Nielsen MA, Chuang I. Quantum Computation and Quantum İnformation. Cambridge: Cambridge University Press; 2002.

[33] Get a Trial of MATLAB and Simulink Product. Accessed onJul 12, 2024. Available at: https://uk.mathworks.com/

[34] D'Errico J. HPF - a big decimal class. MATLAB Central File Exchange. Accessed on Jan 10, 2022. Available at: https://www.mathworks.com/matlabcentral/fileexchange/36534-hpf-a-big-decimal-class

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1050–1057, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00087

Research Article

Operational matrix for multi-order fractional differential equations with hermite polynomials

Hatice YALMAN KOSUNALP[1,*], Mustafa GULSU[2]

[1]*Department of Accounting and Tax, Gonen Vocational School, Bandırma Onyedi Eylul University, Balıkesir, 10200, Türkiye*
[2]*Department of Mathematics, Faculty of Science, Mugla Sitki Kocman University, Mugla, 48000, Türkiye*

ARTICLE INFO

Article history
Received: 26 December 2022
Revised: 12 March 2023
Accepted: 04 April 2023

Keywords:
Fractional Differential Equations; Orthogonal Polynomials; Operational Matrix

ABSTRACT

In this article, a new operational matrix of fractional integration of Hermite polynomials is derived to solve multi-order linear fractional differential equations (FDEs) with spectral tau approach. We firstly convert the FDEs into an integrated-form through multiple fractional integration in association with the Riemann-Liouville sense. This integral equation is then formulated as an algebraic equation system with Hermite polynomials. Finally, linear multi-order FDEs with initial conditions are solved with this method. We present exact and approximated solutions for a number of representative examples. Numerical results indicate that the proposed method provides a high degree of accuracy to solve the linear multi-order FDEs.

Cite this article as: Yalman Kosunalp H, Gulsu M. Operational matrix for multi-order fractional differential equations with hermite polynomials. Sigma J Eng Nat Sci 2024;42(4):1050–1057.

INTRODUCTION

The calculus of fractional order can be considered as a generalization of ordinary differentiation

and integration to arbitrary order. The fractional calculus was born in 1695 with G.W.Leibniz's question arising the uncertainty of the rational order of the derivation [1]. Fractional calculus' history can be found in [1-3]. Fractional differential equations (FDEs) have gained an increasing interest with a wide range of significant applications within science domain [4-7]. Examples of application areas are mechanics [8], biology [9], signal processing [10], economics [11] and control theory [12]. The main motivation behind the research of FDEs is its high accuracy when compared with integer order models, providing a high level of flexibility for choosing degree of derivation. This is because FDEs ensure more realistic models for complex real-world problems. In order to solve the FDEs, efficient solutions are required accurately in which different methods have attempted to solve FDEs. In recent years, spectral methods have been an effective method for numerical solutions of FDEs, particularly in the area of computational fluid dynamics. A typical example of spectral methods attempts to formulate Jacobi pseudospectral scheme to solve multi-dimensional fractional Schrodinger equations in association with various boundary conditions [13]. It solves the variable-order FDEs by deriving the operational matrices for fractional variable-order of the derivative and integral with Jacobi polynomials. A series of attempts

***Corresponding author.**
*E-mail address: hkosunalp@bandirma.edu.tr

This paper was recommended for publication in revised form by Editor in-Chief Ahmet Selim Dalkilic

employing Laguerre polynomials have been carried out to solve the FDEs with different spectral methods in the scope of numerical methods [14]. The work conducted in [15] resulted in the development of a novel algorithm targeting time-dependent problems under the basis of spectral Laguerre approximations. A recent work considered the modified Laguerre functions by proposing a novel tau method [16]. It is based upon the operational matrix of fractional integration (OMFI) inspired by Riemann-Liouville paradigm, highlighting the efficiency of the proposed idea using illustrative examples. Another work exploited the Chebyshev polynomials in order to present a fractional radiative transfer equation, whereby the multi-dimensional issue is approximated by FDEs system [17]. In [18], a new explicit solution, which is targetted for shifted Chebyshev polynomials with flexible degree and fractional order, is formed to figure out multi-term FDEs. To solve the same linear problem, the work in [19] combined the shifted Chebyshev polynomials and extended spectral operational tau approach.

Jacobi polynomials have recently gained an important interest in both theory and practice. A derivation of shifted Jacobi operational matrix of fractional derivatives and spectral tau approach are applied together for solving the multi-term FDEs [20].To solve the nonlinear Langevin equation, on the other hand, a jacobi Gauss Lobatto collocation method is proposed in [21]. Authors in [22] included shifted Jacobi polynomials for a derivation of an OMFI using Riemann-Liouville, resulting in a direct solution of FDEs. An operational version of Legendre- tau technique to numerically solve the multi-term FDEs is also proposed in [23]. An extended work of Legendre polynomials presents an implementation of operational matrix with the sense of Riemann-Liouville [24]. Recently, a numerical solution for solving linear and non-linear FDEs is presented using Bernoulli polynomials with the methods of tau and collocation [25]. In this study, we attempt to present a new solution for the integrated form of FDEs with Hermite polynomials along with the fractional integration' operational matrix with the sense of Riemann Liouville. To do this, FDEs are initially re-written in the integral form which is then converted into an algebraic equation system with the introduction of the OMFI of Hermite polynomials. Upon the solution of the algebraic equations with initial conditions, we obtain exact and approximated solutions for a number of illustrative problems. The organization of the paper is as follows. Section II introduces the required notations and preliminaries, particularly Riemann-Liouville. The derivation of the Hermite OMFI is presented in section III. The operational matrix derived in the previous section is applied to solve linear FDEs in section IV. In section V, the proposed methods are implemented to various representative examples. Finally, the paper is concluded in section VI.

PRELIMINARIES AND NOTATION

The Fractional Integration in Rieman-Liouville Sense

The most common definition of Rieman-Liouville integration is:

$$J^{v}f(x)=\frac{1}{\Gamma(v)}\int_{0}^{x}(x-t)^{v-1}f(t)dt \quad v>0, x>0 \tag{2.1}$$

and if $v = 0$, then

$$J^{0}f(x)=f(x) \tag{2.2}$$

A significant property of Rieman-Liouville integration part is:

$$J^{v}x^{\beta}=\frac{\Gamma(\beta+1)}{\Gamma(\beta+1+v)}x^{\beta+v}. \tag{2.3}$$

The definition of Rieman-Liouville fractional derivation of order v is:

$$D^{v}f(x)=\frac{d^{m}}{dx^{m}}(J^{m-v}f(x)), \tag{2.4}$$

where $m-1<v\le m, m\in N$ and m is the smallest integer greater than v.

Lemma1. If $m-1<v\le m, m\in N$ then,

$$\begin{aligned} &D^{v}J^{v}f(x)=f(x) \text{ and} \\ &J^{v}D^{v}f(x)=f(x)-\sum_{i=0}^{m-1}f^{(i)}(0^{+})\frac{x^{i}}{i!}, \quad x>0. \end{aligned} \tag{2.5}$$

The Properties of Hermite Polynomials

Let $\Lambda=(-\infty,\infty)$ and $w(x)=e^{-x^{2}}$ be the weight function on Λ. The analytic form of Hermite polynomials of degree i is defined [26]

$$H_{i}(x)=\sum_{k=0}^{\left\lfloor\frac{i}{2}\right\rfloor}\frac{(-1)^{k}\,i!(2x)^{i-2k}}{k!(i-2k)!} \tag{2.6}$$

where $H_{0}(x)=1$ and $H_{1}(x)=2x$.

Hermite polynomials satisfy this recurrence relation

$$H_{i+1}(x)=2xH_{i}(x)+2iH_{i-1}(x). \tag{2.7}$$

The set of Hermite polynomials are orthogonal polynomials is an orthogonal system, namely

$$\int_{-\infty}^{\infty}H_{i}(x)H_{j}(x)w(x)dx=h_{j}\delta_{ij} \tag{2.8}$$

where $\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$ indicates the function of Kronecker and $h_j = 2^j j!\sqrt{\pi}$.

HERMITE OPERATIONAL MATRIX of FRACTIONAL INTEGRATION

In this section, we aim to derive an OMFI for Hermite polynomials. Let $u(x) \in L^2(\Lambda)$, then $u(x)$ can be defined in terms of Hermite polynomials as

$$u(x) = \sum_{j=0}^{\infty} a_j H_j(x) \tag{3.1}$$

Then, coefficient a_j can be written as

$$a_j = \frac{1}{2^j j!\sqrt{\pi}} \int_{-\infty}^{\infty} u(x) H_j(x) w(x) dx\,, \quad j = 0,1,\ldots \tag{3.2}$$

The initial (N + 1) terms of Hermite polynomials are only taken into consideration, such that

$$u_N(x) = \sum_{j=0}^{N} a_j H_j(x) = A^T \phi(x) \tag{3.3}$$

where

$$A^T = [a_0 \quad a_1 \quad \ldots \quad a_N] \text{ and}$$
$$\phi(x) = [H_0(x) \quad H_1(x) \quad \ldots \quad H_N(x)]^T \tag{3.4}$$

When we define q-step repeating integration of Hermite vector $\phi(x)$ by $J^q \phi(x)$ it will be

$$J^q \phi(x) = \boldsymbol{P}^{(q)} \phi(x) \tag{3.5}$$

where q indicates a fixed integer value and $\boldsymbol{O}^{(q)}$ represents the actual operational matrix of integration of $\phi(x)$.

Theorem 1. Let $\phi(x)$ be Hermite vector and $v > 0$ then

$$J^v(x) \cong O^{(v)} \phi(x) \tag{3.6}$$

where $O^{(v)}$ shows $(N+1) \times (N+1)$ OMFI of order v in the Rieman-Liouville sense which can be given as follows:

$$\boldsymbol{O}^{(v)} = \begin{bmatrix} \psi_v(0,0) & \psi_v(0,1) & \psi_v(0,2) & \ldots & \psi_v(0,n) \\ \psi_v(1,0) & \psi_v(1,1) & \psi_v(1,2) & \ldots & \psi_v(1,n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_v(i,0) & \psi_v(i,1) & \psi_v(i,2) & \cdots & \psi_v(i,n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_v(N-1,0) & \psi_v(N-1,1) & \psi_v(N-1,2) & \cdots & \psi_v(N-1,N) \\ \psi_v(N,0) & \psi_v(N,1) & \psi_v(N,2) & \cdots & \psi_v(N,N) \end{bmatrix} \tag{3.7}$$

where

$$\psi_v(n,j) = \sum_{i=0}^{\lfloor n/2 \rfloor} \sum_{r=0}^{\lfloor j/2 \rfloor} \frac{1}{2^j j!\sqrt{\pi}} \frac{(-1)^{i+r} n! j! 2^{n-2i+j-2r} \Gamma\left(\frac{n-2i+v+j-2r+1}{2}\right)}{2.i!r!(j-2r)!\Gamma(n-2i+v+1)}, \quad j = 0,1,2 \ldots N \tag{3.8}$$

Proof: We will apply Rieman-Liouville integration to the analytic form of Hermite polynomials as:

$$J^v(H_n(x)) = \sum_{i=0}^{\lfloor n/2 \rfloor} \frac{(-1)^i J^v[(2x)^{n-2i}]}{(n-2i)!i!} = \sum_{i=0}^{\lfloor n/2 \rfloor} \frac{(-1)^i 2^{n-2i} x^{n-2i+v}}{i!\Gamma(n-2i+1+v)} \tag{3.9}$$

If we approximate x^{n-2i+v} by N + 1 Hermite polynomial series; we obtain

$$x^{n-2i+v} = \sum_{j=0}^{N} c_j H_j(x) \tag{3.10}$$

where c_j is given from (3.2) and it is

$$c_j = \frac{1}{2^j j!\sqrt{\pi}} \sum_{r=0}^{\lfloor j/2 \rfloor} \frac{(-1)^r j! 2^{j-2r} \Gamma\left(\frac{n-2i+v+j-2r+1}{2}\right)}{2.r!(j-2r)!\Gamma(n-2i+v+1)}. \tag{3.11}$$

Then in virtue of (3.9) and (3.10), we obtain

$$J^v(H_n(x)) = \sum_{j=0}^{N} \psi_v(n,j) H_j(x) \quad i = 0,1,\ldots,N, \tag{3.12}$$

where

$$\psi_v(n,j) = \sum_{i=0}^{\lfloor n/2 \rfloor} \sum_{r=0}^{\lfloor j/2 \rfloor} \frac{1}{2^j j!\sqrt{\pi}} \frac{(-1)^{i+r} n! j! 2^{n-2i+j-2r} \Gamma(n-2i+v+j-2r+1)}{2.i!r!(j-2r)!\Gamma(n-2i+v+1)}, \quad j = 0,1,2,\ldots,N \tag{3.13}$$

HERMITE TAU METHOD WITH OPERATIONAL MATRIX

In practice, various problems are driven by initial value conditions of multi-term FDEs. This section modifies the Hermite tau method with the operational matrix for solving the FDEs. Each step of the whole process is given below.

$$D^v u(x) = \sum_{J=1}^{k} \gamma_j D^{\beta_J} u(x) + \gamma_{k+1} u(x) + f(x), \tag{4.1}$$

with initial conditions

$$u^{(i)}(0) = d_i \tag{4.2}$$

where $\gamma_{j,(j=0,1,2,\ldots,k+1)}$ are real constants and m-1 < v ≤ m, and $0 < \beta_1 < \beta_2 < \cdots < \beta_k < v$ and $f(x)$ is source function [16]. Rieman-Liouville integral of order v is applied to (4.1) after utilization (2.4), an integrated form of (4.1) is obtained, such as

$$u(x) - \sum_{j=0}^{m-1} u^{(j)}(0^+) \frac{x^j}{j!} = \sum_{i=1}^{k} \gamma_i J^{v-\beta_i} \left[u(x) - \sum_{j=0}^{m_i-1} u^{(j)}(0^+) \frac{x^j}{j!}\right] + \gamma_{k+1} J^v u(x) + J^v f(x) \tag{4.3}$$
$$u^{(i)}(0) = d_i, i = 0, \ldots, m-1$$

where $m_i - 1 < \beta_i \leq m_i,\ m_i \in N$. This states that

$$u(x) = \sum_{i=1}^{k} \gamma_i J^{v-\beta_i} u(x) + \gamma_{k+1} J^v u(x) + g(x), \tag{4.4}$$
$$u^{(i)}(0) = d_i, i = 0, \ldots, m-1$$

where

$$g(x) = J^{v} f(x) + \sum_{j=0}^{m-1} u^{(j)}(0^{+}) \frac{x^{j}}{j!} + \sum_{i=1}^{k} \gamma_i J^{v-\beta_i} \left[u(x) - \sum_{j=0}^{m_i-1} u^{(j)}(0^{+}) \frac{x^{j}}{j!} \right] \quad (4.5)$$

To apply Tau method with OMFI for Hermite polynomials to solve the fully-integrated problem (4.4) given by initial conditions (4.3), $u(x)$ and $g(x)$ are approximated by the Hermite polynomials as

$$u_N(x) \cong \sum_{i=0}^{N} c_i H_i(x) = C^T \phi(x) \quad (4.6)$$

$$g(x) \cong \sum_{i=0}^{N} g_i H_i(x) = G^T \phi(x) \quad (4.7)$$

where the vector $G = [g_0 \;\; g_1 \;\; \dots \;\; g_N]^T$ can be calculated from (4.7), whereas $[c_0 \;\; c_1 \;\; \dots \;\; c_N]^T$ is unkown vector. We then appply Rieman-Liouville integral of order v and ($v - \beta_j$) of the approximate solution, it is re-written as

$$J^{v} u_N(x) \cong C^T J^{v} \phi(x) \cong C^T \boldsymbol{P}^{(v)} \phi(x) \quad (4.8)$$

and

$$J^{v-\beta_j} u_N(x) \cong C^T J^{v-\beta_j} \phi(x) \cong C^T \boldsymbol{P}^{(v-\beta_j)} \phi(x), \;\; j = 1,2,\dots,k. \quad (4.9)$$

The residual $R_N(x)$ will be given as [24-25]

$$R_N(x) = \left(C^T - C^T \sum_{i=1}^{k} \gamma_i J^{v-\beta_i} - \gamma_{k+1} P^{(v)} - G^T \right) \phi(x) \quad (4.10)$$

with Tau method, by applying

$$\langle R_N(x), H_j(x) \rangle = \int_{-\infty}^{\infty} R_N(x).H_j(x)\, w(x) dx = 0, \quad j = 0,1,\dots,N-m. \quad (4.11)$$

N - m + 1 linear algebraic equations are generated. Then by using (3.2) and (4.6) for (4.3) we generate m linear equations. Then by solving these two sets of equations, we get the vector C. From the vector C, we obtain the approximate solution $u_N(x)$.

ILLUSTRATIVE EXAMPLES

Example 1. The first example is this following problem

$$D^2 u(x) + D^{3/2} u(x) + u(x) = x^2 + 2 + \frac{4x^{1/2}}{\Gamma(0.5)}, \quad u(0) = 0, u'(0) = 0, x \epsilon \Lambda. \quad (5.1)$$

whose exact solution is $u(x) = x^2$.

By applying our method for N = 2, we can write the approximate solution as

$$u_N(x) = \sum_{i=0}^{2} c_i H_i(x) = C^T \phi(x) \text{ and } g(x) = \sum_{i=0}^{2} g_i H_i(x) = G^T \phi(x)$$

From (3.8), we can find

$$\boldsymbol{P}^{(2)} = \begin{bmatrix} \mathbf{0.1250} & \mathbf{0.1410} & \mathbf{0.0625} \\ \mathbf{0.0940} & \mathbf{0.1250} & \mathbf{0.0705} \\ \mathbf{-0.1250} & \mathbf{-0.0940} & \mathbf{0} \end{bmatrix} \text{ and}$$

$$\boldsymbol{P}^{0.5} = \begin{bmatrix} \mathbf{0.3901} & \mathbf{0.2885} & \mathbf{0.0488} \\ \mathbf{0.3847} & \mathbf{0.3901} & \mathbf{0.1443} \\ \mathbf{-0.1580} & \mathbf{0.1923} & \mathbf{0.2925} \end{bmatrix} \text{ and } \boldsymbol{G} = \begin{bmatrix} \boldsymbol{g_1} \\ \boldsymbol{g_2} \\ \boldsymbol{g_3} \end{bmatrix}$$

By using equation (4.10) and (4.11) we obtain

$$1.5151 c_0 + 0.4787 c_1 - 0.2801 c_2 = 0.7185 \quad (5.2)$$

Now by applying (4.3) for the initial condition we have

$$c_0 - 2c_2 = 0, \quad 2c_1 = 0 \quad (5.3)$$

By solving linear system (5.2) and (5.3) we get

$$[c_0 \quad c_1 \quad c_2] = [0.5 \quad 0 \quad 0.25]$$

thus our solution $u_N(x) = [c_0 \quad c_1 \quad c_2] \begin{bmatrix} H_0(x) \\ H_1(x) \\ H_2(x) \end{bmatrix} = x^2$

which is the same as the exact solution.

Example 2. We now consider the following initial value problem as follow

$$D^{3/2} u(x) + 3u(x) = 3x^3 + \frac{8x^{3/2}}{\Gamma(0.5)} \; u(0) = 0, u'(0) = 0 \quad (5.4)$$

whose exact solution is given by $u(x) = x^3$.

For N = 3, if we apply our tecnique to this problem, the approximate solution

$$u_N(x) = \sum_{i=0}^{3} c_i H_i(x) = C^T \phi(x)$$

and

$$\boldsymbol{P}^{(\frac{3}{2})} = \begin{bmatrix} \boldsymbol{\psi}_{(0,0)} & \boldsymbol{\psi}_{(0,1)} & \boldsymbol{\psi}_{(0,2)} & \boldsymbol{\psi}_{(0,3)} \\ \boldsymbol{\psi}_{(1,0)} & \boldsymbol{\psi}_{(1,1)} & \boldsymbol{\psi}_{(1,2)} & \boldsymbol{\psi}_{(1,3)} \\ \boldsymbol{\psi}_{(2,0)} & \boldsymbol{\psi}_{(2,1)} & \boldsymbol{\psi}_{(2,2)} & \boldsymbol{\psi}_{(2,3)} \\ \boldsymbol{\psi}_{(3,0)} & \boldsymbol{\psi}_{(3,1)} & \boldsymbol{\psi}_{(3,2)} & \boldsymbol{\psi}_{(3,3)} \end{bmatrix} \text{ and } \boldsymbol{G} = \begin{bmatrix} \boldsymbol{g_0} \\ \boldsymbol{g_1} \\ \boldsymbol{g_2} \\ \boldsymbol{g_3} \end{bmatrix}$$

By using (4.10) and (4.11) we get

$$1.5770 c_0 + 0.4681 c_1 - 0.4946 c_2 - 1.5602 c_3 = 0.1560 \quad (5.5)$$

$$0.5851c_0 + 1.5770c_1 - 0.2340c_2 - 1.4838c_3 = 0.9973 \quad (5.6)$$

Now by applying (4.3) we get

$$c_0 - 2c_2 = 0 \quad (5.7)$$

$$2c_1 - 12c_3 = 0 \quad (5.8)$$

By solving Eqs. (5.5)-(5.8) we have 4 unknown coefficients, which are found as

$$[c_0 \quad c_1 \quad c_2 \quad c_3] = [0 \quad 0.75 \quad 0 \quad 0.125]$$

Thereby we can write our solution as

$$u_N(x) = \sum_{i=0}^{3} c_i H_i(x) = C^T \phi(x) = x^3$$

Example 3. Another example is considered

$$D^{0.5}u(x) + u(x) = \frac{2x^{1.5}}{\Gamma(2.5)} - \frac{x^{0.5}}{\Gamma(1.5)} + x^2 - x \quad u(0) = 0,$$

The exact solution of this example is:

$$u(x) = x^2 - x$$

With our method for N=2, we obtain the following equations

$$1.3901c_0 + 0.3847c_1 - 0.1560c_2 = 0.4637$$

$$0.2885c_0 + 1.3901c_1 + 0.1923c_2 = -0.5027$$

$$c_0 - 2c_2 = 0$$

Upon solution of these algebraic equations, we present the following values of C parameters

$$[c_0 \quad c_1 \quad c_2] = [0.5 \quad -0.5 \quad 0.25]$$

Therefore, our proposed method successfully finds the exact solution as

$$u_N(x) = [c_0 \quad c_1 \quad c_2] \begin{bmatrix} H_0(x) \\ H_1(x) \\ H_2(x) \end{bmatrix} = x^2 - x$$

Example 4. Following problem is also considered

$$D^{3/2}u(x) + u(x) = x^2 + \frac{4x^{1/2}}{\Gamma(0.5)}, \quad u(0) = 0, u'(0) = 0 \quad (5.9)$$

whose exact solution is $u(x) = x^2$.

For N=2 if we apply our method and we obtain

$$[c_0 \quad c_1 \quad c_2] = [0.5 \quad 0 \quad 0.25].$$

From here, the approximate solution is obtained the same as the exact solution like

$$u_N = u(x) = x^2.$$

Example 5. Consider this initial value problem

$$D^2u(x) + D^{3/2}u(x) + u(x) = x^2 + 2 + \frac{4x^{1/2}}{\Gamma(1.5)},$$
$$u(0) = 0, u'(0) = 0 \quad (5.11)$$

whose exact solution is $u(x) = x^3$.

After applying our technique for $N = 3$ we get $[c_0 \quad c_1 \quad c_2 \quad c_3] = [0 \quad 0.75 \quad 0 \quad 0.125]$.

The approximate solution is $u_N(x) = u(x) = x^3$.

Example 6. The following inital value problem is considered

$$D^{\alpha}u(x) + u(x) = 0, \quad (5.12)$$

with conditions

$$0 < \alpha \leq 2, \quad u(0) = 1, u'(0) = 0$$

Also, we have the second initial condition valid for only α > 1 [27]. The exact solution is given as:

$$u(x) = \sum_{k=0}^{\infty} \frac{(-x^{\alpha})^k}{\Gamma(\alpha k + 1)}$$

The proposed method solves this problem, and the absolute error is given in fig. 1 for α = 0.75, 0.85, 0.95 and N = 4. It is important to note that the exact solution converges to the analytical solution of exp(-x) with α = 1. It can

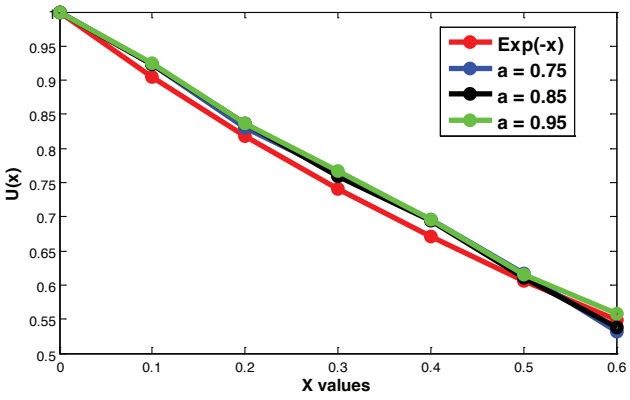


Figure 1. Comparisons of u(x) with varying α = 0.75, 0.85, 0.95.

be seen that our method presents a good approximation in comparison to results given in [23].

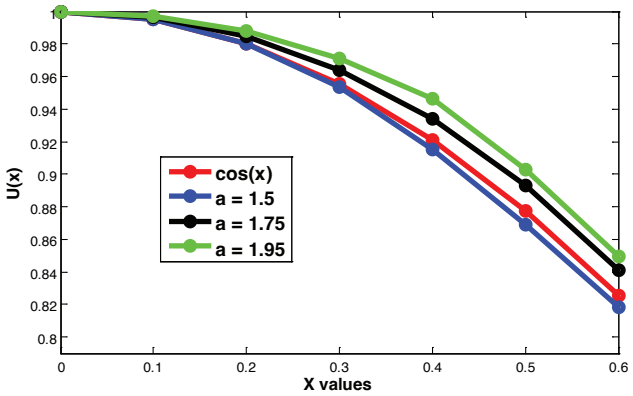


Figure 1. Comparisons of u(x) with varying α = 1.5, 1.75, 1.95.

In order to show the results for α > 1, where exact solution is given as cos(x) with α = 2, we selected three values of α, as indicated in fig. 2 below. A good balance between the exact solutions and obtained solutions is achieved. For α = 1.5, the best results are achieved with a very low error ratio.

Example 7. Consider the following equation

$$D^2u(x) - aD^v u(x) - bu(x) = 8, \tag{5.13}$$

with conditions

$$0 < v < 2, u(0) = 0, u'(0) = 0$$

We obtain approximated solutions for v = 0.5 and 1.5 with varying N, which are illustrated in Table 1 and Table 2 compared with the exact solution. In this solution, we use a special case for a = b = -1. The results exhibit a satisfactory approximation solution with solutions presented in [24].

Table 1. Numerical results in comparison to exact solution for v = 0.5

x	N=2	N=3	N=4	N=5	Exact Solution
0	0	0	0	0	0
0.1	0.0291	0.0343	0.0378	0.378	0.03975
0.2	0.1164	0.1352	0.1488	0.1492	0.157036
0.3	0.2619	0.2998	0.3293	0.3306,	0.347370
0.4	0.4656	0.5251	0.5746	0.5779	0.604695
0.5	0.7275	0.8080	0.8798	0.8863	0.921768
0.6	1.0476	1.1458	1.2393	1.2502	1.290452
0.7	1.4259	1.5353	1.6468	1.6633	1.702008
0.8	1.8624	1.9736	2.0955	2.1185	2.147287
0.9	2.3571	2.4578	2.578	2.6078	2.617001
1	2.91	2.9848	3.0863	3.1224	3.101906

Table 2. Numerical results in comparison to exact solution for v = 1.5

x	N=2	N=3	N=4	N=5	Exact Solution
0	0	0	0	0	0
0.1	0.0291	0.0304	0.0319	0.0321	0.125221
0.2	0.1164	0.1187	0.1245	0.1246	0.033507
0.3	0.2619	0.2607	0.2731	0.2717	0.267609
0.4	0.4656	0.4518	0.4729	0.4675	0.455435
0.5	0.7275	0.6879	0.7186	0.7061	0.684335
0.6	1.0476	0.9646	1.0049	0.9816	0.950393
0.7	1.4259	1.2775	1.3260	1.2884	1.249959
0.8	1.8624	1.6223	1.6762	1.6213	1.579557
0.9	2.3571	1.9947	2.0491	1.9752	1.935832
1	2.91	2.3903	2.4385	2.3458	2.315528

CONCLUSION

This paper presented a general derivation for the OMFI of the Hermite polynomials. Riemann-Liouville sense is exploited to define the FDE as a form of fully integrated integration. The operational matrix obtained is a key part of the idea, in order to approximate the numerical solutions of the linear FDEs. A number of signals existed in the integrated form equation are treated as linear combinations of the Hermite polynomials. Then, a final algebraic equation is obtained with the integrated form equation introducing the OMFI of the Hermite polynomials. The numerical solutions obtained showed the accuracy of the proposed method.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICTS OF INTEREST

The author(s) declare that there are no conflicts of interest regarding the publication of this paper.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Carpinteri A, Mainardi F. Fractals and fractional calculus in continium mechanics. Vienna: Springer-Verlag Wien; 1997. [CrossRef]

[2] Miller KS, Ross B. An introduction to the fractional calculus and fractional differential equations. Toronto: John Wiley and Son;1993.

[3] Oldham KB, Spainer J. The fractional calculus, theory and applications of differentiation and integration to arbitrary order. Mineola, New York: Dover Publications; 2006.

[4] Podlubny I. Fractional differential equations. 1st ed. Cambridge, Massachusetts: Academic Press; 1999.

[5] Das S. Functional fractional calculus for system identification and controls. Heidelberg, Berlin: Springer; 2008.

[6] Kilbas AA, Srivastava HM, Trujillo JJ. Theory and applications of fractional differential equations. Amsterdam, Netherlands: Elsevier; 2006.

[7] Baleanu D, Diethelm K, Scalas E, Trujillo JJ. Fractional calculus models and numerical methods. Hackensack, New Jersey; World Scientific Publishing; 2012. [CrossRef]

[8] Plonka A. Recent developments in dispersive kinetics. Progr React Kinet Mech 2000;25:109-127. [CrossRef]

[9] Allegrini P, Buiatti M, Grinolini P, West BL. Fractional brownian motion as a nonstationary process: An alternative paradigm for DNA sequences. J West Phys Rev 1998;57:558-567. [CrossRef]

[10] Bisquert J. Fractional diffusion in the multiple-trapping regime and revision of the equivalence with the continuous time random walk. Phys Rev Lett 2003;91:010602. [CrossRef]

[11] Bailie RT, King ML. Fractional differencing and long memory processes. J Econom 1996;73:1-3. [CrossRef]

[12] Outsaloup A. La Commande CRONE: Commande robuste d'ordre non nntiere. Paris, France: Hermes; 1991.

[13] Bhrawy AK, Abdelkawy MA. A fully spectral collocation approximation for multi-dimensional fractional Schrödinger equations. J Comput Phys 2015;294:462-483. [CrossRef]

[14] Guo BY, Zhang XY. A new generalized laguerre approximation and its applications. J Comput Appl Math 2005;181:342-363. [CrossRef]

[15] Mikhailenko BG. Spectral laguerre method for the approximate solution of time-dependent problems. Appl Math Lett 1999;12:105-110. [CrossRef]

[16] Bhrawy AH, Alghamdi MM, Taha MT. A new modified generalized laguerre operational matrix of fractional integration for solving fractional differential equations on the half line. 2012;2012:179. [CrossRef]

[17] Ding XL, Jiang YL. Waveform relaxation methods for fractional differential equations with the Caputo derivatives. J Comput Appl Math 2012;16:573-594. [CrossRef]

[18] Bhrawy AH, Tharwat MM, Yıldırım A. A new formula for fractional integrals of chebyshev polynomials: application for solving multi-term fractional differential equations. Appl Math Model 2013;37:4245-4252. [CrossRef]

[19] Bhrawy AH, Alofi AS. The operational matrix of fractional integration for shifted chebyshev polynomials. Appl Math Lett 2013;26:25-31. [CrossRef]

[20] Doha EH, Bhrawy AH, Ezz-Eldien SS. A new jacobi operational matrix: An application for solving fractional differential equations. Appl Math Model 2013;36:4931-4943. [CrossRef]

[21] Bhrawy AH, Alghamdi MA. A shifted jacobi-gauss-lobatto collocation method for solving nonlinear fractional langevin equation. Bound Value Probl 2012;2012:62. [CrossRef]

[22] Bhrawy AH, Tharwat MM, Alghamdi MA. A new operational matrix of fractional integration for shifted jacobi polynomials. Bull Malays Math Sci Soc 2014;37:983-995.

[23] Saadatmandi A, Dehghan M. A new operational matrix for solving fractional-order differential equations. Comput Math Appl 2014;3:1326-1336. [CrossRef]

[24] Akrami MH, Atabakzadeh MH, Erjaee GH. The operational matrix of fractional integration for shifted legendre polynomials. Iran J Sci Technol 2013;37:439-444. [CrossRef]

[25] Belgacem R, Bokhari A, Amir A. Bernoulli operational matrix of fractional derivative for solution of fractional differential equations. Gen Lett Math 2018;5:32-46. [CrossRef]

[26] Poularikas AD. The handbook of formulas and tables for signal processing. 1st ed. Boca Raton: CRC Press; 1999. [CrossRef]

[27] Kumar P, Agrawal OP. An approximate method for numerical solution of fractional differential equations. Signal Process 2006;86:2602-2610. [CrossRef]

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1058–1066, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2023.00033

Research Article

A new improvement technique for expansive soils

İlyas ÖZKAN[1,*], Erdal ÇOKÇA[1]

[1]*Middle East Technical University, Department of Civil Engineering, Ankara, 06800, Türkiye*

ARTICLE INFO

Article history
Received: 10 January 2023
Revised: 16 March 2023
Accepted: 04 May 2023

Keywords:
Calcium Lignosulphonate; CBR Swell, Expansive Soil; Lime Pile, Prewetting; Soil Improvement; Swell

ABSTRACT

In this study, a Unified Soil Improvement Technique (USIT) that consists of prewetting, calcium lignosulphonate lime piles, fabric, and soil nail was applied to an expansive soil specimen to both accelerate the process and increase the effect of these techniques. 19 boreholes that had 0.45 cm diameter and 1.9 cm height were opened into an expansive soil specimen in an oedometer ring. Then, these boreholes were filled with a calcium lignosulphonate-water mixture prepared in a liquid form during the prewetting process. Thereafter, these boreholes were filled with a mixture including calcium lignosulphonate, water, and lime. Finally, a fabric was laid down over this specimen and seven pieces of soil nails were used for connecting the fabric and the lime piles. To sum up, the swelling potential of expansive soil specimens under 7 kPa decreased from 43.95 % to 1.58 % after 28 day curing period. The CBR swell of the USIT specimens, which waited in the humid room for 90 days, was measured as 1.14% under 7 kPa surcharge pressure and was measured as 0.4% under 25 kPa surcharge pressure. The significant changes in the swell potential values suggest that the Unified Soil Improvement Technique's performance is better than calcium lignosulphonate-added lime piles in the improvement of expansive clays.

Cite this article as: Özkan İ, Çokça E. A new improvement technique for expansive soils. Sigma J Eng Nat Sci 2024;42(4):1058–1066.

INTRODUCTION

Expansive soils, which are called problematic soils due to their volume change capacity due to the change in water contents of them, is present in many parts of the world. The damages of these soils can be seen on lightweight structures such as roads, pipelines, one-story buildings, etc. Although these damages do not lead to the loss of life, they annually give rise to a significant economic loss in the whole world [1,2].

Many improvement techniques have been developed to minimize the damage caused by the expansive soil for the last decades. These techniques include chemical additives, prewetting, soil replacement with compaction control, moisture control, surcharge loading, and thermal methods [3].

The prewetting technique is that the water content of the soil is increased before the construction to block the volume change of the soil. Under field conditions, enabling the soaking process to be executed within a short time period and maintaining the soil at high moisture levels constantly is far from attainable as stated by Nelson et al. (2015)

***Corresponding author.**
*E-mail address: iozkan@erbakan.edu.tr

This paper was recommended for publication in revised form by Editor in-Chief Ahmet Selim Dalkilic

[1]. Time is a considerable factor in the achievement of this technique applied in expansive soils, as it may take a long time to see the effect of this method. Since it takes a long time to transmit water to deeper soil particles due to low hydraulic conductivity. Therefore, the sand drain technique has been generally used to solve this problem [3,4].

Lignosulphonates, which are by-products of the paper industry, are materials that are usually used in many different industries from oil production to construction. These waste materials with water-soluble properties are derived from lignin, a wood part. Moreover, the basis of the categorization process depends on the complex polymer of the wood as the properties and chemical composition of lignin vary to wood types. This material is generally categorized as calcium lignosulphonate, magnesium lignosulphonate, sodium lignosulphonate, etc concerning their chemical compositions of them. Annual production of lignosulphonate is 50 million tonnes in the world. However, very little amounts of these materials have been used as agriculture chemicals, industrial binders, and water reducers in different areas. For example, the addition of these waste materials into soils results in the reduction of water evaporation. In another example, concrete that includes lignosulphonate is more fluid than concrete without lignosulphonate. They are mostly found in two different types that are powder and liquid in the market [5-8].

Some studies related to the utilization of lignosulphonate in geotechnical engineering have been done for the last decades. Tingle and Santoni (2003) explained that the unconfined compressive strength of low- plasticity clay improved with the addition of lignosulphonate [9]. For example, the unconfined compressive strength of this clay was increased approximately five times with the addition of lignosulphonate. An unconfined compressive strength of marl soil specimens was increased with the addition of lignosulphonate. In addition to this, the improvement percentage occurred of this parameter of marl soils also heightened with the curing period [10]. Tingle (2007) stated that the swelling potential of clay soil decreased with the addition of any type of lignosulphonate [11].

The addition of lignosulphonate into erodible soil resulted in both the rise of critical shear stress and the reduction of soil erosion coefficient [12]. Under cyclic loading, the deformation behavior of sandy silt was enhanced with the addition of lignosulphonate [13]. The lignosulphonate stabilization process of clay soil ended up with both the reduction of diffuse double-layer thickness and the declination of soil dispersivity [14].

The lime pile technique, which has been applied either dry or wet, has been generally used for the improvement of soft soil. Studies related to this technique have been divided into two parts that are soil stabilization and lime diffusion into the soil. Lime diffusion into the soil is the main problem of this technique as diffusion velocity is very slow. For example, the velocity of lime diffusion into soil that has no extensive cracks and fissures was determined as 12.5 mm/year. A study shows that the addition of lignosulphonate into the lime pile increased the velocity of lime diffusion. A study done by Gnana et al. (2022) stated that the fly ash addition to lime pile enhanced the strength properties of expansive soil specimens [15-19].

Geotextile is a material that has been utilized for many geotechnical applications such as reinforcement, drainage, slope stability, and separation. Geotextile application on expansive soil results in the reduction of swelling pressure. For example, the lime-treated coir geotextile reduces the upward swelling pressure of the expansive soil layer by approximately 52.19% in single-layer and 81.89% in double-layer. The percentage swell rate of expansive soil also reduced by 77.59% and 94.53% with a single-layer and double-layer of lime-treated coir geotextile, respectively [20-24].

Soil nails that are referred to as ground anchors have been reinforced for the enhancement of soft soil [25]. In general, the soil nail is not used in the improvement of the stabilization of expansive soil. In this study, the soil nail will be used as a connecting material between the lignosulfonate lime column and the geotextile.

Madhyannapu, R.S., and Puppala, A.J., (2014) studied the effectiveness of Deep Soil Mixing (DSM) technology in mitigating the swell behavior of an expansive subsoil under actual field conditions and found that the DSM technology was effective in mitigating the swell behavior of the expansive subsoil [26].

Techniques of expansive soil stabilization that are mentioned above have been commonly preferred and applied in the field. However, the achievement of these techniques

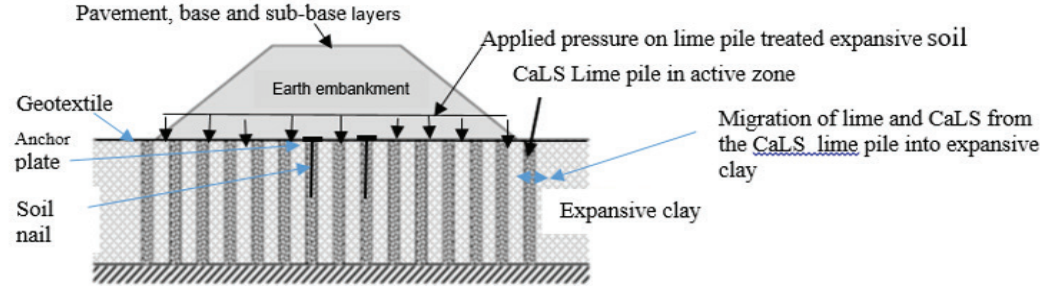


Figure 1. Unified Soil Improvement Technique under the pavement and embankment load.

Table 1. Properties of kaolinite and bentonite (KB) mixture

Properties, (Unit)	Value	Properties, (Unit)	Value
Specific gravity	2.56	Maximum dry density (Mg/m^3)	1.495
Liquid limit (%)	95.2	Swelling potential under 7 kPa (%)	43.95
Plastic limit (%)	23.2	Swelling potential under 25 kPa (%)	29.47
Plasticity index (%)	72.3	Swelling pressure (kPa)	265
Clay content (%)	57	Unsoaked CBR (%)	51.52
Activity	1.27	Soaked CBR (%)	1.31
Optimum water content (%)	26	CBR swell at soaked condition (%) under 7 kPa	40.3

depends on many factors, especially time. An essential aim of this study is the solution of the time factor by combining improvement techniques that are prewetting, lignosulphonate lime column, and geotextile-soil nailing (Figure 1).

Expansive Soil Specimen

The expansive soil specimen used in this study was obtained by mixing kaolinite and bentonite in the laboratory (it is called KB). The dry weight percentages of both kaolinite and bentonite were selected as 85% and 15%, respectively. KB is high plasticity clay according to the Unified Soil Classification System. The dry density and water content of KB were selected as 1.495 Mg/m^3 and 15%, respectively. The properties of KB are given in Table 1.

KB is classified as having high swelling potential concerning Seed et al. (1962) [27]. The soaked CBR value and CBR swell value of KB specimens do not fulfill the criteria for the subgrade material specified in Highways Technical Specification (HTS, 2013), and the soil stabilization/improvement required (in a case where the soaked CBR value < 10% and CBR swell % > 3%) [28]. The methods, which consisted of prewetting, calcium lignosulphonate lime pile, and fabric-soil nails, were used to improve the properties of KB.

EXPERIMENTAL WORK

The first part of the experimental work was based on the free swell test and the second part is related to Unified Soil Improvement Techniques (USIT).

The Free Swell Test

The swelling potential is a considerable parameter of expansive soil, free swell tests were done on specimens prepared at this part for measuring the achievement of improvement methods used in this study. All specimens prepared at this part had 63.5 mm diameter and 19 mm height. 7 kPa surcharge pressure was applied to all specimens prepared in this study. All tests were done in this part according to ASTM D4546 [29].

The part of this study formed three subtitles that were prewetting, lignosulphonate lime pile, and fabric-soil nails.

Prewetting

The dry density of KB was prepared at maximum dry density, 15% water content. The water content of KB (15%) is too low considering its optimum water content (26%). For this reason, the water content of KB was increased from 15% to 26.5% (optimum water content) by applying a prewetting process. During the process, a calcium lignosulphonate (CaLS) -water mixture instead of water was added to KB. The water loss of soils that resulted from evaporation could be diminished by adding lignosulphonate [30]. However, the density of this mixture was increased with the addition of lignosulphonate. For this reason, the calcium lignosulphonate-water mixture consisted of 98% water and 2% calcium lignosulphonate. To investigate the effect of the curing process, the specimens prepared in this study waited in a humid room that maintains a humidity greater than 95% for 1 day, 7 days, and 28 days.

Ca Lignosulphonate Lime Pile (LLP)

Treated specimen with Ca lignosulphonate lime pile (LLP) involved 37 pieces pile that had 4.5 mm diameter and 19 mm height. Each pile was filled with a CaLS lime pile mixture. The properties of lignosulphonate lime piles constructed in KB are given in Table 2.

This mixture consisted of lime, calcium lignosulphonate (CaLS), and water. Figure 2 illustrates both Ca

Table 2. The properties of Ca lignosulphonate lime piles (LLP)

Pile distance center to center (1.9 D) (cm)	Pile diameter (D) (cm)	Pile Length (H) (cm)	Pile area ratio (%)	All piles volume (cm^3)
0.855	0.45	1.9	20.91	11.18

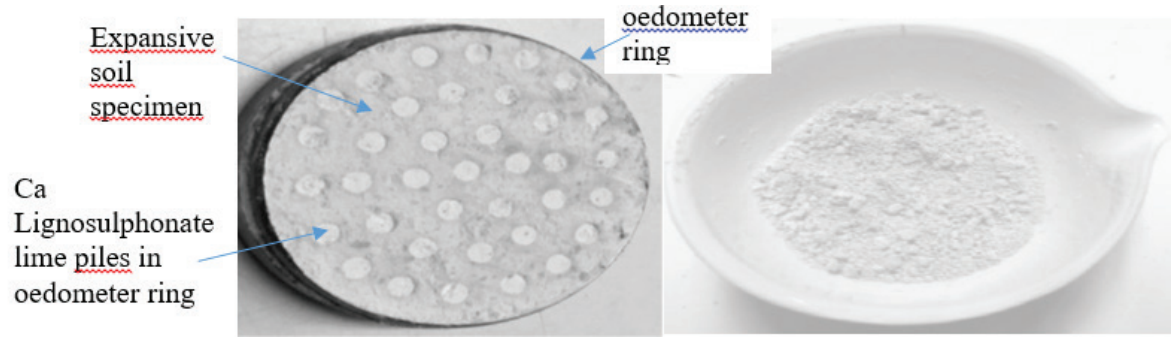


Figure 2. Ca Lignosulphonate lime pile (LLP) (left) in oedometer ring and CaLS lime mixture (right).

Table 3. The weight percentage of the calcium lignosulphonate lime pile mixture

	Water	Calcium lignosulphonate (CaLS)	Lime
Percentage by weight (%)	30%	3.75%	66.25%

Lignosulphonate LP (LLP) in the oedometer ring and CaLS lime mixture.

The weight percentages of the CaLS lime pile mixture are given in Table 3.

The amount of both water and calcium lignosulphonate in this mixture increased lime diffusion into expansive soil [18]. Treated specimens prepared at this part were waited in the humid room for 1 day, 7 days, and 28 days to observe the effect of the curing period.

Fabric and Soil Fasteners

The free swell test was done on specimens that had 63.5 mm diameter and 19 mm height, these dimensions were very small concerning field conditions. For this reason, a piece of nonwoven fabric instead of geotextile was used to model geotextile. A fastener, which was nonrusting metal and had 8 mm height and 9 mm top diameter, was used to model soil nailing and an anchor plate for connecting fabric and Ca lignosulphonate lime piles. The specimen treated with both fabric and soil fasteners/soil nail+nut was called GSN in this study.

Swelling Potentials of Treated Specimens

The swelling potentials of treated specimens were calculated according to data obtained from the free swell test done under 7 kPa surcharge pressure. These swelling potentials of treated specimens with either prewetting or Ca Lignosulphonate Lime Pile (LLP) techniques are given in Figure 3. It was observed that there are significant changes in the swelling percent of the treated soils due to the clay–lime reactions and clay-CaLS reactions in the Ca lignosulphonate lime pile technique.

The swelling potentials of KB decreased with both prewetting and LLP techniques. In addition to this, these falls continued with the curing period for both techniques. LLP technique was more effective than prewetting for this study. However, the falls obtained from the 28-day curing

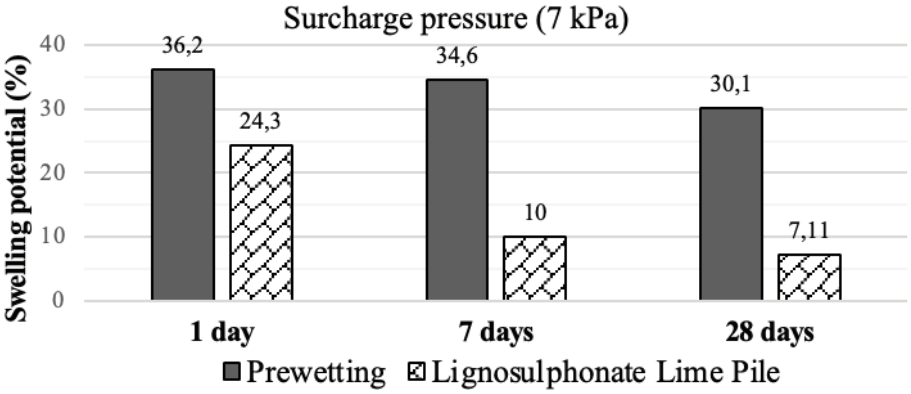


Figure 3. Swelling potentials of treated specimens with either prewetting or LLP.

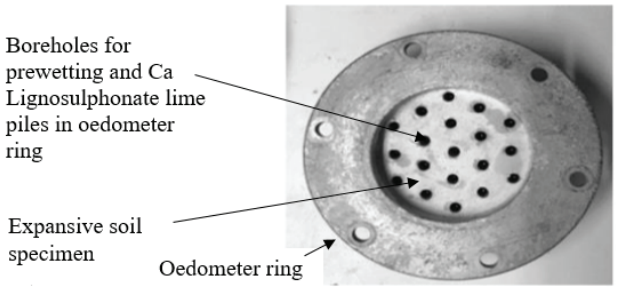


Figure 4. Number of piles in oedometer ring for USIT technique.

period were not sufficient for the stabilization of expansive soils, and do not meet the required criteria.

Unified Improvement Technique (USIT)

The Unified Improvement Technique (USIT) was formed by combining techniques that are prewetting, Ca lignosulphonate lime pile, and fabric+soil fasteners. This technique was named The Unified Improvement Technique (USIT). But, during the installation of soil fasteners to the lime piles, damage occurred to the nearby lime piles, to solve this problem, the number of piles at the LLP technique was reduced from 37 to 19 (Figure 4) and the distance between fasteners was determined as 6D (2.7 cm). Thus, the quantities of piles installed in specimens were felt by half of the LLP technique.

The reduction of pile number installed in KB causes some parameters related to piles. Table 4 shows the parameters related to piles installed in the USIT specimen in this study.

Table 4. Parameters related to calcium lignosulphonate lime piles installed in USIT specimen

Parameters	USIT
Pile distance center to center (cm)	1.35 (3D)
Pile area ratio (%)	10
All piles volume (cm^3)	5.74

The center-to-center distance of piles installed in the USIT specimen increased concerning the LLP specimen. Thus, both the area ratio and volume of piles installed in the USIT specimen are lower than in the LLP specimen. The installation cost of the USIT specimen is reduced compared to the LLP specimen.

Preparation of USIT specimen

Nineteen boreholes were opened into the USIT specimen before Prewetting, and the LLP and GSN techniques were applied, respectively.

Each borehole was filled with the calcium lignosulphonate-water mixture during the prewetting process. Boreholes were saturated by this mixture (Figure 5). After this process, the water content of KB was increased from 15% to 26% (optimum water content of KB). This process took a while due to the low permeability of KB.

Each borehole placed in the USIT specimen was filled with the Ca lignosulphonate-lime mixture. The achievement of the GSN technique of this study depends on the condition that the fastener (to model soil nails and anchor plates) should be installed properly.

The fabric (which model geotextile) was laid down on the treated specimen. The fastener was not installed in each borehole of the USIT specimen. The distance between fasteners was determined as 6D (2.7 cm). Figure 6 illustrates both plans of fasteners and the final top view of the USIT specimen.

Swelling Potential of USIT Specimen

Free swell tests were done under both 7 kPa and 25 kPa surcharge pressures. To observe the curing effect, USIT specimens were waited in the humid room for 1 day, 7 days, 28 days, 90 days, and 180 days respectively.

The swelling potentials of the USIT specimens were shown in Figure 7 concerning curing periods.

Under 7 kPa surcharge load, the CaLS lime pile case gives 2.63 times higher swell percent (1-day curing), 3.62 times higher swell percent (7 days curing), and 4.5 times higher swell percent (28 days curing) than the USIT case.

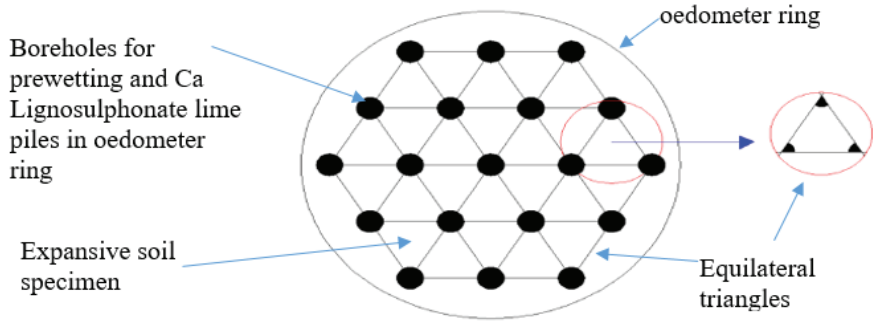


Figure 5. Equilateral triangles into the expansive soil specimen in oedometer ring.

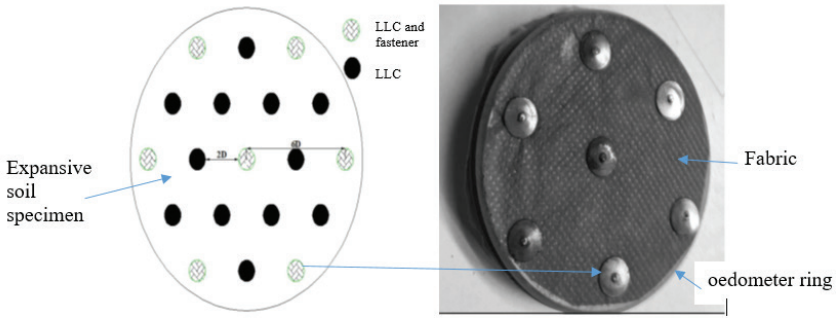


Figure 6. The plan of fasteners and final top view of the USIT specimen in oedometer ring.

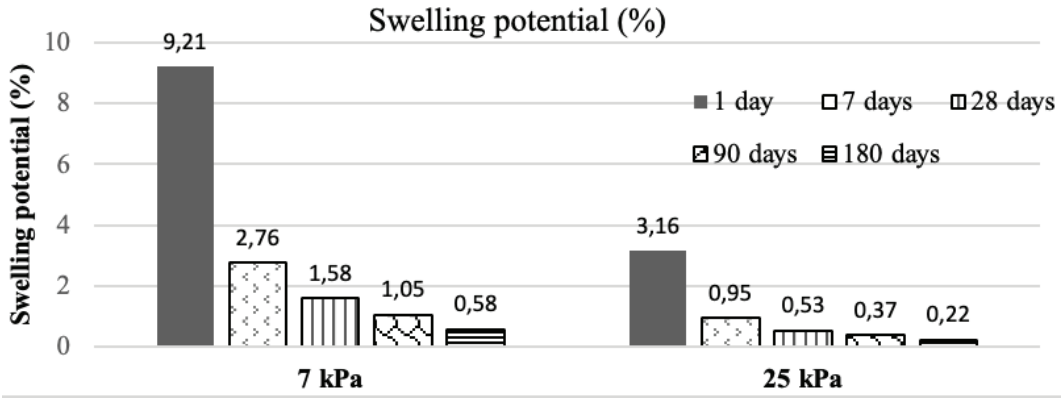


Figure 7. Swelling potentials of USIT specimens.

Alazigha et al. (2018) stated that CaLS lime pile stabilization is due to the migration of Ca 2+ ions from the lime piles into the clay, flocculation of particles, pozzolanic reactions, and stabilizing mechanisms of CaLS admixture with smearing and agglomeration of clay particles [31]. Additional improvement (decrease in swell percent) in this study in USIT compared to CaLS lime pile stabilization is due to the effect of fabric and soil nailing.

The swelling potential of the USIT specimen was improved with the curing period. While the swelling potential of USIT specimens under 7 kPa was measured as 1.58%, the swelling potentials of treated specimens with prewetting, LLP, and GSN techniques separately were determined as 30.1%, 7.11%, and 41.5%. When the techniques that were mentioned above were separately applied to KB, the reduction of the swelling potential of KB was either very little or insufficient. However, these techniques were applied together and then the swelling potential of KB was decreased from 43.95% to 1.58% after 28 day curing period. Under 25 kPa surcharge pressure, the swelling potential of KB treated with USIT in this study was determined as 0.53% after 28 day curing period.

California Bearing Ratio (CBR) Swell Test

California Bearing Ratio (CBR) swell test has been widely done for determining the swell percent of soil used in pavement design. USIT specimen was tested at the soaked CBR conditions. This test specimen has been waited in the water for four days to measure its swelling potential. Since the test specimen formed high plasticity clay that had high swelling potential. Two different surcharge pressures that are 7 kPa and 25 kPa were applied at both the soaked condition and test stage.

USIT specimen prepared for soaked CBR swell test had 19 pieces of piles that had 1.05 cm diameter and six pieces

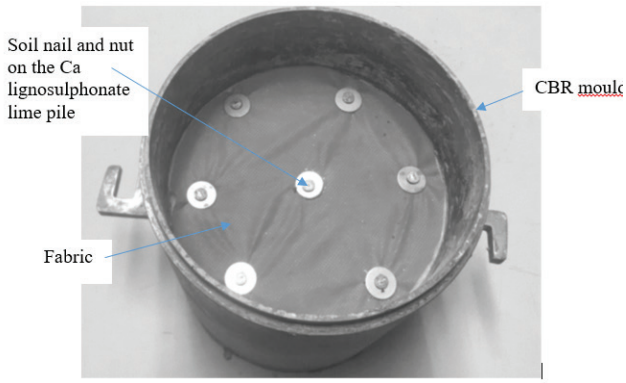


Figure 8. USIT specimen in CBR mold prepared for CBR swell test.

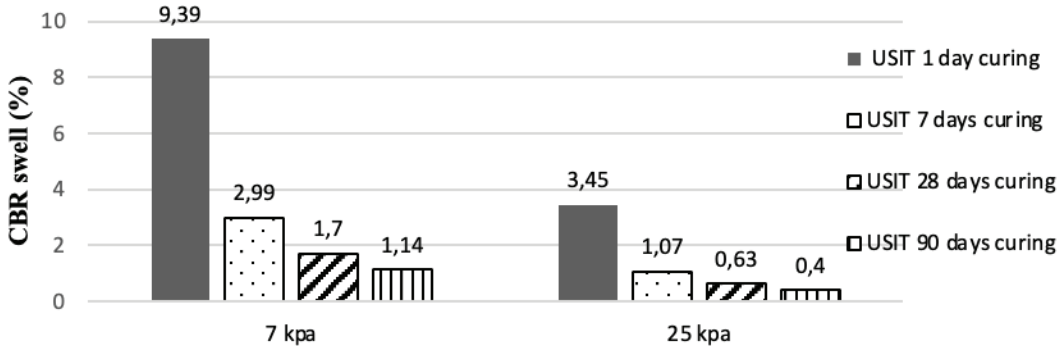


Figure 9. CBR swell values of USIT specimens.

of nails that had 6 cm length. Both diameters of the pile and the length of the nails have adjusted to the mold of the CBR test. In addition to this, six pieces of nut (to model an anchor plate) are put on the nails to block the tearing of the fabric and to transfer swell pressure to lime piles. The specimen used in this step is illustrated in Figure 8.

These specimens had been kept in the humid room for 7 days, 28 days, and 90 days for investigating the curing effect of this improved method. After the curing period, the CBR swell values of the treated specimens were determined by the data obtained from these tests (Figure 9).

The CBR swell values of USIT decreased with the curing period for each surcharge pressure applied in this study. The CBR swell of the USIT specimens (after the soil stabilization/improvement), which waited in the humid room for 90 days, was measured as 1.14% under 7 kPa surcharge pressure and was measured as 0.4% under 25 kPa surcharge pressure and fulfilled the criteria for the subgrade material specified in Highways Technical Specification (HTS, 2013) (i.e. CBR swell % < 3% after 7 days curing) [28].

CONCLUSION

Unified Improvement Techniques (USIT) include three techniques that are prewetting, lignosulfonate lime piles, and fabric-soil nails. Firstly, the prewetting technique is the saturation step of the soil specimen with a mixture formed calcium lignosulfonate and water. The lignosulfonate lime pile technique (LLP) forms a rigid layer for fasteners (or soil nails and anchor plates). Then a piece of fabric was laid down over the expansive soil specimen and fasteners were used for the connection between the fabric and the lignosulphonate lime pile. The fabric and soil fasteners are used to ensure both uniform pressures and to transfer some part of the swell pressure to lime piles. To sum up, the expansive soil is improved in both vertical and horizontal directions.

The CBR swell of the USIT specimens which waited in the humid room for 90 days, fulfilled the criteria for the subgrade material specified in Highways Technical Specification (HTS, 2013) (i.e. CBR swell % < 3% after 7 days of curing).

Time is a significant factor for the stabilization of expansive soil with either prewetting or lignosulphonate lime piles methods, to solve this problem, the GSN technique was applied with both prewetting and lignosulphonate lime pile methods alone. Therefore, the improvement process of USIT in this study is faster and more effective than prewetting, lignosulphonate lime piles.

The treated specimen with USIT is more stable and resistant to volumetric change than both the other treated specimens and untreated specimens prepared in this study. The improvement percentage at the swelling potential of expansive soil specimens under 7 kPa surcharge pressure was approximately 96.5 % after 28 day curing period.

The significant drop in the swell potential value suggests that the Unified Soil Improvement Technique's performance is better than CaLS lime piles in the improvement of expansive clays.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Nelson JD, Chao KC, Overton DD, Nelson EJ. Foundation Engineering for Expansive Soils Foundation Engineering for Expansive Soils. 1st ed. New Jersey: Wiley; 2015. [CrossRef]

[2] Parhi PS, Garanayak L, Mahamaya M, Das SK. Stabilization of expansive soil using alkali-activated fly ash based geopolymer. In: Hoyos L, McCartney J, editors. Advances in Characterization and Analysis of Expansive Soils and Rocks. New York, USA: Springer; 2017. p. 36-50. [CrossRef]

[3] Nelson J, Miller DJ. Expansive Soils: Problems and Practice in Foundation and Pavement Engineering. New York: Wiley; 1992.

[4] Teng TCP, Mattox RM, Clisby MB. A study of active clays as related to highway design. Available at: https://trid.trb.org/view/285608. Accessed on Jul 4, 2024.

[5] Callebaut F, Gabriels D, Boodt MD. The effect of polymer structure on soil physicochemical properties and soil water evaporation. J Chem Technol Biotechnol 1979;29:723-729. [CrossRef]

[6] Rana D, Neale GH, Hornof V. Surface tension of mixed surfactant systems: Lignosulfonate and sodium dodecyl sulfate. Colloid Polym Sci 2002;280:775-758. [CrossRef]

[7] Yang D, Qiu X, Zhou M, Lou H. Properties of sodium lignosulfonate as a dispersant of coal water slurry. Energy Convers Manage 2007;48:2433-2438. [CrossRef]

[8] Ekinci CE, Ay S, Baykuş N, Ay A. Examination of the impact of lignin sulfonate based structure chemicals on fresh and hardened concrete. Pamukkale Univ Muh Bil Derg 2016;22:478-485. Turkish. [CrossRef]

[9] Tingle J, Santoni R. Stabilization of clay soils with non-traditional additives. Transp Res Rec 2003;1819:72-84. [CrossRef]

[10] Vakili AH, Salimi M, Lu Y, Shamsi M, Nazari Z. Strength and post-freeze-thaw behavior of a marl soil modified by lignosulfonate and polypropylene fiber: An environmentally friendly approach. Constr Build Mater 2022;332:127364. [CrossRef]

[11] Tingle JS, Newman JK, Larson SL, Weiss CA, Rushing JF. Stabilization mechanisms of non-traditional additives. Transp Res Rec 2007;1989-2:59-67. [CrossRef]

[12] Vinod JS, Indraratna B, Mahamud M. Stabilization of an erodible soil using chemical admixtures. Available at: https://ro.uow.edu.au/cgi/viewcontent.cgi?article=1721&context=eispapers. Accessed on Jul 4, 2024.

[13] Chen Q, Indraratna B. Shear behavior of sandy silt treated with lignosulfonate. Can Geotech J 2015;52(8):1180-1185. [CrossRef]

[14] Vakili AH, Kaedi M, Mokhberi M, bin Selamat MR, Salimi M. Treatment of highly dispersive clay by lignosulfonate addition and electroosmosis application. Appl Clay Sci 2018;152:1-8. [CrossRef]

[15] Rogers CDF, Glendinning S. Modification of clay soils using lime. In Proceedings, Seminar on Lime Stabilization; 1996 Sept 25; London, United Kingdom. 1996. p. 99-114.

[16] Muntohar AS. A laboratory test on the strength and load-settlement characteristic of improved soft soil using lime-column. Dinamika Teknik Sipil 2010;10:202-207.

[17] Subbarao RGV. Foundation practices and rehabilitation of structures on expansive soil. Available at: https://www.nbmcw.com/product-technology/construction-chemicals-waterproofing/waterproofing-repair-chemicals/foundation-practices-a-rehabilitation-of-structures-on-expansive-soils.html. Accessed on Jul 4, 2024.

[18] Özkan İ, Çokça E. The effect of lime pile parameters on the improvement of the swelling potential of expansive clay. Arab J Geosci 2022;15:1-13. [CrossRef]

[19] Gnana Prasanna G, Kumar M, Dhanaraj A, Raja Sekhar K. Shear strength behavior of the flyash treated lime columns on expansive soils. In proceedings of the Indian Geotechnical Conference; 2022 Jun 8; New Delhi, India. 2022.

[20] Subaida EA, Chandrakaran S, Sankar N. Laboratory performance of unpaved roads reinforced with woven coir geotextiles. Geotext Geomembr 2009;27:204-210. [CrossRef]

[21] Nguyen TT, Indraratna B. Experimental and numerical investigations into hydraulic behavior of coir fiber drain. Can Geotech J 2017;54:75-87. [CrossRef]

[22] Broda J, Grzybowska-Pietras J, Gawłowski A, Rom M, Przybylo S, Laszczak R. Application of wool geotextiles for the protection of steep slopes. Procedia Eng 2017;200:112-119. [CrossRef]

[23] Prambauer M, Wendeler C, Weitzenböck J, Burgstaller C. Biodegradable geotextiles - An overview of existing and potential materials. Geotext Geomembr 2019;47:48-59. [CrossRef]

[24] Tiwari N, Satyam N. An experimental study on the behavior of lime and silica fume-treated coir geotextile reinforced expansive soil subgrade. Eng Sci Technol Int J 2020;23:1214-1222. [CrossRef]

[25] Lazarte CA, Robinson H, Gómez JE, Baxter A, Cadden A, Berg R. Soil nail walls reference manual. Available at: https://www.fhwa.dot.gov/engineering/geotech/pubs/nhi14007.pdf. Accessed on Jul 4, 2024.

[26] Madhyannapu RS, Puppala AJ. Design and construction guidelines for deep soil mixing to stabilize expansive soils. J Geotech Geoenviron Eng 2014;140:04014051. [CrossRef]

[27] Seed HB, Woodward RJ Jr, Lundgren R. Prediction of swelling potential for compacted clays. J ASCE Soil Mech Found Div 1962;88:53-87. [CrossRef]

[28] Karayolları Genel Müdürlüğü. Karayolları teknik şartnamesi 2013. Available at: https://www.tamyol.com.tr/UserFiles/Content/KGM_Teknik_Sartnamesi_2013.pdf. Accessed on Jul 4, 2024.

[29] ASTM. Standard test methods for one-dimensional swell or collapse of soils. Available at: https://webstore.ansi.org/standards/astm/astmd454621. Accessed on Jul 4, 2024.

[30] Abbasi PA, Soltani N, Cuppels DA, Lazarovits G. Reduction of bacterial spot disease severity on tomato and pepper plants with foliar applications of ammonium lignosulfonate and potassium phosphate. Plant Dis 2002;86:1232-1236. [CrossRef]

[31] Alazigha DP, Indraratna B, Vinod JS, Heitor A. Mechanisms of stabilization of expansive soil with lignosulphonate admixture. Transp Geotech 2018;14:81-92. [CrossRef]

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1067–1074, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00088

Research Article

Effects of cementitious ingredients on long term properties of self compacting concrete

Yuvaraj L. BHIRUD[1,*], Om V. VAIDYA[1], Swapnil A. PATARE[1]

[1]*Department of Civil Engineering, SNJB's LSKBJ College of Engineering, Chandwad, Nashik, 423101, India*

ARTICLE INFO

Article history
Received: 28 October 2022
Revised: 09 February 2023
Accepted: 29 March 2023

Keywords:
Creep; Elastic Shortening; Self Compacting Concrete; Shrinkage

ABSTRACT

The paper "Properties of Self-Compacting Concrete with High Volume of Cementitious Material" investigates the impact of higher binder content on the properties of self-compacting concrete (SCC). The study was conducted using three SCC mixes with binding materials of 550 kg/m^3, 600 kg/m^3, and 650 kg/m^3 with a constant water-to-binder ratio and 40% of the binder being fly ash. The properties of the concrete were evaluated in the fresh stage using tests like the slump flow test and V-funnel test and in the hardened stage using tests like the compressive strength test and shrinkage test. The results showed that while there was a correlation between higher binder content and higher compressive strength, there were also differences in shrinkage and creep values that were estimated using ACI 209 R-92 and BS EN 1992-1-1-2004. The study highlights the need to investigate the effects of high binder content on SCC properties.

Cite this article as: Bhirud YL, Vaidya OV, Patare SA. Effects of cementitious ingredients on long term properties of self compacting concrete. Sigma J Eng Nat Sci 2024;42(4):1067–1074.

INTRODUCTION

Self-compacting concrete (SCC) is a type of concrete that can be placed without the need for external efforts such as vibration, floating, or poking. Its flowing nature requires it to be stable against segregation, which can be achieved by using chemicals such as Viscosity Modifying Agents (VMA type SCC) or a high amount of binder materials like GGBS, fly ash, or limestone powder (Powder type SCC). This change in ingredient proportions results in a change in various properties of SCC compared to normal vibrated concrete (NVC) [1-6]. Among these properties, shrinkage, creep, and elastic shortening can cause losses in prestress force.

In India, fly ash is abundant [8] and its disposal contributes to environmental pollution. Utilizing this available fly ash through high volume fly ash SCC is a way to reduce this pollution. Studies have shown both similarities and differences between the properties of SCC and NVC. While some research has shown that the creep, shrinkage, and elastic modulus properties of SCC are comparable to NVC when the strength of both is kept constant [1, 10, 27-29] other studies have shown that NVC has less shrinkage and creep than SSC [5, 7, 30-32].

***Corresponding author.**
*E-mail address: yuvarajbhirud@gmail.com

This paper was recommended for publication in revised form by Regional Editor in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey

Several codes have been used to predict the shrinkage, creep, and elastic modulus of SCC, with varying degrees of accuracy. ACI 209 R [13] has been shown to accurately predict SCC shrinkage by Jin-Kenn Kim et al [5] and Stefanus A Kristiawan et al [12], while B. Barr et al [17] found that it is accurate for low strength concrete but differs when the concrete strength exceeds 45 N/mm^2. Euro Code 2 [15] has been shown to accurately forecast the elastic modulus of concrete by Lino Maia et al [18], and to accurately forecast SCC creep if the stress level is below 30% of strength.

While studies have been carried out for SCC with cement and filler material quantities of up to 550 kg/m^3, there is limited literature on the behavior of SCC with cementitious substance of 600 kg/m^3 or more. Thus, there is a need to investigate the effect of higher binder content on properties such as shrinkage, creep, and elastic shortening. To address this gap, an experimental investigation was conducted to study the properties of SCC with a high volume of cementitious material. Three mixes with binding materials of 550 kg/m^3, 600 kg/m^3, and 650 kg/m^3 were prepared with the same water-to-binder ratio. The workability, segregation resistance at the fresh stage, and compressive strength, elastic modulus, shrinkage, creep, and elastic shortening at the hardened stage were determined. The shrinkage and creep values were estimated using ACI 209 R-92 and BS EN 1992-1-1-2004, and compared with experimental values obtained in the study.

EXPERIMENTAL PROGRAM

The self-compacting concrete properties of workability, resistance to segregation, and flowability were evaluated in the fresh stage using the Slump Flow Test, V-Funnel Test, and L-Box Test. The mixture proportions were determined based on the EFNARC (European federation dedicated to specialist construction chemicals and concrete systems) guidelines [11] and previous research findings [19-22]. The permissible limits for acceptance of SCC at fresh stage by EFNARC guidelines are mentioned in last column of table 3. The specific surface area of aggregate were determine based on IS 4030(Part-II)-1988. Three SCC mixes with binding materials of 550 kg/m^3, 600 kg/m^3, and 650 kg/m^3 were prepared, with 40% substitute binder being fly ash (SCC550, SCC600, and SCC650, respectively). Class F fly ash and OPC 53 grade cement were used as binders, and crushed basalt stone aggregates with a maximum size of 20mm and fine natural river sand were employed as the coarse and fine aggregates, respectively. The water-to-binder ratio was kept constant for all mixtures, and a polycarboxylate ether superplasticizer was used to enhance workability. Table 1 shows properties of ingredients used while the mixture proportions are presented in Table 2.

In order to study the mechanical properties of concrete, the elasticity, shrinkage, creep, and elastic shortening moduli were evaluated using concrete specimens in their hardened state. To assess the compressive strength, 150mm cubes were produced and evaluated at seven, twenty-eight,

Table 1. Binding substance and aggregate properties

Types	Cement	Fly Ash	Fine aggregates	Coarse aggregates	
				10 mm	20mm
Specific Surface area	325 m^2/kg	374 m^2/kg	-	-	-
Specific gravity	3.15	2.3	2.79	2. 9	2.9
Fineness Modulus	-	-	2.93	6.09	8.11

Table 2. Mix proportion used for concrete

		Mix		
		SCC550	SCC600	SCC650
Cement (kg)		330	360	390
Fly ash (40%) (kg)		220	240	260
Natural sand (kg)		1130	1090	1060
Coarse aggregates (kg)	10 mm (70%)	321	308	359
	20 mm (30%)	137	136	154
Water/Binder ratio		0.33	0.33	0.33
Super-plasticizer		1.30%	0.95%	0.85%

Figure 1. Shrinkage test setup

surfaces of the specimen at the prestressing reinforcement location with epoxy adhesive (Figure 2).

The prestressing cable was gradually released after 7 days of casting at a rate of 0.8 kN/s and the strain gauge readings were taken just before and after cable release. The difference between these two readings is defined as the elastic shortening (ε_e).

After the release of the tendon, the total strain (ε_t) that occurred in the specimens was recorded periodically. This total strain includes shrinkage strain, elastic shortening and creep. From this total strain, shrinkage strain, which was measured separately, and elastic shortening is subtracted to get creep strain and given in equation 1.

$$\varepsilon_c = \varepsilon_t - \varepsilon_s - \varepsilon_e \quad (1)$$

Readings were taken once a week for the first 4 weeks and then every four weeks thereafter. The shrinkage and creep measurements were taken up to 112 days of age to obtain a comprehensive understanding of the concrete's behaviour over time.

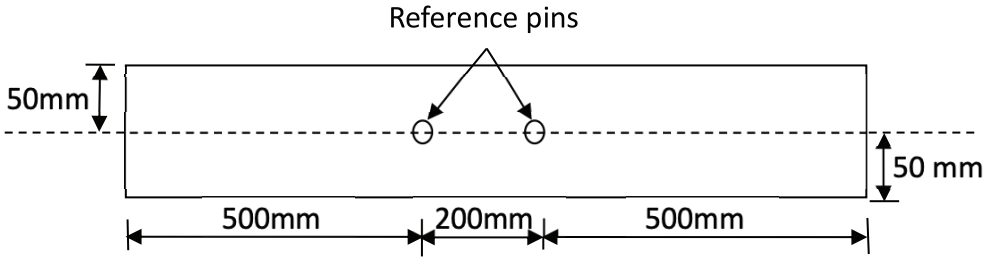


Figure 2. Positioning of the reference pins and the prestressing wire on sample.

and fifty-six days of age. The elasticity modulus was determined using 150mm diameter cylinders.

Prismatic specimens with a cross section of 100x100mm were used to determine shrinkage. The shrinkage behaviour of the self-compacting concrete samples was evaluated using a dial gauge with an accuracy of 0.001mm. The samples were vertically stacked and kept in a controlled environment with a temperature of 20±1°C and relative humidity of 50%. In this research endeavour, the objective was to determine the total shrinkage of the concrete samples (Figure 1).

100x100x1200mm prismatic specimens were specifically designed for the pretension prestressing system. In the study, a 12.5mm diameter high tensile strand wire with a tensile strength of 1750 N/mm² was utilized throughout the experiment. The cable was positioned at the centre of the sample cross section (100x100x1200mm) to eliminate eccentricity and ensure that the concrete section remains under compressive stress when the prestressing cable is released. Strain gauge points were attached to the lateral

RESULTS AND DISCUSSION

Fresh Characteristics of Self-Compacting Concrete

In this study, the fresh characteristics of self-compacting concrete were evaluated. Initially, trials were conducted on the SCC550 mix, and upon achieving satisfactory results, the percentage of superplasticizer was adjusted for the specific performance of the other mixes. The workability results are presented in Table 3. It was observed that as the binder concentration increased, the required amount of superplasticizer decreased. The SCC550 required 1.3% superplasticizer, the SCC600 required 0.95%, and the SCC650 required 0.85%. The higher binder concentration resulted in an increase in the available paste, which provided better lubrication and reduced the need for superplasticizer to achieve the desired workability. The visual examination of the concrete flow provided insight into segregation, as significant segregation results in a concentration of coarse particles in the centre of the concrete flow, while only the cement paste reaches the edges. The Slump flow test revealed that the coarse aggregates

Table 3. Results for workability of concrete

Test	Mix			Permissible Limits
	SCC550	SCC600	SCC650	
Slump flow diameter (mm)	680	710	690	600 - 750
V-funnel Time T_0 (Sec)	8.2	7.2	7.42	6 - 12
V-funnel at $T_{5minute}$ (Sec)	10.9	9.9	10.25	£T_0 + 3
L-box T_{20cm} (Sec)	0.89	0.9	0.92	1 ± 0.5
L-box T_{40cm} (Sec)	2.34	2.2	2.09	2 ± 0.5
L-box blocking ratio (H_2/H_1)	0.86	0.9	0.85	0.8 – 1.0

were evenly distributed within the concrete flow, and the presence of coarse aggregates and mortar paste at the edge of the flow indicated the stability of the mix. These findings are illustrated in Figure 3.

Hardened Characteristics of Self-Compacting Concrete

(a) Compressive Strength of Self-Compacting Concrete

The compressive strength of each of the SCC mixes, SCC550, SCC600, and SCC650, was evaluated and the results are presented in Figure 4. The results showed that the compressive strength of the concrete increased with an increase in the concentration of binding material. The compressive strength of SCC550 was found to be the lowest among the three mixes, while that of SCC650 was the highest at all ages. This can be attributed to the fact that SCC600 and SCC650 mixes contain a higher concentration of binding material, thus resulting in greater compressive strength.

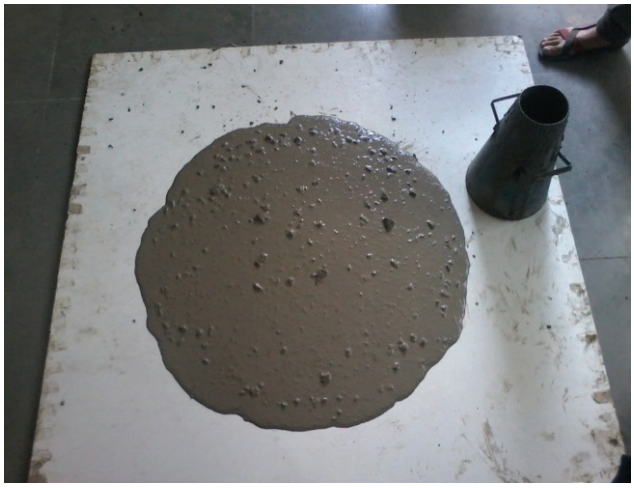


Figure 3. Slump flow of concrete.

(b) Elastic modulus

The elastic modulus of the concrete samples at 28 days of age is depicted in Figure 5. The results indicate a slight increase in the elastic modulus for SCC600 and SCC650 mixes compared to SCC550. The elastic modulus of concrete is commonly estimated by multiplying a constant by the square root of its characteristic strength ($\sqrt{f_{ck}}$). The ACI 318-89 [16] recommends the relation 4700 $\sqrt{f_{ck}}$ for predicting concrete's elastic modulus. Table 4 presents the results of the experimentally obtained constants for the calculation of the elastic modulus by multiplying with $\sqrt{f_{ck}}$ It can be seen that the constant decreases as the powder content in the SCC mix increases. Higher powder content results in a larger volume of paste and a smaller quantity of coarse aggregates, causing the aggregates to be stiffer than the paste [9], [23]. This leads to a higher strain in the concrete at the same stress level, causing the constant to decrease.

The experimental values and the predicted values of the modulus of elasticity using the ACI 318-89, ACI 363R-84, and BS EN 1992-1-1-2004 methods are shown in Figure 4.

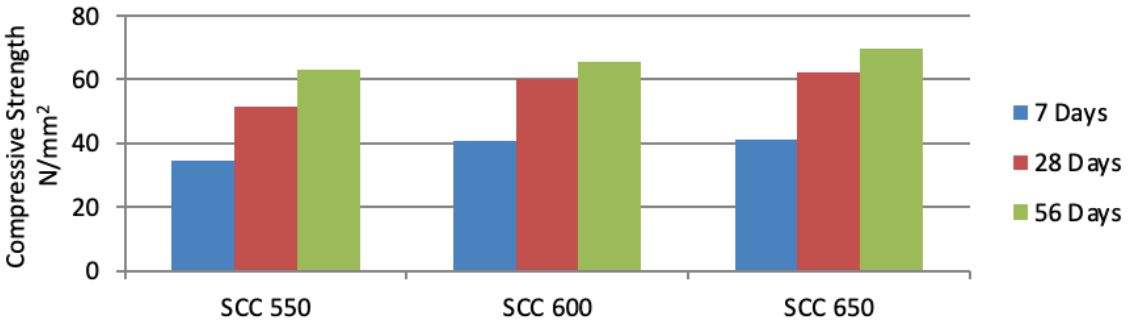


Figure 4. Compressive strength of concrete at various ages.

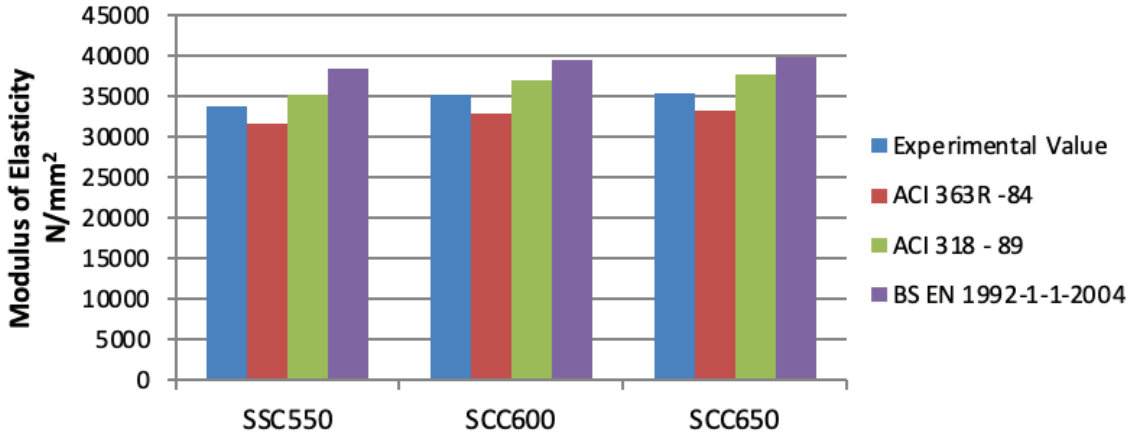


Figure 5. Modulus of elasticity comparison between experimental and projected values

Table 4. Multiplying constant for $\sqrt{f_{ck}}$

Mix	Value of multiplying constant for √fck
SSC550	4503
SCC600	4464
SCC650	4410

The results indicate that the estimated elastic modulus values using the ACI 363R-89 approach are found to be lower, while the values obtained using the ACI 318-89 and BS EN 1992-1-1-2004 methods are higher than the experimental values. The experimental analysis reveals a high degree of agreement between the predicted and measured modulus of elasticity using the ACI 363R-89 approach.

(b) Shrinkage, Elastic Shortening and Creep

The results of the shrinkage, as measured by the change in length, are shown in Figure 6. It was observed that the shrinkage increased with an increase in the binder content of the concrete. In comparison to SCC550 at 112 days, the shrinkage for SCC600 was 13.04% higher and for SCC650 it was 34.78% higher. The predicted values of shrinkage based on ACI 209 R-92 and BS EN 1992-1-1-2004 were found to be lower than the experimental values. The difference between the predicted and experimental values of shrinkage was found to be 17% for SCC550 concrete based on ACI 209 R-92 and 54% based on BS EN 1992-1-1-2004. The gap between the predicted and experimental values of shrinkage was found to be highest for SCC650 concrete, where the difference was 34%.

The results obtained for elastic shortening are presented in Figure 7. The results indicate that concrete with higher binder content has a higher elastic shortening, with SCC600 and SCC650 having 6.52% and 13.25% more elastic shortening, respectively, compared to SCC550.

After the release of the tendon, the total strain (ε_t) that occurred in the specimens was recorded periodically. Readings were taken once a week for the first month and then every four weeks thereafter till for total 112 days. The creep strain (ε_c) was calculated using Equation 1.

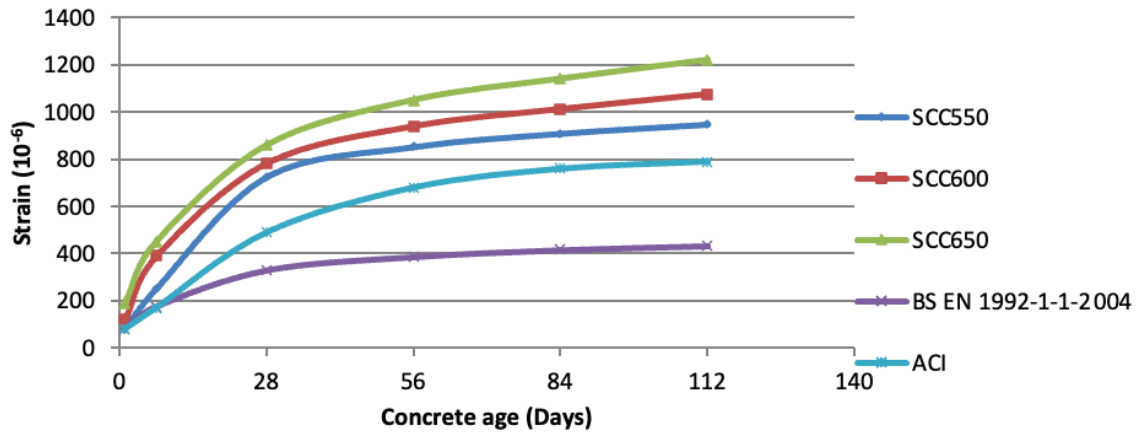


Figure 6. Comparison of shrinkage of various mixes.

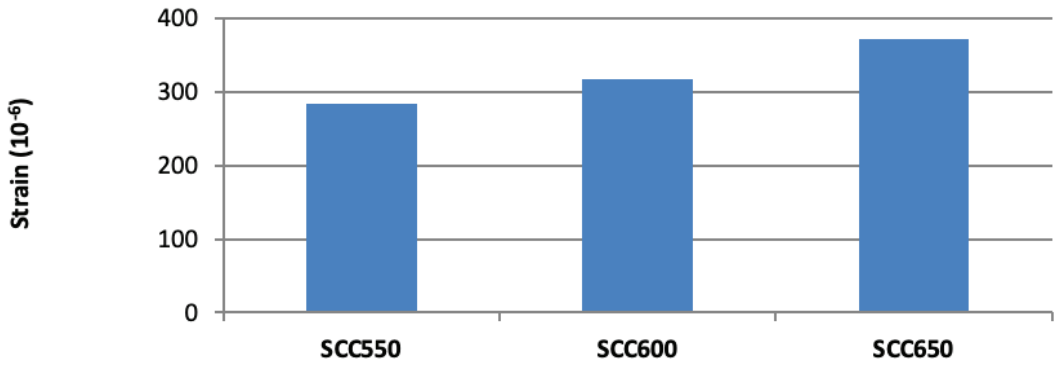


Figure 7. Elastic shortening of mixes.

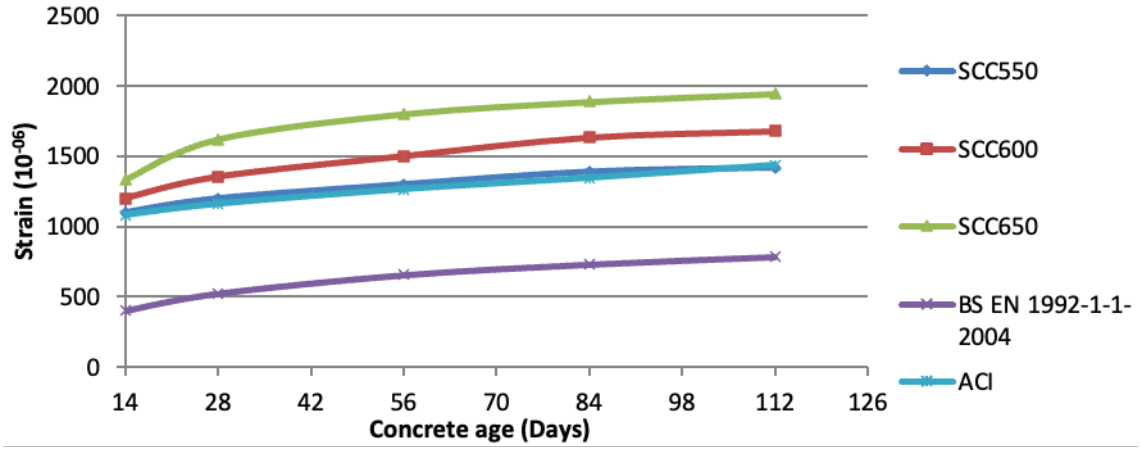


Figure 8. Comparison of creep of mixes.

Table 5. Binding material and coarse aggregate ratio in concrete

Mix	Binding material ratio to concrete	Coarse aggregates ratio to Concrete
SCC550	0.233	0.234
SCC600	0.254	0.223
SCC650	0.275	0.217

Figure 8 shows the results of creep in concrete and indicates that concrete with a higher binding material has more creep. The results also show that the rate of creep in the first week is rapid for all mixes and decreases in subsequent weeks. In comparison with SCC550, creep for SCC600 is 18.3% more and for SCC650 it is 36.6% more at 112 days of concrete age. The predicted values for creep by BS EN 1992-1-1-2004 for SCC550 are lower than the experimental values, but the predicted values by ACI 209 R-92 are nearly identical to the experimental values. However, for SCC600 and SCC650, the experimental values are more than the prediction by ACI 209 R-92 by 14% and 24% respectively.

Table 5 presents the data on the relationship between the amounts of binding material and coarse aggregates in concrete. The results indicate that as the amount of binding material increases, the proportion of coarse aggregate decreases. This shift in the composition of concrete has significant implications on its properties.

Firstly, an increase in the volume of paste in concrete results in finer pore structures and smaller capillaries. This process is known as self-desiccation, which leads to a higher surface tension and greater attraction forces between the pores. This results in higher shrinkage of concrete, as the increased surface tension causes the pores to pull together [24, 26].

Secondly, the decrease in coarse aggregate volume and increase in paste volume also contributes to increased creep and elastic shortening. These properties can cause concrete to deform over time, affecting its strength and durability.

In conclusion, the results from Table 5 highlight the importance of the balance between binding material and coarse aggregates in concrete. An increase in the volume of binding material results in a reduction of coarse aggregate, which can lead to increased shrinkage, creep, and elastic shortening. These factors must be considered when designing concrete mixtures to ensure their durability and performance over time.

CONCLUSION

In summary, this study investigated the impact of varying binder content on the properties of Self-Compacting Concrete (SCC) mixes. Three SCC mixes were cast with binder contents of 550 kg/m^3, 600 kg/m^3, and 650 kg/m^3, respectively, and all mixes were stored under identical conditions. The results showed that all mixes met the workability requirements of SCC, and that the amount of superplasticizer required decreased with an increase in binder content. Additionally, all mixes demonstrated stability against segregation.

The modulus of elasticity was found to decrease with an increase in binder content, and the predicted values by the ACI 363R-89 method were in close agreement with experimental values. Meanwhile, the predicted values by the ACI 318-89 and BS EN 1992-1-1-2004 methods were higher compared to the experimental values.

The shrinkage was found to increase with an increase in binder content, and the estimated values by the ACI 209 R-92 and BS EN 1992-1-1-2004 codes were lower than the experimentally measured values. The difference between the predicted and measured values increased with the powder content, reaching 34% for the SCC650 mix.

The study also found that creep increased with an increase in binder content. The creep predicted by the ACI 209 R-92 method was in close agreement for the SCC550 mix, but was lower for the higher binder content mixes (SCC600 and SCC650). Similarly, the estimated values by the BS EN 1992-1-1-2004 method were lower than the measured values for all mixes.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Persson B. A comparison between properties of self-compacting concrete and the corresponding properties of normal concrete. Cem Concr Res 2001;31:193–198. [CrossRef]

[2] Aslani F, Nejadi S. Mechanical properties of conventional and self-compacting concrete: An analytical study. Constr Build Mater 2012;36:330–347. [CrossRef]

[3] Uysal M, Yilmaz K. Effect of mineral admixtures on properties of self-compacting concrete. Cem Concr Compos 2011;33:771–776. [CrossRef]

[4] Georgiadis AS, Sideris KK, Anagnostopoulos NS. Properties of SCC produced with limestone filler or viscosity modifying admixture. J Mater Civ Eng 2010;22:352–360. [CrossRef]

[5] Kim JK, Han SH, Park YD, Noh JH. Material properties of Self-flowing concrete. J Mater Civ Eng 1998;10:244–249. [CrossRef]

[6] Gesoglu M, Guneyisi E, Ozbay E. Properties of self-compacting concretes made with binary, ternary, and quaternary cementitious blends of fly ash, blast furnace slag, and silica fume. Constr Build Mater 2009;23:1847–1854. [CrossRef]

[7] Roziere E, Granger S, Turcry Ph, Loukili A. Influence of paste volume on shrinkage cracking and fracture properties of self-compacting concrete. Cem Concr Compos 2007;29:626–636. [CrossRef]

[8] Central Electricity Authority. Report on fly ash generation at coal/lignite based thermal power stations and its utilization in the country for the year 2014-15. Available at: https://cea.nic.in/wp-content/uploads/2020/04/flyash_final_1415.pdf. Accessed on Jun 6, 2024.

[9] Dinakar P, Reddy MK, Sharma M. Behaviour of self compacting concrete using Portland pozzolana cement with different levels of fly ash. Mater Des 2013;46:609–616. [CrossRef]

[10] Turcry P, Loukili A, Haidar K, Pijaudier-Cabot G, Belarbi A. Cracking tendency of Self-Compacting Concrete subjected to restrained shrinkage: Experimental study and modelling. J Mater Civ Eng 2006;18:46–54. [CrossRef]

[11] The Self-Compacting Concrete European Project Group. The European Guidelines for self compacting concrete - Specification, production and use. Available at: https://www.theconcreteinitiative.eu/images/ECP_Documents/EuropeanGuidelinesSelfCompactingConcrete.pdf. Accessed on Jun 6, 2024.

[12] Kristiawan SA, Aditya MTM. Effect of high volume fly ash on shrinkage of Self compacting concrete. Procedia Eng 2015;125:705–712. [CrossRef]

[13] ACI Committee 209. Prediction of creep, shrinkage and temperature effect in concrete structures. Available at: http://civilwares.free.fr/ACI/MCP04/209r_92.pdf. Accessed on Jun 6, 2024.
[14] ACI Committee 363. State-of-the-art report on high-strength concrete. Available at: https://www.silicafume.org/pdf/reprints-363rtoc.pdf. Accessed Jun 6, 2024.
[15] European Committee for Standardization. Design of concrete structures - Part 1-1: General rules and rules for buildings. Available at: https://www.phd.eng.br/wp-content/uploads/2015/12/en.1992.1.1.2004.pdf. Accessed on Jun 6, 2024.
[16] ACI Committee 318. Building code requirements for structural concrete (ACI 318-95) and commentary (ACI 318-89). Available at: https://myyardimci.weebly.com/uploads/1/6/3/4/16347790/building_design_code_english.pdf. Accessed on Jun 6, 2024.
[17] Barr B, Hoseinian SB, Beygi MA. Shrinkage of concrete stored in natural environments. Cem Concr Res 2003;25:19–29. [CrossRef]
[18] Maia L, Figueiras J. Early-age creep deformation of a high strength self-compacting concrete. Constr Build Mater. 2012;34:602–610. [CrossRef]
[19] Naik TR, Kumar R, Rammeb BW, Canpolat F. Development of high-strength, economical self-consolidating concrete. Constr Build Mater 2012;30:463–469. [CrossRef]
[20] Li Z, Ohkubo T, Tanigawa Y. Flow performance of high-fluidity concrete. J Mater Civ Eng 2004;16:588–596. [CrossRef]
[21] El Chabib H, Syed A. Properties of self-consolidating concrete made with high volumes of supplementary cementitious materials. J Mater Civ Eng 2013;25:1579–1586. [CrossRef]
[22] Khaleel OR, Al-Mishhadani SA, Abdul Razak H. The effect of coarse aggregate on fresh and hardened properties of self-compacting concrete (SCC). Procedia Eng. 2011;14:805–813. [CrossRef]
[23] Parra C, Valcuende M, Gomez F. Splitting tensile strength and modulus of elasticity of Self-compacting concrete. Constr Build Mater 2011;25:201–207. [CrossRef]
[24] Valcuende M, Marco E, Serna P. Influence of limestone filler and viscosity-modifying admixture on the shrinkage of self-compacting concrete. Constr Build Mater 2012;42:583–592. [CrossRef]
[25] Leemann A, Lura P, Loser R. Shrinkage and creep of SCC - The influence of paste volume and binder composition. Constr Build Mater 2011;25:2283–2289. [CrossRef]
[26] Bymaster JC, Dang CN, Floyd RW, Hale WM. Prestress losses in pretensioned concrete beams cast with lightweight self-consolidating concrete. Structures 2015;2:50–57. [CrossRef]
[27] Savija B, Lukovic M, Schlangen E. Creep, shrinkage and cracking of self-compacting and conventional vibrated concrete under drying and autogenous shrinkage. Cem Concr Res 2019;120:67–78.
[28] Belheouane M, Khelil N, Kenai S. Shrinkage and creep of self-compacting and vibrated high-strength concrete. J Build Eng. 2020;31.
[29] Yin G, Li Q, Wang L. Experimental investigation on the shrinkage and creep behavior of self-compacting concrete. Adv Mater Sci Eng 2021
[30] Şahin B, Meran C, Gurun E. Shrinkage and creep of normal and self-compacting concrete in different curing conditions. Constr Build Mater 2019;211:348–356.
[31] Erdogan Y, Aydin S, Edil TB. Effect of vibration on shrinkage and creep behavior of self-compacting and normal vibrated concrete. Constr Build Mater. 2020;258.
[32] Dash DP, Kumar P, Patra RC. Time-dependent behavior of self-compacting and normal vibrated high-strength concrete under sustained loading. Constr Build Mater. 2021;282.

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1075–1087, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00089

Research Article

MHD natural convection in a square enclosure using carbon nanotube-water nanofluid with two isothermal fins

Mohamed EL HATTAB[1,2,*], Mustapha BOUMHAOUT[2], Soufiane OUKACH[2]

[1]*Department of Mechanics, Process of Energy and Environment Laboratory, ENSA, Ibn Zohr University, Agadir, 80000, Morocco*
[2]*Research Team, Energy and Sustainable Development, ESTG, Ibn Zohr University, Guelmim, 80000, Morocco*

ARTICLE INFO

Article history
Received: 20 January 2023
Revised: 03 March 2023
Accepted: 13 May 2023

Keywords:
CNT-Water Nanofluid; Magnetic Field; Natural Convection; Thin Fin

ABSTRACT

This paper reports the numerical study of natural convection in a square enclosure filled with CNT-water nanofluid and exposed to a uniform external magnetic field. Heating is ensured by twothin fins. Using the control volume method, the effects of the fins position, their length and spacing as well as the solid volume fraction, the Rayleigh number and the Hartmann number on the thermal performance of the cavity were examined. The results obtained show that the heat transfer rate increases with the Rayleigh number, solid volume fraction and fins length; but decreases with Hartmann numbers. A comparison is also carried out between the results obtained from the Maxwell and Xue models. The results prove that the mean Nusselt number is higher based on the Xue model.

Cite this article as: El Hattab M, Boumhaout M, Oukach S. MHD natural convection in a square enclosure using carbon nanotube-water nanofluid with two isothermal fins. Sigma J Eng Nat Sci 2024;42(4):1075–1087.

INTRODUCTION

Heat transfer phenomena are of decisive importance for the study and operation of devices such as steam generators, ovens, heat exchangers, condensers, etc. Due to economic and environmental issues, the efficient use of a facility (heat exchange) with minimum energy expenditure is the aim in all cases. In this context, many studies have focused on heat transfer by natural convection [1-2], considering its involvement in various industrial systems and processes, such as buildings insulation, solar energy collection, cooling of heat-generating components in the electrical, etc. The improvement of the thermal transfer in enclosures, with the introduction of fins fixed on the walls, has thus been intensely studied theoretically and experimentally in recent years because the applications concerned are extremely varied. Shi and Khodadadi [3] carried out a numerical investigation of natural convection in a square enclosure with a single thin fin placed at the hot wall. It was observed that the heat transfer increases with increasing Rayleigh number and when the fin is placed closer to the adiabatic walls. Heat transfer by convection in a square enclosure with a thin conductive fin located on the hot wall was studied numerically by Tasnim and Collins [4]. The results obtained showed that the heat transfer rate improves by approximately 31.46%, when the fin is fixed on

***Corresponding author.**
*E-mail address: elhattab.1970@gmail.com

This paper was recommended for publication in revised form by Editor in-Chief Ahmet Selim Dalkilic

the hot wall, comparing with a wall without fin for $Ra=10^4$. The influence of the Rayleigh number, the length and the position of the thin fin as well as the ratio of the conductivities (fin/fluid) on the heat transfer by free convection within a differentially heated cavity is examined by Bilgen [5]. It has been found that the Nusselt number increases with Rayleigh number, but decreases with conductivity ratio and fin length. Ben-Nakhi and Chamkha [6] numerically investigated the influence of the length and angle of inclination of a fin on heat transfer by free convection in a square enclosure. The fin was placed at the middle of the hot wall. It was found that the average Nusselt number is sensitive to variation in inclination angle and fin length. An experimental and numerical study on natural convection in a square cavity with two isolated baffles is carried out by Nardini et al. [7]. The cavity with plexiglas walls is heated by four discrete sources. The obtained results show that the flow and heat transfer characteristics are profoundly influenced by different fins lengths. The recent numerical study by Attouchi et al. [8] focused on the analysis of natural convection in a cavity with finned surface where one surface is kept at periodic wall temperature. Their results indicate that the best heat transfer rate is obtained for the case of enclosure with three fins at the hot side wall.

To further improve heat transfer, another innovative technique is to use nanofluids. The latter are obtained by adding nanoparticles to a base fluid in order to improve the thermal properties of the mixture. Nanofluids based on classical nanoparticles such as Al_2O_3, CuO, TiO_2, Cu, AgO, AgO are intensively studied [9-13]. One of the developments in nanofluid is carbon nanotubes (CNTs). The thermal conductivity of CNTs is very high compared to other nanoparticles and, therefore, their use can significantly improve the thermal performance of energy installations. For these reasons, studies on CNT-based nanofluids have been conducted by Ul Haq et al. [14], Tayebi et al. [15,16] and Noranuar et al. [17]. Recently, review studies on heat transfer application of carbon-based nanofluid are carried out by Borode et al. [18] and Ali et al. [19].

Sometimes undesirable factors can alter the beneficial phenomenon of natural convection. Among these undesirable factors is the presence of a magnetic field acting on electrically conductive fluids. Indeed, the Lorentz force significantly influences the heat transfer and the flow of cooling fluids. Ghasemi et al. [20] have numerically studied the natural convection of a water-Al_2O_3 nanofluid confined in an enclosure exposed to a magnetic field. Their results show that heat transfer rate increases with increasing Rayleigh number but decreases as Hartmann number increases. Sourtiji et al. [21] studied the effect of the magnetic field on the heat transfer by natural convection in L-shaped enclosures filled with nanofluid. The problem of natural convection and entropy generation of nanofluids confined in a cavity maintained at a sinusoidal temperature distribution and subjected to a magnetic field is studied numerically by Mejri et al. [22], using the lattice Boltzmann method. Belhaj and Ben-Beya [23] performed a numerical study on heat transfer by natural convection within a square cavity filled with CNT-water nanofluid and subjected to the action of a constant magnetic field. The enclosure is heated from below with a sinusoidal temperature distribution. Hamid et al. [24] studied numerically by the finite element method the hydromagnetic flow and heat transfer of water-based carbon nanotubes (CNTs) inside a rectangular fin-shaped cavity. Recently, Sarala et al. [25] have studied the effects of MHD alumina-water nanofluid flow past an oscillating vertical plate in the presence of radiation and Hall effects.

We note that the majority of research work is carried out on numerical simulation of the flow and thermal transfer of nanofluids based on classical nanoparticles such as Cu, Al_2O_3, TiO_2, etc. However, little studies has been done on nanofluids based on carbon nanotubes. In this context, the present work constitutes a contribution to the numerical study of natural hydromagnetic convection of CNT-water nanofluid confined in an enclosure heated by two isothermal fins.

PHYSICAL MODEL AND MATHEMATICAL FORMULATION

Figure 1 illustrates the configuration studied. It is a square enclosure, filled with water containing different concentrations of single walled carbon nanotubes (SWCNTs) and heated by two horizontal thin fins. The two horizontal walls and the left vertical wall are adiabatic, whereas the right wall is operated at a lower temperature (T_C). The fins are located at the left wall and maintained at high temperature (T_H). (L), (s) and (d) are representing lengths, positions and spacing of two fins, respectively. A horizontal uniform magnetic field B_o is imposed. The nanofluid is assumed to be incompressible and the density in the buoyancy force is evaluated by the Boussinesq approximation [26].This model considers density constant in all basic equations except the buoyancy term in the momentum equation. The equation below obtained using the Boussinesq approximation [26]:

$$\rho_{nf} = \rho_0[1 - \beta(T - T_0)] \tag{1}$$

Table 1 presents the thermophysical properties of water and SWCNT [27]. The viscous dissipation and joule heating are neglect.

Therefore, the conservation equations, in non dimensional form, are established below in two-dimensional [20]:

$$\frac{\partial U}{\partial X} + \frac{\partial V}{\partial Y} = 0 \tag{2}$$

$$U\frac{\partial U}{\partial X} + V\frac{\partial U}{\partial Y} = -\frac{\partial P}{\partial X} + \frac{\mu_{nf}}{\rho_{nf}\alpha_f}\left(\frac{\partial^2 U}{\partial X^2} + \frac{\partial^2 U}{\partial Y^2}\right) \tag{3}$$

$$U\frac{\partial V}{\partial X} + V\frac{\partial V}{\partial Y} = -\frac{\partial P}{\partial Y} + \frac{\mu_{nf}}{\rho_{nf}\alpha_f}\left(\frac{\partial^2 V}{\partial X^2} + \frac{\partial^2 V}{\partial Y^2}\right) + \frac{(\rho\beta)_{nf}}{\rho_{nf}\beta_f}Ra Pr\theta - Ha^2 Pr V \tag{4}$$

$$U\frac{\partial\theta}{\partial X}+V\frac{\partial\theta}{\partial Y}=\frac{\alpha_{nf}}{\alpha_f}\left(\frac{\partial^2\theta}{\partial X^2}+\frac{\partial^2\theta}{\partial Y^2}\right) \quad (5)$$

Where, the following non dimensional parameters are used:

$$X=\frac{x}{H},\quad Y=\frac{y}{H},\quad U=\frac{uH}{\alpha_f},\quad V=\frac{vH}{\alpha_f},$$
$$P=\frac{pH^2}{\rho_{nf}\alpha_f{}^2},\quad \theta=\frac{T-T_C}{T_H-T_C} \quad (6)$$
$$Ra=\frac{g\beta_f H^3(T_H-T_C)}{\upsilon_f\alpha_f},\quad Pr=\frac{\upsilon_f}{\alpha_f},\quad Ha=B_0H\sqrt{\frac{\sigma_{nf}}{\rho_{nf}\upsilon_f}}$$

The non dimensional boundary conditions are as follows:

- On the right wall: $U=V=0,\ \theta=0$ (7.a)
- On the left wall: $U=V=0,\ \frac{\partial\theta}{\partial X}=0$ (7.b)
- On the bottom wall: $U=V=0,\ \frac{\partial\theta}{\partial Y}=0$ (7.c)
- On the top wall: $U=V=0,\ \frac{\partial\theta}{\partial Y}=0$ (7.d)
- On the fins: $U=V=0,\ \theta=1$ (7.e)

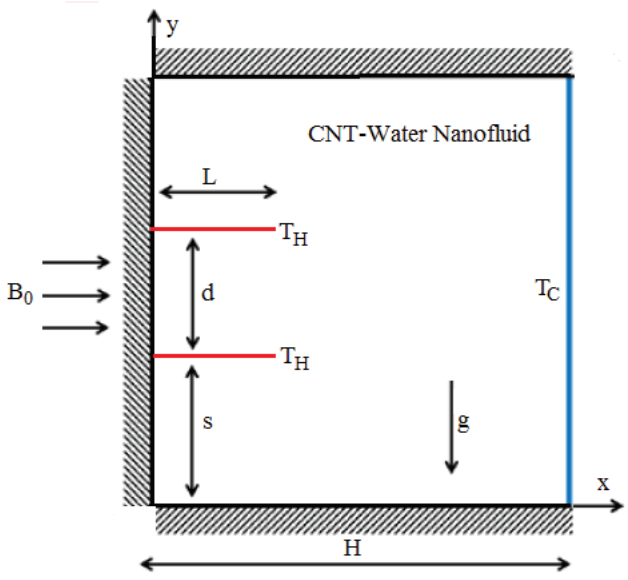


Figure 1. Physical configuration.

The properties of the nanofluid are calculated as follows [28,20]:

$$\rho_{nf}=(1-\phi)\rho_f+\phi\rho_p \quad (8)$$

$$(\rho\beta)_{nf}=(1-\phi)(\rho\beta)_f+\phi(\rho\beta)_p \quad (9)$$

$$(\rho Cp)_{nf}=(1-\phi)(\rho Cp)_f+\phi(\rho Cp)_p \quad (10)$$

$$\alpha_{nf}=\frac{k_{nf}}{(\rho Cp)_{nf}} \quad (11)$$

$$\sigma_{nf}=(1-\phi)\sigma_f+\phi\sigma_p \quad (12)$$

The dynamic viscosity and thermal conductivity of the nanofluid can beestimated by Brinkman [29] and Maxwell [30] models, respectively.

$$\mu_{nf}=\frac{\mu_f}{(1-\phi)^{2.5}},\ k_{nf}=k_f\left[\frac{(k_p+2k_f)-2\phi(k_f-k_p)}{(k_p+2k_f)+\phi(k_f-k_p)}\right], \quad (13)$$

The stream function is evaluated as following:

$$U=\frac{\partial\psi}{\partial Y}\ \text{and}\ \ V=-\frac{\partial\psi}{\partial X} \quad (14)$$

Local and average Nusselt numbers along the cold wall are evaluated as:

$$Nu_Y=-\frac{k_{nf}}{k_f}\frac{\partial\theta}{\partial X}\bigg|_{X=1} \quad (15)$$

$$Nu_m=\int_0^1 Nu_Y\,dY \quad (16)$$

NUMERICAL METHODOLOGY

The system of equations (2) - (5) associated with the boundary conditions (7.a) - (7.e) are discretized by the control volume method established by Patankar [31] employing the power law scheme. The velocity-pressure couplingis realised by the SIMPLE algorithm [31]. The tridiagonal matrix algorithm method is used to solve algebraic equations iteratively.

Table 1. Properties of water and SWCNT at 298 K [27]

	ρ (kg.m^{-3})	C_p (J.kg^{-1}.K^{-1})	k (W.m^{-1}.K^{-1})	β (K^{-1})	σ (S.m^{-1})
Pure water	997.1	4179	0.613	21×10^{-5}	0.05
SWCNT	2600	425	6600	0.16×10^{-5}	4.8×10^{7}

Table 2. Influence of mesh size on Nu_m et $|\psi|_{max}$

Grid size (X×Y)	Nu_m	Error %	$\|\psi\|_{max}$	Error %
41×41	6.9373	–	13.3122	–
61×61	6.8159	1.750	13.1135	1.492
81×81	6.7596	0.826	13.0382	0.574
101×101	6.7272	0.479	13.0083	0.229
121×121	6.7090	0.270	12.9921	0.124
141×141	6.6971	0.177	12.9843	0.060

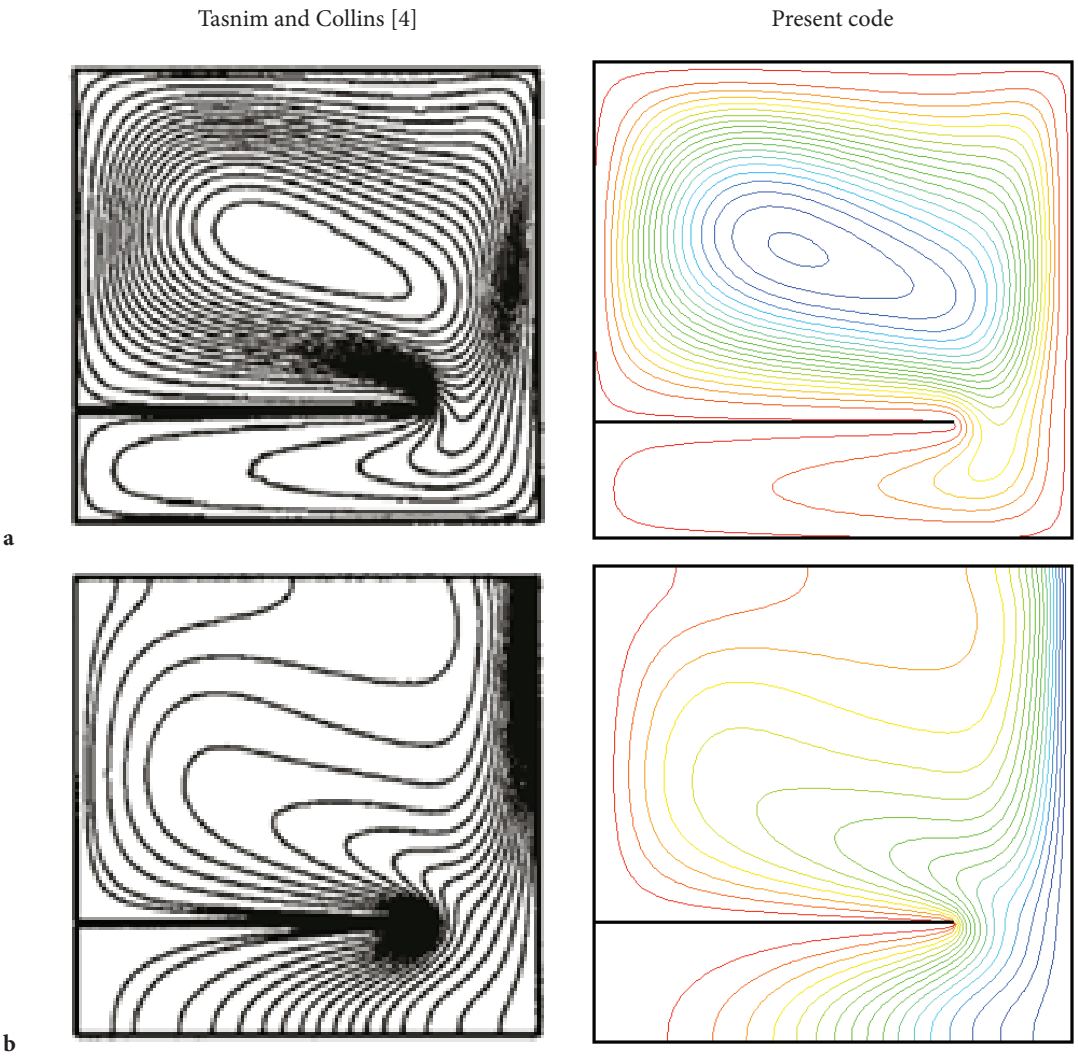


Figure 2. (a) Streamlines, (b) isotherms for $Ra=10^5$ and Pr=0.7.

Table 2 regroup the tests carried out to follow the sensitivity of the results to the mesh and this, for $Ra=10^6$, Ha=30, ϕ=0.03, L=0.25, S=0.4, D=0.2 and it appears that Nu_m and $|\psi|_{max}$ become insensitive to the number of nodes from the grid 121×121. In the continuation of our work, the 121×121 mesh is chosen to perform the simulations due to the precision/computation time compromise.

To validate the computer code developed and programmed in FORTRAN language, we compare our results with those of the literature. First, we qualitatively compared the structure of the flow and the isotherms obtained by Tasnim and Collins [4], for the case of free convection in a square enclosure with a baffle placed on the hot wall. From the Figure 2, a great similarity is observed between the

streamlines and the isotherms, obtained by our computer code and those of the reference [4].

In addition, a quantitative validation is carried out by confronting our results with those obtained by Ghasemi et al. [20] and Teamah et al. [32], who numerically treated the case of natural convection within a closed square cavity, filled with a nanofluid and exposed to a magnetic field. The comparison illustrated in Figure 3 shows the evolution of the mean Nusselt number as a function of the Hartmann number. It is clear from this figure, that there is a very satisfactory agreement between our results and those obtained by the authors.

RESULTS AND DISCUSSION

In this study, the Rayleigh number (Ra), the Hartmann number (Ha), the volume fraction of nanoparticles (ϕ), the fins position (S), the fins length (L), and the spacing of two fins (D) are considered to be within the following ranges: $10^3 \le Ra \le 10^6$, $0 \le Ha \le 60$, $0 \le \phi \le 0.06$, $0.25 \le L \le 0.75$, $0.1 \le S \le 0.7$, and $0.2 \le D \le 0.8$.

Effect of Hartmann Number (Ha) and Rayleigh Number (Ra)

The streamlines and isotherms are plotted in Figure 4a and 4b, respectively, for Rayleigh number (Ra $=10^4$, 10^5, 10^6) and Hartmann number (Ha=0, 30, 60). It is assumed that ϕ=0.03, L=0.25, S=0.4, and D=0.2. Figure 4a clearly shows the buoyancy driven circulating flows in the cavity for all Rayleigh and Hartmann numbers. As the Hartmann number increase, the strength of the circulations decreases. Indeed, the horizontal magnetic field generates a Lorentz force which acts in the vertical direction and in the opposite direction to the buoyancy force, which has result of reducing the flow intensity of the nanofluid in the enclosure. On the other hand, the circulation cell decomposes to double-eye pattern at Ha=60 for Ra=10^4 and 10^5, however this behavior disappears when the buoyancy force becomes strong at high Rayleigh number Ra=10^6. The results also show that the application of the magnetic field affects the isotherms (Figure 4b), in particular at Ra $=10^5$. As the Hartmann number increases, the isotherms straighten almost vertically. This is an indication of transformation of heat transfer mode almost to conduction. At Ra=10^6, the effect of the Lorentz force becomes less significant and the convection mode is evident. A thermal stratification is observed in the center of the cavity and below the bottom fin.

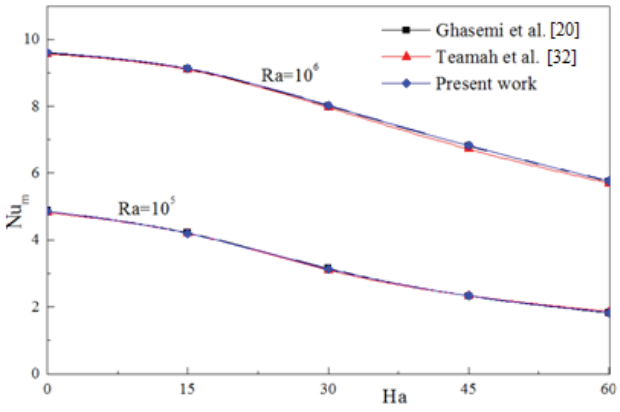


Figure 3. Comparison with the results of Ghasemi et al. [20] and Teamah et al. [32].

Figure 5 presents the influence of the Lorentz force, expressed by the variation of the Hartmann number, on the velocity (Figure 5 (a), (b) and (c)) and temperature (Figure 5 (d), (e) and (f)) profiles along the horizontal mid-span of the cavity at three Rayleigh numbers (Ra $=10^4$, 10^5 and 10^6) and for a volume fraction ϕ =0.03. It can be observed from the figure, that the maximum velocity increases under the effect of the buoyancy force, when the Rayleigh number increases and decreases when the Hartmann number increases, due to the influence of the magnetic force on the flow. The velocity profiles show that the flow between the two fins is almost stagnant for Ra=10^4 and 10^5. This is due to the convection flux which is weak. The corresponding temperature profiles show that the temperature at this location varies slowly. Increasing the Rayleigh number to 10^6 allows the nanofluid to circulate between the two fins and greatly reduces the effect of the magnetic field which manifests for the other values of the Rayleigh numbers.

The variation of the mean Nusselt number ratio ($Nu_m/Nu_{m,Ha=0}$) with the Rayleigh number at different Hartmann numbers is shown in Figure 6. It is assumed that ϕ =0.03. The $Nu_m/Nu_{m,Ha=0}$ is insensitive to the variation of the Hartmann number at Ra=10^3 because the heat transfer regime is dominated by conduction. It is clearly seen that for all Hartmann numbers, when the Rayleigh number increases up to Ra=10^4, the $Nu_m/Nu_{m,Ha=0}$ decreases because the magnetic field reduces the effect of convection flows. Moreover, we notice that for a given Rayleigh number, except Ra=10^3, the $Nu_m/Nu_{m,Ha=0}$ decreases when Hartmann number increases.

Effect of Position (S), Length (L) and Spacing (D) of Thin Fins

Figure 7 presents the streamlines (left) and the isotherms (right) for Ra=10^5 and for different fins positions (S=0.1, 0.3, 0.5, and 0.7). We observe from the figure that, when the two fins are placed very close to the bottom wall (S=0.1), the center of the recirculation cell is almost in the middle of the cavity. As the position increases, the center of the recirculation cell moves right up and the flow intensity decreases (see Table 3) due to flow clogging between the top wall and the top fin. At S=0.7, the fluid in the lower part of the cavity becomes almost motion less and colder, as shown in Figure 7.

Figures 8–10 illustrate, respectively, the effects of the fins position, the fins length and the spacing of two fins on the mean Nusselt number at different Rayleigh numbers,

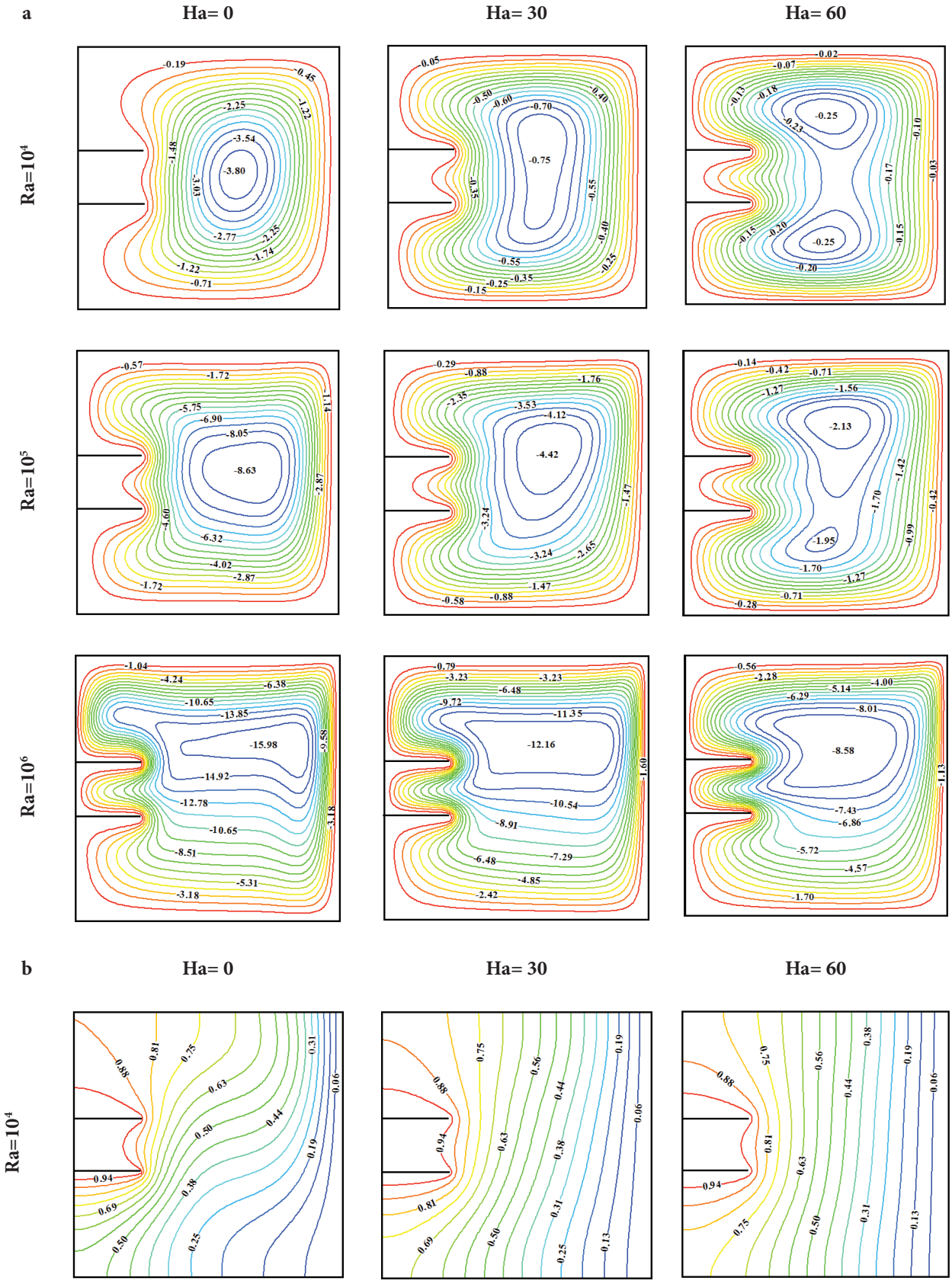


Figure 4. Continued

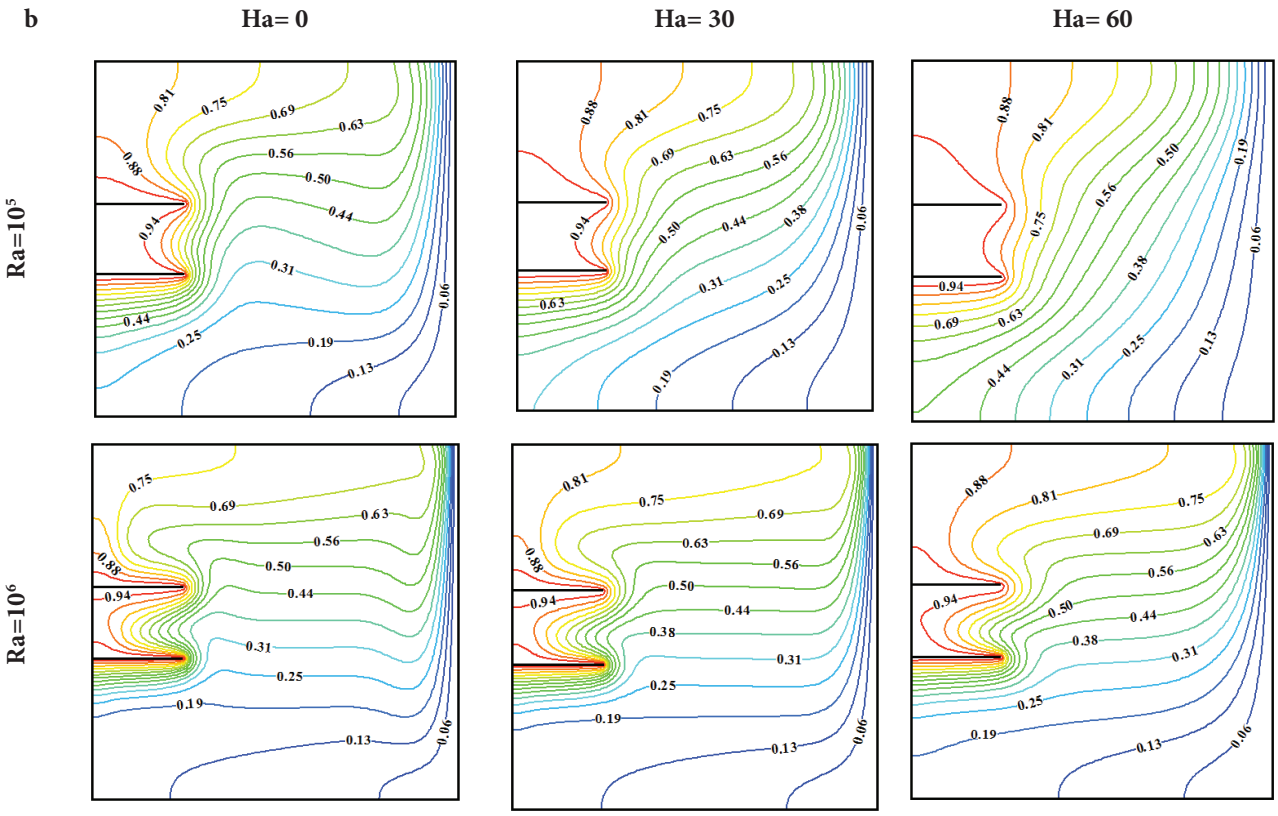


Figure 4. (a) Streamlines, (b) Isotherms for different Rayleigh and Hartmann numbers.

for ϕ=0.03 and Ha=30. It can be seen from Figure 8 that at any position S, mean Nusselt number increases as Rayleigh number increases. At low Rayleigh numbers (Ra=10^3 and 10^4), the variation in position has no significant effect on the heat transfer rate. However, at high Rayleigh numbers (Ra=10^5 and 10^6), where heat transfer is dominated by convection, the influence of position on the heat transfer rate is discernible. Indeed, the mean Nusselt number increases with increasing position until it reaches a maximum then it begins to decrease with further increase in position. As shown in Table 3 and Figure 8, the maximum mean Nusselt number occurs at S=0.3.

The influence of fins length on mean Nusselt number is presented in Figure 9. It can be observed that the heat transfer rate increases with Rayleigh number for any fins length, as the increase in the Rayleigh number improves buoyancy force and therefore augments the heat transfer rate. The mean Nusselt number is also improved with increasing fins length. This improvement can be attributed to the increase in heat transfer surface area as the length of the fins increases. The effect of fins spacing on the mean Nusselt number is shown in Figure 10. We note from the figure that the heat transfer rate increases with Rayleigh number independently of the value of fins spacing on the one hand, and that the variation of the mean Nusselt number with the fins spacing is not significant at low Rayleigh numbers, on the other hand. The maximum mean Nusselt number corresponds respectively to spacing D=0.5 and 0.3 for Ra=10^5 and 10^6.

Effect of Solid Volume Fraction (ϕ)

To examine the effect of nanoparticle concentration on heat transfer rate, the ratio between the mean Nusselt number and mean Nusselt number without volume fraction ($Nu_m/Nu_{m,\ \varphi=0}$) is illustrated in Figure 11. In the figure, for a given Ra, the Nusselt number ratio increases with increasing volume fraction. This enhancement is due to the improvement of the nanofluid thermal conductivity as the volume fraction increases. The results also show that for Ra=10^3, the volume fraction has a major effect on the heat transfer rate, however it has a minor effect at Ra=10^5. For example, when the volume fraction increases from 0 to 0.06 , the mean Nusselt number enhances about 21% at Ra=10^3 and about 1% at Ra=10^5.

Maxwell Versus Xue Models

This section presents acomparison between two models that are used to calculate the thermal conductivity of nanofluids: Maxwell model [30] and Xue model [33] which is defined as follows:

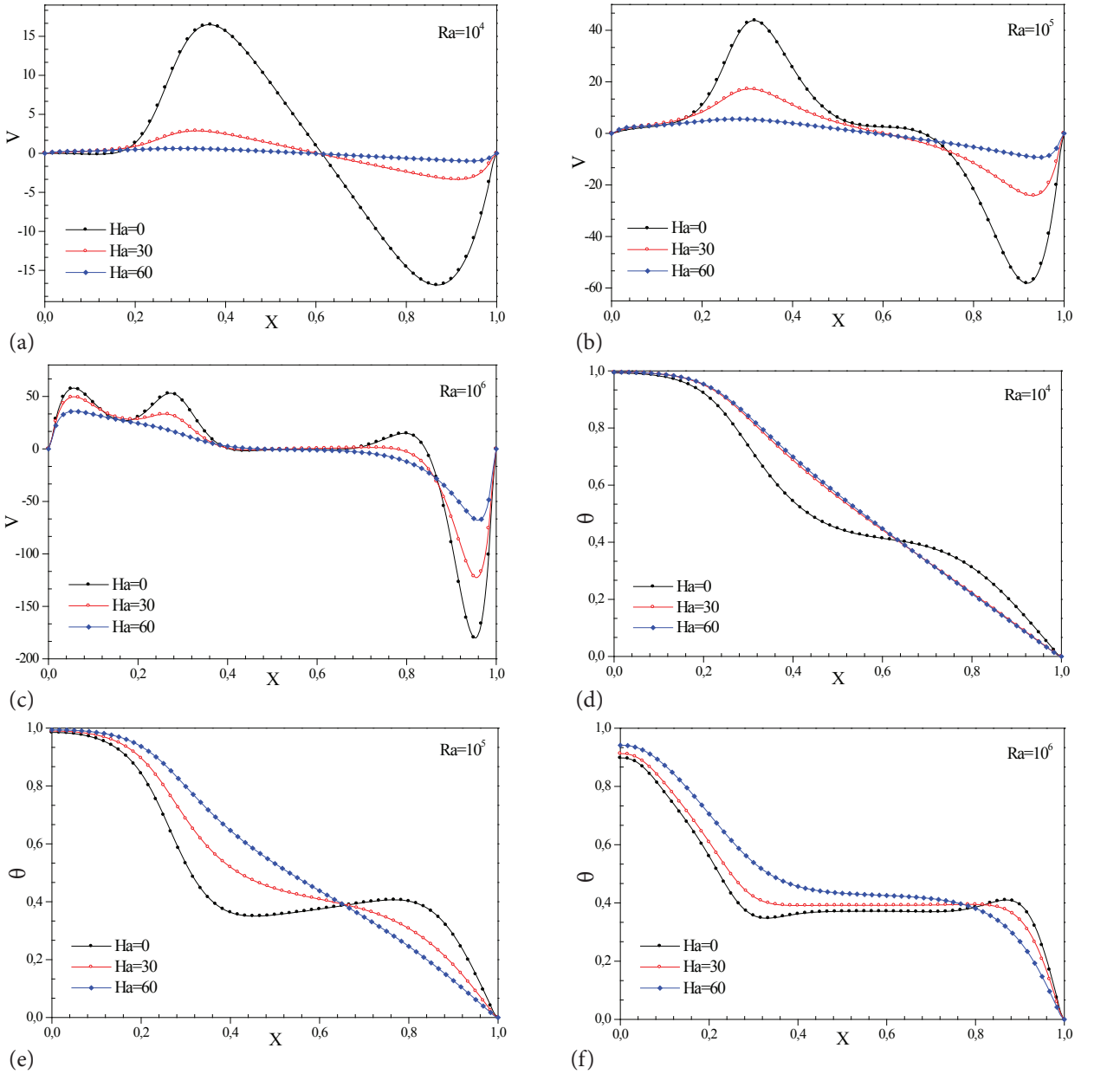


Figure 5. (a), (b) and (c) vertical velocity profile, (d), (e) and (f) temperature profile at Y = 0.5 for ϕ = 0.03, L = 0.25, S = 0.4 and D = 0.2.

$$k_{nf} = k_f \left[\frac{1 - \phi + 2\phi \frac{k_p}{k_p - k_f} ln \frac{k_p + k_f}{2k_f}}{1 - \phi + 2\phi \frac{k_f}{k_p - k_f} ln \frac{k_p + k_f}{2k_f}} \right] \quad (17)$$

In fact, the models used to evaluate the properties of nanofluids and in particular those relating to thermal conductivity, considerably affect the thermal performance of nanofluids [34, 35]. The theoretical model proposed by Xue is based on Maxwell theory and takes into consideration the rotating elliptical nanotubes with a important axial ratio [17]. The results obtained so far are based on the Maxwell model [30].

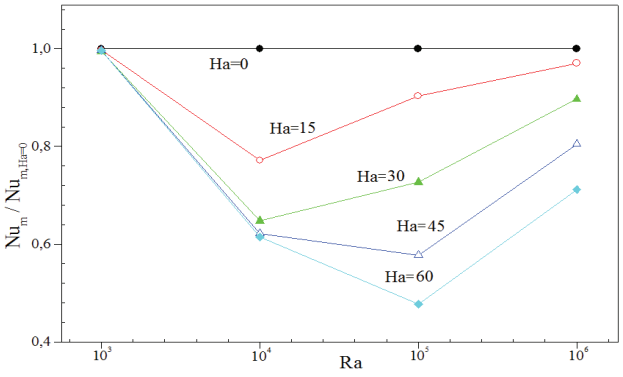


Figure 6. Evolution of mean Nusselt number ratio with Ra for L=0.25, S=0.4 and D=0.2.

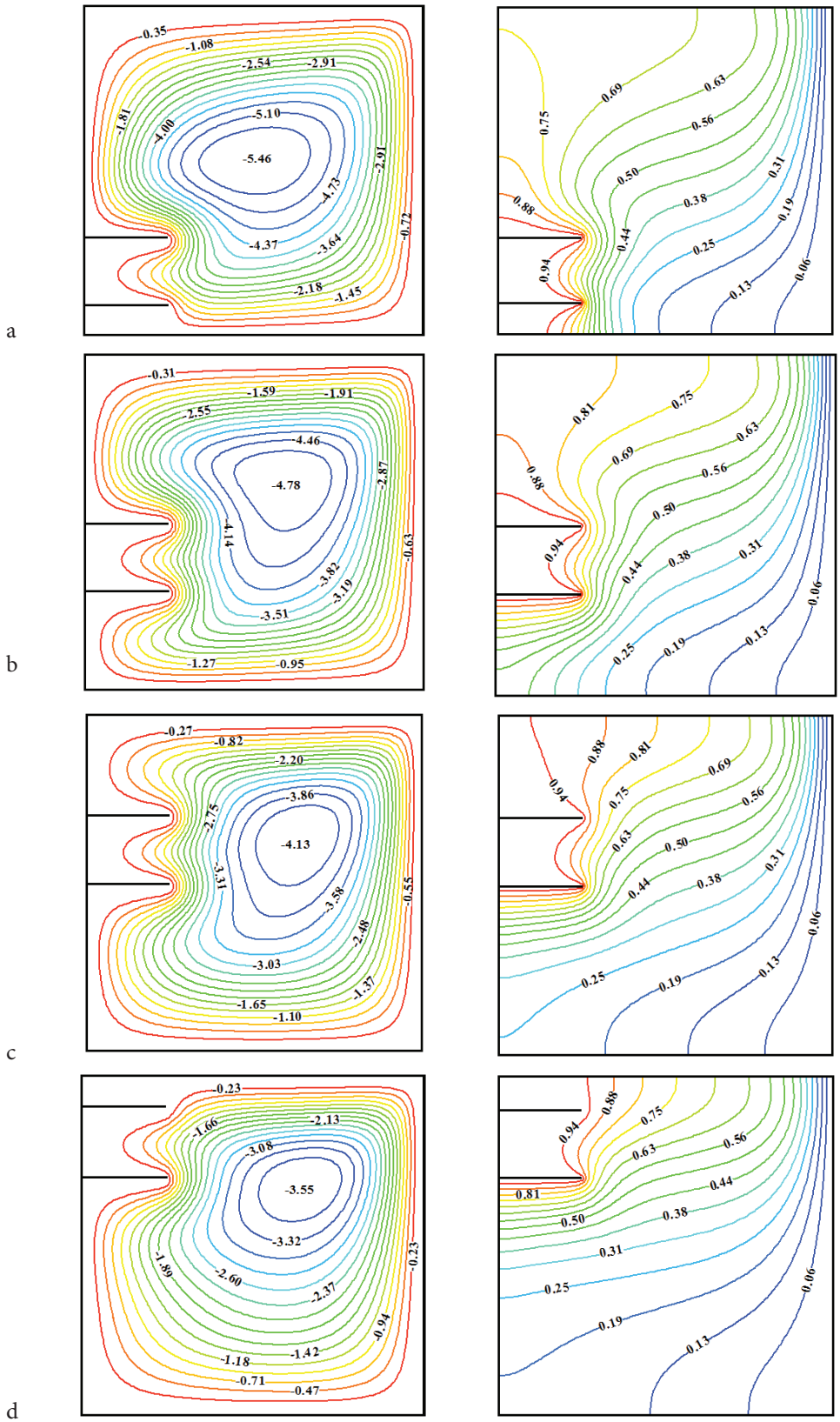


Figure 7. Streamlines (left) and isotherms (right) for Ra=10^5, Ha=30, L=0.25, D=0.2 and ϕ=0.03, (a) S=0.1, (b) S=0.3, (c) S=0.5, (d) S=0.7.

Table 3. Nu_m and $|\psi|_{max}$ at various Ra and S (L=0.25, D=0.2, ϕ =0.03 and Ha=30)

		S=0.1	S=0.3	S=0.5	S=0.7
$Ra=10^3$	Nu_m	1.0904	1.2021	1.1753	1.0420
	$\|\psi\|_{max}$	0.0915	0.0877	0.0857	0.0873
$Ra=10^4$	Nu_m	1.1772	1.2763	1.2527	1.1138
	$\|\psi\|_{max}$	0.9609	0.8772	0.7914	0.7758
$Ra=10^5$	Nu_m	2.6945	2.8288	2.5418	2.0462
	$\|\psi\|_{max}$	5.8383	5.1110	4.4148	3.7932
$Ra=10^6$	Nu_m	6.8154	7.0009	6.2250	4.6662
	$\|\psi\|_{max}$	17.8165	14.7509	11.6565	9.2761

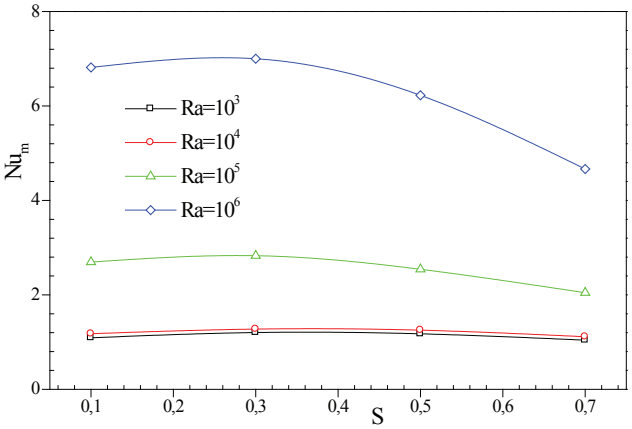


Figure 8. Evolution of Nu_m with S at different Ra for L=0.25, D=0.2and Ha=30.

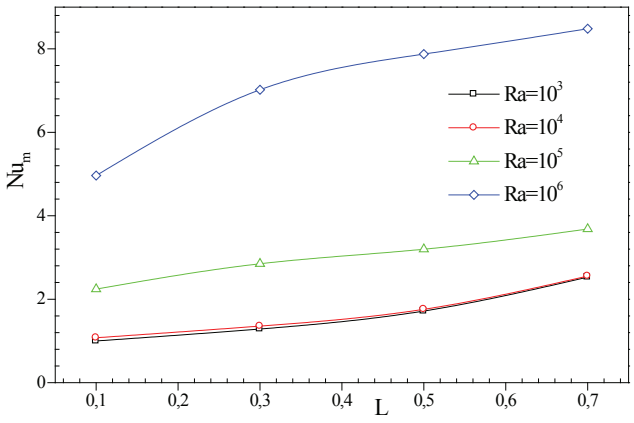


Figure 9. Evolution of Nu_m with L at different Ra for S=0.4, D=0.2 and Ha=30.

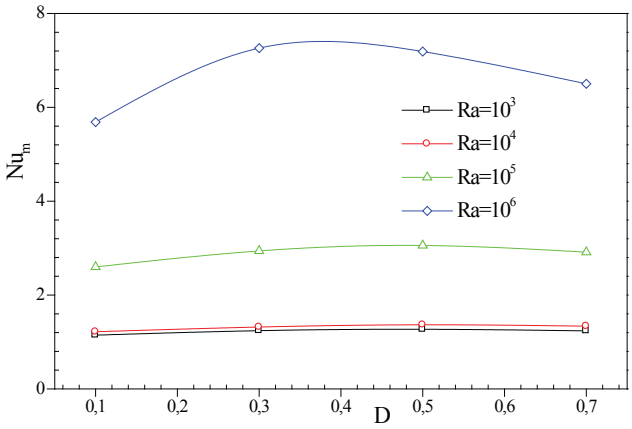


Figure 10. Evolution of Nu_m with D at different Ra for L=0.25 and Ha=30.

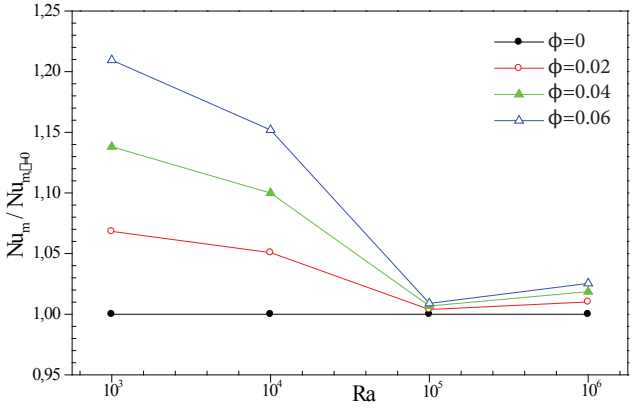
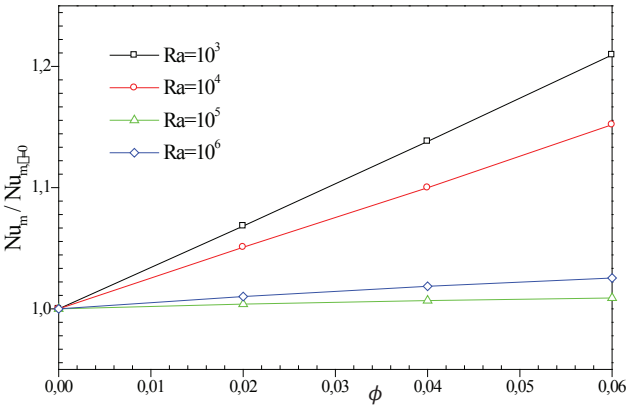
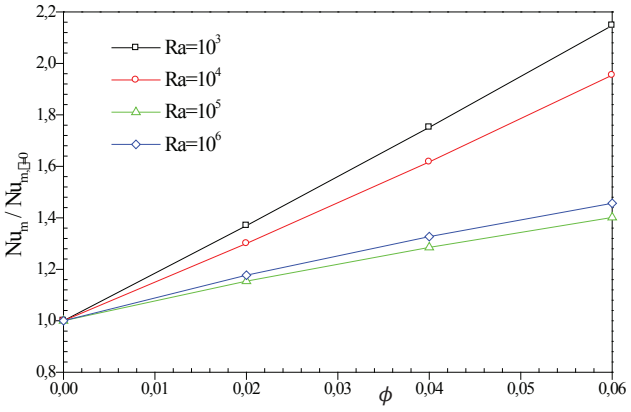


Figure 11. Evolution of mean Nusselt number ratio with Ra for Ha=30, L=0.25, S=0.4 and D=0.2

Figure 12 examines the variation in the mean Nusselt number ratio ($Nu_m/Nu_{m,\varphi=0}$) with the volume fraction at different Rayleigh numbers, for the two models. The results, which are presented for Ha=30, L=0.25, S=0.4, and D=0.2 show that the mean Nusselt number ratio obtained with Xue's model is higher than that obtained using Maxwell's model. Furthermore, the results reveal that for all Rayleigh numbers and for both models, when the volume fraction increases the mean Nusselt number ratio increases. This increase is



(a)



(b)

Figure 12. Evolution of mean Nusselt number ratio with ϕ for Ha=30, L=0.25, S=0.4, and D=0.2, (a) Maxwell model, (b) Xue model.

accentuated at low Rayleigh numbers because the introduction of nanoparticles improves the thermal conductivity of nanofluid which increases the mode of heat transfer by conduction. The results relating to the Xue model for ϕ=0.06 indicates that the mean Nusselt number ratio is about 2.14, 1.95, 1.40 and 1.45 for Ra =10^3, 10^4, 10^5 and 10^6, respectively. However, the results obtained with Maxwell's model indicate that the mean Nusselt number ratio is about 1.20, 1.15, 1.01 and 1.02 for Ra=10^3, 10^4, 10^5 and 10^6, respectively.

CONCLUSION

In this study, MHD natural convection heat transfer of carbon nanotube-water nanofluid in square enclosure under a magnetic field and including two isothermal thin fins is numerically investigated. The effects of Hartmann number, Rayleigh number, positon of the fins, length of the fins, spacing of the two fins and volume fraction of nanoparticles are analysed. The results obtained revealed the following points:

- Application of the magnetic field results in significant changes in the flow structure, especially if the Hartmann number is high and the Rayleigh numbers are low.
- The flow strength increases with the increase in the Rayleigh number and decreases when the Hartmann number increases.
- The application of the magnetic field also modifies the temperature distribution in the cavity, in particular at Ra =10^5, however the effect of the magnetic force becomes less significant At Ra=10^6.
- When the Rayleigh number is between 10^4 and 10^5, the heat transfer rate increases or decreases depending on the value of the Hartmann number. But, beyond Ra=10^5, there is an improvement in heat transfer for any value of the Hartmann number.
- At low Rayleigh numbers (Ra=10^3 and 10^4), the average Nusselt number is not sensitive either to the variation in position or to the spacing of the two fins.
- The increase in fins length leads to increase in mean Nusselt number for any Rayleigh number, Whereas the optimal heat transfer rate is obtained at the position S=0.3 for all Rayleigh numbers.
- The heat transfer rate improves with increasing concentration of nanoparticles, the effect of which is important at low Rayleigh number.
- The heat transfer rates obtained using the Maxwell model are lower than those obtained using the Xue model.

As future extension to this work, the heterogeneous dynamic model due to the heterogeneity of the nanofluid induced by thermophoresis diffusions and brownian motion, may be investigated.

NOMENCLATURE

B_0	Strength of the magnetic field, T
Cp	Specific heat, $J.kg^{-1}.K^{-1}$
d	Dimensional spacing of two fins, m
D	Non dimensional spacing of two fins
g	Acceleration of gravity, $m.s^{-2}$
H	Cavity height, m
Ha	Hartmann number
k	Thermal conductivity, $W.m^{-1}.K^{-1}$
l	Dimensional thin fins length, m
L	Non dimensional thin fins length
Nu	Nussel number
p	Pressure, Pa
P	Non dimensional pressure
Pr	Prandtl number
Ra	Rayleigh number
s	Dimensional thin fins position, m
S	Non dimensional thin fins position
T	Temperature, K
u, v	velocity components, $m.s^{-1}$
U, V	Non dimensional velocities
x, y	Cartesian coordinates, m
X, Y	Non dimensional cartesian coordinates

Greeksymbols

α	Thermal diffusivity, $m^2.s^{-1}$
β	Thermal expansion coefficient, K^{-1}
θ	Non dimensional temperature
μ	Dynamic viscosity, Pa.s
υ	Kinematic viscosity, $m^2.s^{-1}$
ρ	Density, $kg.m^{-3}$
σ	Electrical conductivity, $S.m^{-1}$
ϕ	Nanoparticle volume fraction
ψ	Non dimensional stream function

Subscripts

C	Cold
f	Fluid (pure water)
H	Hot
m	Mean
max	Maximum
nf	Nanofluid
o	Reference value
p	Nanoparticle

Abbreviations

CNT	Carbon nanotube
MHD	Magnetohydrodynamic
SWCNT	Single walled carbon nanotube

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] De Vahl Davis G. Natural convection of air in a square cavity: A bench mark numerical solution. Int J Numer Methods Fluids 1983;3:249–264. [CrossRef]

[2] Yildiz S. Investigation of natural convection heat transfer at constant heat flux along with a vertical and inclined plate. J Therm Eng 2018;4:2432–2444. [CrossRef]

[3] Shi X, Khodadadi JM. Laminar natural convection heat transfer in a differentially heated square cavity due to a thin fin on the hot wall. J Heat Transf 2003;125:624–634. [CrossRef]

[4] Tasnim SH, Collins MR. Numerical analysis of heat transfer in a square cavity with a baffle on the hot wall. Int Commun Heat Mass Transf 2004;31:639–650. [CrossRef]

[5] Bilgen E. Natural convection in cavities with a thin fin on the hot wall. Int J Heat Mass Transf 2005;48:3493–3505. [CrossRef]

[6] Ben-Nakhi A, Chamkha AJ. Effect of length and inclination of a thin fin on natural convection in a square enclosure. Numer Heat Transf A Appl 2006;50:381–399. [CrossRef]

[7] Nardini G, Paroncini M, Vitali R. Natural convection in a square cavity with two baffles on the vertical walls: Experimental and numerical investigation. Int J Mech 2015;9:120–127.

[8] Attouchi MT, Larbi S, Khelladi S. Effect of some parameters on natural convection heat transfer in finned enclosures - A case study. Int J Thermofluid Sci Technol 2022;9:090102. [CrossRef]

[9] Santra AK, Sen S, Chakraborty N. Study of heat transfer augmentation in a differentially heated square cavity using copper water nanofluid. Int J Therm Sci 2007;47:1113–1122. [CrossRef]

[10] Ho CJ, Chen MW, Li ZW. Numerical simulation of natural convection of nanofluid in a square enclosure: Effects due to uncertainties of viscosity and thermal conductivity. Int J Heat Mass Transf 2008;51:4506–4516. [CrossRef]

[11] Aminossadati SM, Ghasemi B. Natural convection cooling of a localized heat source at the bottom of a nanofluid-filled enclosure. Eur J Mech B Fluids 2009;28:630–640. [CrossRef]

[12] Suneetha S, Subbarayudu K, Bala Anki Reddy P. Hybrid nanofluids development and benefits: A comprehensive review. J Therm Eng 2022;8:445–455. [CrossRef]

[13] El Hattab M, Lafdaili Z. Turbulent natural convection heat transfer in a square cavity with nanofluids in presence of inclined magnetic field. Therm Sci 2022;26:3201–3213. [CrossRef]

[14] Ul Haq R, Nadeem S, Khan ZH, Noor NFM. Convective heat transfer in MHD slip flow over a stretching surface in the presence of carbon nanotubes. Phys B Condens Matter 2015;457:40–47. [CrossRef]

[15] Tayebi T, Ferhat CE, Rezig N, Djezzar M. Free convection in a carbon nanotube-water nanofluid filled enclosure with power-law variation wall temperature. J Nanofluids 2016;5:531–542. [CrossRef]

[16] Tayebi T, Chamkha AJ, Djezzar M. Natural convection of CNT-water nanofluid in an annular space between confocal elliptic cylinders with constant heat flux on inner wall. Sci Iran 2019;26:2770–2783.

[17] Noranuar WNN, Mohamad AQ, Shafie S, Khan I. Unsteady free convection flow of water-based carbon nanotubes due to non-coaxial rotations of moving disk. J Appl Sci Eng 2022;25:501–510.
[18] Borode AO, Ahmed NA, Olubambi PA. A review of heat transfer application of carbon-based nanofluid in heat exchangers. Nano Struct Nano Objects 2019;20:100394. [CrossRef]
[19] Ali N, Bahman AM, Aljuwayhel NF, Ebrahim SA, Mukherjee S, Alsayegh A. Carbon-based nanofluids and their advances towards heat transfer applications-a review. Nanomaterials (Basel) 2021;11:1628. [CrossRef]
[20] Ghasemi B, Aminossadati SM, Raisi A. Magnetic field effect on natural convection in a nanofluid-filled square enclosure. Int J Therm Sci 2011;50:1748–1756. [CrossRef]
[21] Sourtiji E, Hosseinizadeh SF. Heat transfer augmentation of magnetohydrodynamics natural-convection in L-shaped cavities utilizing nanofluids. Therm Sci 2012;16:489–501. [CrossRef]
[22] Mejri I, Mahmoudi A, Abbassi MA, Omri A. Magnetic field effect on entropy generation in a nanofluid-filled enclosure with sinusoidal heating on both side walls. Powder Technol 2014;266:340–353. [CrossRef]
[23] Belhaj S, Ben-Beya B. Numerical simulation of unsteady MHD natural convection of CNT-water nanofluid in square cavity heated sinusoidally from below. Particul Sci Technol 2019;37:851–870. [CrossRef]
[24] Hamid M, Khan ZH, Khan WA, Ul Haq RU. Natural convection of water-based carbon nanotubes in a partially heated rectangular fin-shaped cavity with an inner cylindrical obstacle. Phys Fluids 2019;31:103607. [CrossRef]
[25] Sarala S, Geetha E, Nirmala M. Numerical investigation of heat transfer & hall effects on mhd nanofluid flow past over an oscillating plate with radiation. J Therm Eng 2022;8:757–771. [CrossRef]
[26] Gray DD, Giorgini A. The validity of the boussinesq approximation for liquids and gases. Int J Heat Mass Transf 1976;19:545–551. [CrossRef]
[27] Job VM, Gunakala SR, Rushi Kumar B, Sivaraj R. Time-dependent hydromagnetic free convection nanofluid flows within a wavy trapezoidal enclosure. Appl Therm Eng 2017;115:363–377. [CrossRef]
[28] Xuan Y, Roetzel W. Conceptions for heat transfer correlation of nanofluids. Int J Heat Mass Transf 2000;43:3701–3707. [CrossRef]
[29] Brinkman HC. The viscosity of concentrated suspensions and solution. J Chem Phys 1952;20:571–581. [CrossRef]
[30] Maxwell JC. A Treatise on Electricity and Magnetism. Oxford: Clarendon Press; 1891.
[31] Patankar SV. Numerical Heat Transfer and Fluid-Flow. New York, USA: Hemisphere Publishing; 1980.
[32] Teamah MA, El-Maghlany WM. Augmentation of natural convective heat transfer in square cavity by utilizing nanofluids in the presence of magnetic field and uniform heat generation/absorption. Int J Therm Sci 2012;58:130–142. [CrossRef]
[33] Xue QZ. Model for thermal conductivity of carbon nanotube-based composites. Physica B Condens Matter 2005;368:302–307. [CrossRef]
[34] Lee S, Choi SUS, Li S, Eastman JA. Measuring thermal conductivity of fluids containing oxide nanoparticles. J Heat Transf 1999;121:280–289. [CrossRef]
[35] Abu-Nada E. Effects of variable viscosity and thermal conductivity of Al2O3-water nanofluid on heat transfer enhancement in natural convection. Int J Heat Fluid Flow 2009;30:679–690. [CrossRef]

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1088–1100, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00090

Research Article

Performance estimation of parallel Syste under online and offline preventive maintenance

Hamisu Ismail AYAGI[1,4,*], Zhong WAN[1], Ibrahim YUSUF[2,*], Abdullahi SANUSI[3]

[1]*School of Mathematics and Statistics, Central South University, Changsha, 410083, China*
[2]*Department of Mathematical Sciences, Bayero University, Kano, 700006, Nigeria*
[3]*Department of Science, School of Continuing Education, Bayero University, Kano, 700006, Nigeria*
[4]*Department of Mathematics, Yusuf Maitama Sule University, Kano, 700271, Nigeria*

ARTICLE INFO

Article history
Received: 24 October 2022
Revised: 04 January 2023
Accepted: 14 February 2023

Keywords:
Cost Analysis; Mean Time to Failure; Reliability

ABSTRACT

In this paper, the reliability characteristics of a parallel system are investigated. The parallel system under consideration is made up of three active units that run in parallel, with two of them having to be operational in order for the system to work. The main purpose of this study is to quantify/examine the effect of online and offline preventive maintenance. Preventive maintenance is carried out on the systems in two ways: online and offline preventive maintenance. After the first unit of each system fails, online preventive maintenance is performed. Following the failure of the second unit of each system, offline preventive maintenance is performed. Partial and complete failures are the two types of failures that may occur. Both systems can undergo exponential failure and repair. Using supplementary variable technique, Laplace transform, and Copula repair approach, the system of first-order differential equations associated with system effectiveness, which are crucial to this research, is established and resolved. Tables and graphs are used to illustrate the important findings based on assumed numerical values. System designers, programmers, and maintenance supervisors will be able to create and maintain more crucial systems with the assistance of this research paper.

Cite this article as: Ayagi HI, Wan Z, Yusuf I, Sanusi A. Performance estimation of parallel Syste under online and offline preventive maintenance. Sigma J Eng Nat Sci 2024;42(4):1088–1100.

INTRODUCTION

One of the most important characteristics used to assess system performance in engineering is reliability and is also regarded as the best starting point for any system improvement. Due to lack of adequate probabilistic knowledge, convectional reliability's binary state i.e., performance or failure state assumptions, are inadequate for analysing the reliability of complex industrial systems. The dubiety of each individual parameter in large industrial systems adds to the overall system reliability dubiety.

As a result, a number of researchers have created various industrial systems to investigate reliability, availability, mean

***Corresponding author.**
*E-mail address: hamisu@csu.edu.cn, wanmath@csu.edu.cn, iyusuf.mth@buk.edu.ng, asanusi.sce@buk.edu.ng

This paper was recommended for publication in revised form by Editor in Chief Ahmet Selim Dalkılıç

Published by Yıldız Technical University Press, İstanbul, Turkey

time to system failure, and cost function, as well as the impact of coverage variables, failure and repair rates. Researchers such as; [1] explored a method for evaluating the reliability and profitability of a warranty-dependent industrial system. [2] focused on evaluating the performance of serial systems with a variety of failure types and repair policies. Using Kolmogrov differential equations, [3] proposed a method for analyzing the reliability of industrial systems. [4] investigated a novel method for reliability analysis of series-parallel systems via credibility theory. [5] presented their research on the availability of standby complex systems in the event of human failure or repair. [6] presented Markov chain profit modeling and evaluation between two dissimilar systems under two forms of failures. [7] have published a paper on the performance assessment of a series-parallel industrial method. [8] presented the findings of an analysis of the multi-objective non-linear programming problem for reliability (GSA) and compared them to results computed using the practice swarm optimization (PSO) methodology. The reliability study of a two-dissimilar unit warm standby repairable system with priority in usage was addressed by [9]. [10] addressed the availability and cost analysis of an engineering system with series-connected subsystems. [11] discussed the evaluation of some reliability characteristics of a single unit system requiring two types of supporting device for operations. [12] used the Gumbel-Hougaard family copula to study the cost assessment of a complex repairable system made up of two subsystems in a series arrangement. [13] used copula distribution to offer a stochastic study of an N-unit plant with various failure kinds. Improving system performance can also be attained either by increasing the reliability of each component/subsystem in the system or by providing proper maintenance (preventive maintenance).

Preventive maintenance is a type of maintenance that is carried out on a piece of equipment on a regular basis to minimize the risk of failing. It is done when the equipment is still running to keep it from breaking down unexpectedly. Cleaning, lubrication, oil changes, modifications, repairs, checking and removing components, as well as partial and full overhauls, are all part of preventive maintenance schedule in operation. Preventive maintenance is used in many industries to help maintenance managers and reliability engineers increase system reliability, availability, and revenue while lowering costs. The primary goal of every manufacturing company or factory is to increase profits by lowering production costs. In today's competitive age, we can't afford to lose production or have facilities fail due to increased demand. As a result, appropriate maintenance procedures must be used to achieve the best results.

Online and offline preventive maintenance techniques are two of the most common and universal maintenance techniques among the many. The aim of online preventive maintenance is to inspect the system or any output on a regular basis. It is a non-interruptible operation, which means the system won't be shut down until it's done. This assists in cost-cutting, reducing equipment downtime, improving asset lifespan and efficiency, and avoids unexpected expenses. Offline preventive maintenance is a type of inspection that is done manually. Regular asset scanning is sufficient for offline preventive maintenance.

Many research papers on system reliability and availability in the presence of preventive maintenance have been published, including the following; [14] investigated cost-effective coal-fired power plant preventive maintenance scheduling. [15] conducted a probabilistic study of two single-unit device reliability models with preventive maintenance beyond the warranty period and deterioration. [16] examined the best preventive maintenance warranty strategy for repairable goods with growing failure rates on a regular basis. [17] proposed a stochastic study of a reheating-furnace system undergoing preventive maintenance and repair. [18] presented a report on a single unit system with preventive maintenance and repair that was subjected to the fastest possible operation and repair times. For single machine framework, [19] suggested an integrated model of production planning and incomplete preventive maintenance strategy. [20] proposed a parallel system reliability assessment with two types of preventive maintenance. According to maximum operation and repair timeframes, [21] developed a single unit system with preventive maintenance and repair. Stochastic modeling of non-identical redundant systems with priority, preventive maintenance, and Weibull failure and repair distributions was presented by [22].

The literature review above reveals that the majority of previous studies have focused on preventive maintenance and other types of repair policies. Preventive maintenance policies, both online and offline under copula distribution approach, have received little or no attention in the literature. As a result, the impact of online and offline preventive maintenance has been highlighted in this paper.

The reminder of the paper is organised such that section 2 contains notations, assumptions, and description of the model under consideration. The formulation and solution of the model were captured in section 3. The model analysis was discussed in section 4. The findings are discussed in section 5, and finally the conclusion part of the paper is given in section 6.

NOTATIONS, ASSUMPTIONS AND DESCRIPTION OF MODEL

Notations

Here, we provide a table with the notations and definitions used in this study.

Assumptions

Below are listed all the presumptions that were made for this study.

i. √Firstly, all subsystems are assumed to be operational.
ii. Two units from each subsystem are needed for operation.
iii. When any component fails, the system's output suffers.
iv. If a subsystem unit fails, it can be repaired when it is still operational, or it can fail completely.

Table 1. Notations and Definitions

Notations	Definitions
T	Time variable on a time scale
S	Laplace transform variable for all expressions
λ_1	System A's failure rate
λ_2	System B's failure rate
$\emptyset(x)$/ $\emptyset(y)$	Repair rate of system A / system B
$\mu_0(x)/\mu_0(y)$	Repair rate for a fully failed states of system A and system B respectively
$p_i(t)$ For i =0 to 10,	The probability that the system is in S_i state at any given time
$\overline{P}(s)$	Laplace transformation of state transition probability $p(t)$
P_i (x, t)	The probability that a system is in state S_i such that for i=1..., the system is under repair, and the elapsed repair time is (x, t) with x denoting repair and t denoting time.
P_i (y, t)	The probability that a system is in state S_i such that for i=1..., the system is under repair, and the elapsed repair time is (y, t) with y being the repair variable and t being the time variable.
P_i (m, t)	The probability that a system is in state S_i for i=1…, the system is under repair, and the elapsed repair time is (m, t) with repair variable being m and time variable being t.
P_i (n, t)	The probability that a system is in state S_i such that for i=1…, the system is under repair, and the elapsed repair time is (n, t) with n representing the repair variable and t representing the time variable.
$E_p(t)$	Expected profit over the course of the time interval [0, t)
K_1, K_2	Revenue and service cost per unit time, respectively
$\mu_0(x)$	According to the Gumbel-Hougaard family copula definition, joint probability is expressed as: $c_\theta(u_1(x), u_2(x)) = exp\left(x^\theta + \{log\ \phi\ (x)^\theta\}^{\frac{1}{\theta}}\right)$, $1 \leq \theta \leq \infty$. Where $\mu_1 = \emptyset(x)$ and $u_2 = e^x$

v. All failure rates are expected to follow exponential distributions and are constant.
vi. Partially failed states are restored using general distribution, while entirely failed state distribution is handled by the Gumbel Hougaard family copula.
vii. The repaired machine unit should function as new and the repair process should have caused no damage.
viii. The load will be ready for the system's successful performance as soon as the failed unit is repaired.

The Model Overviews

In this research, we looked at a mathematical model of parallel system with two subsystems A and B, each consists of three active units in parallel that can fail in two ways: partially or fully. These two systems are separated

Table 2. Description of States

States	Description
S_0	Is the ideal state, in which both systems are fully functional. The system works since one unit is on standby in both systems.
S_1	In this state, one unit in system A has failed, two units in both systems A and B are operational, and one unit in system B is on standby. The system is up and running.
S_3	In this state, one unit in system B has failed, two units in both systems A and B are operational, and one unit in system A is on standby. The system works.
S_4	Here, one unit in system B has previously failed, one unit in system A suddenly failed, and two units in system A are operational. The system is functional.
S_5	Previously, a unit has failed in system A, suddenly a unit failed in system B, then the system is operational.
S_6	Previously, a unit has failed in system B, suddenly a unit failed in system A, then the system is functional.
S_7	After the failure of the first unit in both systems A and B, this state denotes the online preventive maintenance state. In this state, the system is operational.
S_8	This state denotes system's A total failure as a result of two of its unit failing. The system is not functioning.
S_9	This state represents system's B total failure as a result of two of its unit failing. The system is not working.
S_{10}	This state denotes the offline preventive maintenance state where the system is neither down nor working.

from one another and are located in different locations. When two units from both systems A and B fail, the systems continue to function, but any more failure will bring the systems to total failure. In the case of total failure, both systems A and B can be fixed simultaneously using Gumbel-Hougaard family copula, whereas general repair distribution can be used to bring back partially failed states. The systems go through online preventive maintenance immediately after the failure of first unit with a rate of δ_0 and immediately fixed with a rate of $\emptyset(m)$, as well as offline preventive maintenance after the failure of the second unit with a rate of δ_1 and immediately returned with a rate of $\emptyset(n)$. The system has eleven states including perfect state, seven of which are operational and three of which are not. The detailed overview of the states is given below in table 2:

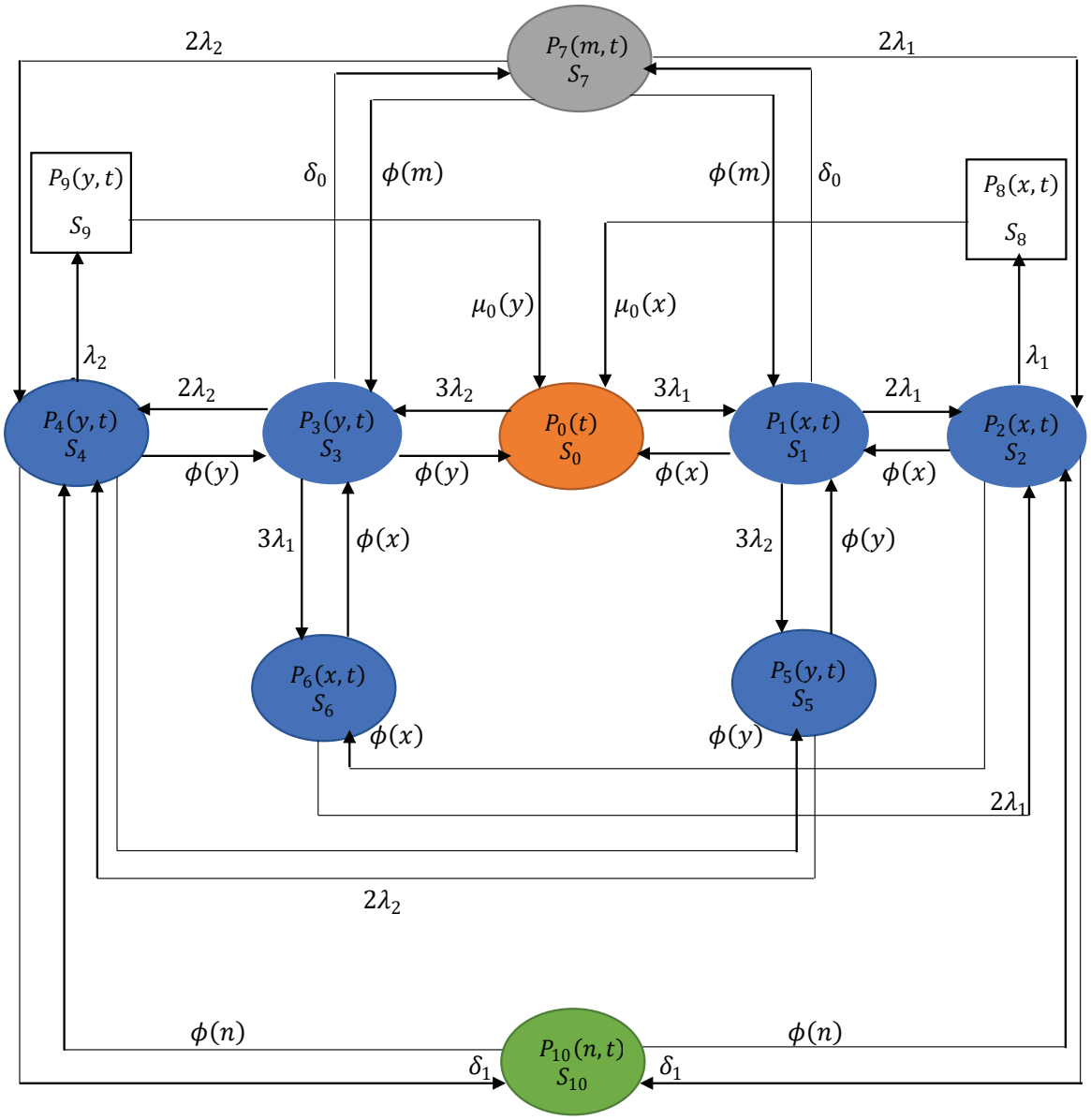
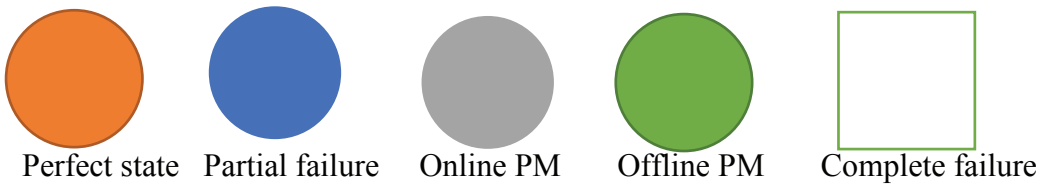


Figure 1. Transition diagram of the model.



MODELS' FORMULATION AND SOLUTION

Model Formulation

As in Lado and Singh (2019), and Singh and Rawal (2015), we obtain the following differential difference equations as:

$$\left\{\frac{\partial}{\partial t}+3\lambda_1+3\lambda_2\right\}P_0(t)=\int_0^\infty \emptyset(x)P_1(x,t)dx+\int_0^\infty \emptyset(y)P_3(y,t)dy + \int_0^\infty \mu_0(x)P_8(x,t)dx+\int_0^\infty \mu_0(y)P_9(y,t)dy, \quad (1)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+2\lambda_1+3\lambda_2+\delta_0+\phi(x)\right\}P_1(x,t)=0, \quad (2)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+\lambda_1+\delta_1+2\phi(x)\right\}P_2(x,t)=0, \quad (3)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial y}+3\lambda_1+2\lambda_2+\delta_0+\phi(y)\right\}P_3(y,t)=0, \quad (4)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial y}+\lambda_2+\delta_1+2\phi(y)\right\}P_4(y,t)=0, \quad (5)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial y}+2\lambda_2+\phi(y)\right\}P_5(y,t)=0, \quad (6)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+2\lambda_1+\phi(x)\right\}P_6(x,t)=0, \quad (7)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial m}+2\lambda_1+2\lambda_2+2\phi(m)\right\}P_7(m,t)=0, \quad (8)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+\mu_0(x)\right\}P_8(x,t)=0, \quad (9)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial y}+\mu_0(y)\right\}P_9(y,t)=0, \quad (10)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial n}+2\phi(n)\right\}P_{10}(n,t)=0. \quad (11)$$

The study's boundary conditions are generated from figure 2 by analyzing the rates of exit from each state in terms of the probability of being in the initial state, $P_0(t)$:

$$P_1(0,t)=3\lambda_1P_0(t), \quad (12)$$

$$P_2(0,t)=\{6\lambda_1^2+6\lambda_1^2\delta_0+6\lambda_1\lambda_2\delta_0+18\lambda_1^2\lambda_2\}P_0(t), \quad (13)$$

$$P_3(0,t)=3\lambda_2P_0(t), \quad (14)$$

$$P_4(0,t)=\{6\lambda_2^2+6\lambda_2^2\delta_0+6\lambda_1\lambda_2\delta_0+18\lambda_1\lambda_2^2\}P_0(t), \quad (15)$$

$$P_5(0,t)=9\lambda_1\lambda_2P_0(t), \quad (16)$$

$$P_6(0,t)=9\lambda_1\lambda_2P_0(t), \quad (17)$$

$$P_7(0,t)=\{3\lambda_1\delta_0+3\lambda_2\delta_0\}P_0(t), \quad (18)$$

$$P_8(0,t)=\{6\lambda_1^3+6\lambda_1^3\delta_0+6\lambda_1^2\lambda_2\delta_0+18\lambda_1^3\lambda_2\}P_0(t), \quad (19)$$

$$P_9(0,t)=\{6\lambda_2^3+6\lambda_2^3\delta_0+6\lambda_1\lambda_2^2\delta_0+18\lambda_1\lambda_2^3\}P_0(t), \quad (20)$$

$$P_{10}(0,t)=\{6\lambda_1^2\delta_1+6\lambda_1^2\delta_0\delta_1+12\lambda_1\lambda_2\delta_0\delta_1 + 18\lambda_1^2\lambda_2\delta_1+6\lambda_2^2\delta_1+6\lambda_2^2\delta_0\delta_1 + 18\lambda_1\lambda_2^2\delta_1\}P_0(t). \quad (21)$$

Initial conditions

$P_0(0) = 1$, i.e., in a perfect state, the probability of a state transition at time $t = 0$ is zero.

With the above initial condition, we can obtain Laplace transforms of equations (1) to (11) as:

$$\{s+3\lambda_1+3\lambda_2\}\overline{P_0}(s)=1+\int_0^\infty \emptyset(x)\overline{P_1}(x,s)dx+\int_0^\infty \emptyset(y)\overline{P_3}(y,s)dy + \int_0^\infty \mu_0(x)\overline{P_8}(x,s)dx+\int_0^\infty \mu_0(y)\overline{P_9}(y,s)dy, \quad (22)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+2\lambda_1+3\lambda_2+\delta_0+\phi(x)\right\}\overline{P_1}(x,s)=0, \quad (23)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+\lambda_1+\delta_1+2\phi(x)\right\}\overline{P_2}(x,s)=0, \quad (24)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial y}+3\lambda_1+2\lambda_2+\delta_0+\phi(y)\right\}\overline{P_3}(y,s)=0, \quad (25)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial y}+\lambda_2+\delta_1+2\phi(y)\right\}\overline{P_4}(y,s)=0, \quad (26)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial y}+2\lambda_2+\phi(y)\right\}\overline{P_5}(y,s)=0, \quad (27)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+2\lambda_1+\phi(x)\right\}\overline{P_6}(x,s)=0, \quad (28)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial m}+2\lambda_1+2\lambda_2+2\phi(m)\right\}\overline{P_7}(m,s)=0, \quad (29)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+\mu_0(x)\right\}\overline{P_8}(x,s)=0, \quad (30)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial y}+\mu_0(y)\right\}\overline{P_9}(y,s)=0, \quad (31)$$

$$\left\{\frac{\partial}{\partial t}+\frac{\partial}{\partial n}+2\phi(n)\right\}\overline{P_{10}}(n,s)=0. \quad (32)$$

Laplace transform of boundary conditions:

$$\overline{P_1}(0,s)=3\lambda_1\overline{P_0}(s), \quad (33)$$

$$\overline{P_2}(0,s)=\{6\lambda_1^2+6\lambda_1^2\delta_0+6\lambda_1\lambda_2\delta_0+18\lambda_1^2\lambda_2\}\overline{P_0}(s), \quad (34)$$

$$\overline{P_3}(0,s)=3\lambda_2\overline{P_0}(s), \quad (35)$$

$$\overline{P_4}(0,s)=\{6\lambda_2^2+6\lambda_2^2\delta_0+6\lambda_1\lambda_2\delta_0+18\lambda_1\lambda_2^2\}\overline{P_0}(s), \quad (36)$$

$$\overline{P_5}(0,s)=9\lambda_1\lambda_2\overline{P_0}(s), \quad (37)$$

$$\overline{P_6}(0,s)=9\lambda_1\lambda_2\overline{P_0}(s), \quad (38)$$

$$P_7(0,s)=\{3\lambda_1\delta_0+3\lambda_2\delta_0\}P_0(s), \quad (39)$$

$$P_8(0,s)=\{6\lambda_1^3+6\lambda_1^3\delta_0+6\lambda_1^2\lambda_2\delta_0+18\lambda_1^3\lambda_2\}P_0(s), \quad (40)$$

$$P_9(0,s)=\{6\lambda_2^3+6\lambda_2^3\delta_0+6\lambda_1\lambda_2^2\delta_0+18\lambda_1\lambda_2^3\}P_0(s), \quad (41)$$

$$\begin{aligned}P_{10}(0,s)=\{&6\lambda_1^2\delta_1+6\lambda_1^2\delta_0\delta_1+12\lambda_1\lambda_2\delta_0\delta_1\\&+18\lambda_1^2\lambda_2\delta_1+6\lambda_2^2\delta_1+6\lambda_2^2\delta_0\delta_1\\&+18\lambda_1\lambda_2^2\delta_1\}P_0(s).\end{aligned} \quad (42)$$

Solution of the models

Combining equation (22) and (32) with the aid of (33) to (42), we get:

$$\overline{P_1}(s)=3\lambda_1\left\{\frac{1-S_\emptyset(s+2\lambda_1+3\lambda_2+\delta_0)}{(s+2\lambda_1+3\lambda_2+\delta_0)}\right\}\overline{P_0}(s), \quad (43)$$

$$\overline{P_2}(s)=M_1\left\{\frac{1-\overline{S_{2\emptyset}}(s+\lambda_1+\delta_1)}{(s+\lambda_1++\delta_1)}\right\}\overline{P_0}(s), \quad (44)$$

$$\overline{P_3}(s)=3\lambda_2\left\{\frac{1-\overline{S_\emptyset}(s+3\lambda_1+2\lambda_2+\delta_0)}{(s+3\lambda_1+2\lambda_2+\delta_0)}\right\}\overline{P_0}(s), \quad (45)$$

$$\overline{P_4}(s)=M_2\left\{\frac{1-S_{2\emptyset}(s+\lambda_2+\delta_1)}{(s+\lambda_2+\delta_1)}\right\}\overline{P_0}(s), \quad (46)$$

$$\overline{P_5}(s)=9\lambda_1\lambda_2\left\{\frac{1-S_\emptyset(s+2\lambda_2)}{(s+2\lambda_2)}\right\}\overline{P_0}(s), \quad (47)$$

$$\overline{P_6}(s)=9\lambda_1\lambda_2\left\{\frac{1-S_\emptyset(s+2\lambda_1)}{(s+2\lambda_1)}\right\}\overline{P_0}(s), \quad (48)$$

$$\overline{P_7}(s)=\{3\lambda_1\delta_0+3\lambda_2\delta_0\}\left\{\frac{1-\overline{S_{2\emptyset}}(s+2\lambda_1+2\lambda_2)}{(s+2\lambda_1+2\lambda_2)}\right\}\overline{P_0}(s), \quad (49)$$

$$\overline{P_8}(s)=M_3\left\{\frac{1-\overline{S_{\mu_0}}(s)}{s}\right\}\overline{P_0}(s), \quad (50)$$

$$\overline{P_9}(s)=M_4\left\{\frac{1-S_{\mu_0}(s)}{s}\right\}\overline{P_0}(s), \quad (51)$$

$$\overline{P_{10}}(s)=M_5\left\{\frac{1-\overline{S_{2\emptyset}}(s)}{s}\right\}\overline{P_0}(s), \quad (52)$$

From (21), we obtain:

$$\overline{P_0}(s)=\frac{1}{\Delta(s)}, \quad (53)$$

Where:

$$\begin{aligned}\Delta(s)=&(s+3\lambda_1+3\lambda_2)-3\lambda_1\overline{S_\emptyset}(s+2\lambda_1+3\lambda_2+\delta_0)\\&-3\lambda_2\overline{S_\emptyset}(s+3\lambda_1+2\lambda_2+\delta_0)\text{-}(M_3+M_4)\overline{S_{\mu_0}}(s),\\M_1=&\{6\lambda_1^2+6\lambda_1^2\delta_0+6\lambda_1\lambda_2\delta_0+18\lambda_1^2\lambda_2\},\\M_2=&\{6\lambda_2^2+6\lambda_2^2\delta_0+6\lambda_1\lambda_2\delta_0+18\lambda_1\lambda_2^2\},\\M_3=&\{6\lambda_1^3+6\lambda_1^3\delta_0+6\lambda_1^2\lambda_2\delta_0+18\lambda_1^3\lambda_2\},\\M_3=&\{6\lambda_2^3+6\lambda_2^3\delta_0+6\lambda_1\lambda_2^2\delta_0+18\lambda_1\lambda_2^3\},\\M_5=&\{6\lambda_1^2\delta_1+6\lambda_1^2\delta_0\delta_1+12\lambda_1\lambda_2\delta_0\delta_1+18\lambda_1^2\lambda_2\delta_1\\&+6\lambda_2^2\delta_1+6\lambda_2^2\delta_0\delta_1+18\lambda_1\lambda_2^2\delta_1\},\end{aligned}$$

The probability that the system is operational is:

$$\begin{aligned}P_{up}(s)=&P_0(s)+P_1(s)+P_2(s)+P_3(s)+P_4(s)\\&+P_5(s)+P_6(s)+P_7(s),\\\overline{P}_{up}(s)=&\frac{1}{\Delta(s)}\Big\{1+3\lambda_1\left(\frac{1-\overline{S_\emptyset}(s+2\lambda_1+3\lambda_2+\delta_0)}{(s+2\lambda_1+3\lambda_2+\delta_0)}\right)\\&+M_1\left(\frac{1-\overline{S_{2\emptyset}}(s+\lambda_1+\delta_1)}{(s+\lambda_1++\delta_1)}\right)+3\lambda_2\left(\frac{1-\overline{S_\emptyset}(s+3\lambda_1+2\lambda_2+\delta_0)}{(s+3\lambda_1+2\lambda_2+\delta_0)}\right)\\&+M_2\left(\frac{1-\overline{S_{2\emptyset}}(s+\lambda_2+\delta_1)}{(s+\lambda_2+\delta_1)}\right)+9\lambda_1\lambda_2\left(\frac{1-\overline{S_\emptyset}(s+2\lambda_2)}{(s+2\lambda_2)}\right)+9\lambda_1\lambda_2\left(\frac{1-\overline{S_\emptyset}(s+2\lambda_1)}{(s+2\lambda_1)}\right)\\&+(3\lambda_1\delta_0+3\lambda_2\delta_0)\left(\frac{1-\overline{S_{2\emptyset}}(s+2\lambda_1+2\lambda_2)}{(s+2\lambda_1+2\lambda_2)}\right)\Big\}.\end{aligned} \quad (54)$$

MODEL ANALYSIS

Availability Analysis

To make a distinction, we looked at the system's availability in two different ways, viz:

$$P_{up}(s) = \{0.00005276490045e^{-2.718445492t} - 0.002813764515e^{-2.060000000t} - 0.007805509296e^{-1.185732893t} - 0.00001540029829e^{-1.106406617t} + 1.012602413e^{-0.007714997787t} - 0.0007566858053e^{-1.040000000t} - 0.0003698765456e^{-2.050000000t} - 0.0008939418739e^{-2.060000000t}\}. \quad (55)$$

Table 3 and figure 2 illustrate the system's availability for copula repair distribution when $t = 0,10,20, \ldots \ldots \ldots ,100$.

Table 3. Availability against time t with respect to copula distribution

Time	0	10	20	30	40	50	60	70	80	90	100
Availability	1.0000	0.9374	0.8678	0.8034	0.7437	0.6885	0.6374	0.5901	0.5463	0.5057	0.4681

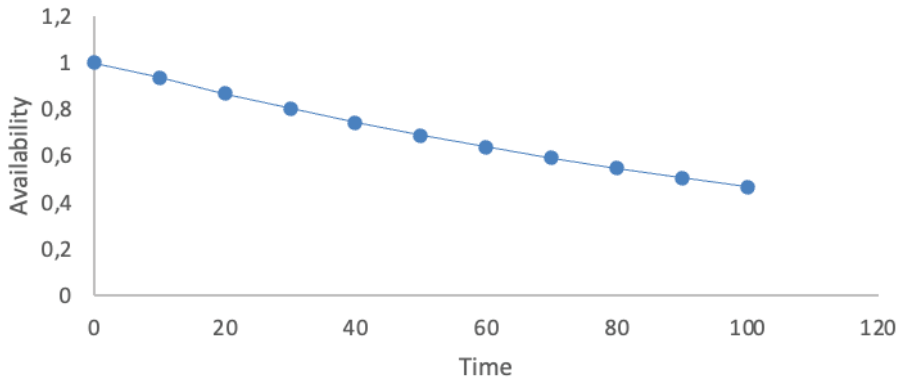


Figure 2. Availability against time with respect to copula distribution.

When repair follows Gumbel-Hougaard Family copula distribution

When the repair follows the copula distribution, the system availability can be determined from (54) for $\overline{S_{\emptyset}}(s) = \frac{\emptyset}{s+\emptyset}$, $\frac{1-\overline{S_{\emptyset}}(s)}{s}$, $\bar{S}_{\mu_0}(s) = \bar{S}_{exp[x^{\theta}+\{log\emptyset(x)\}^{\theta}]^{1/\theta}}(S) = \frac{exp[x^{\theta}+\{log\emptyset(x)\}^{\theta}]^{1/\theta}}{s+exp[x^{\theta}+\{log\emptyset(x)\}^{\theta}]^{1/\theta}}$, with all repairs set to 1, i.e., $\emptyset(x) = \emptyset(y) = \mu_0(x) = \mu_0(y) = \mu_0(m) = \mu_0(n) = 1$, and $\lambda_1 = 0.01$, $\lambda_2 = 0.02$, $\delta_0 = 0.03$, and $\delta_1 = 0.04$ fixed. Then if we take Laplace transforms, we get:

When repair follows general distribution

To obtain the availability of the system when the repair is through general distribution, we set $\overline{S_{\emptyset}}(s) = \frac{1}{s+\emptyset}$ and all repairs to 1, i.e., $\emptyset(x) = \emptyset(y) = \mu_0(x) = \mu_0(y) = \mu_0(m) = \mu_0(n) = 0$, and $\lambda_1 = 0.01$, $\lambda_2 = 0.02$, $\delta_0 = 0.03$, and $\delta_1 = 0.04$, then taking the Laplace transform of equation (54), we get the expression for availability as:

$$P_{up}(s) = \{0.00008717398954e^{-1.000077209t} + 1.012518714e^{-0.007714366521t} - 0.002814812304e^{-2.060000000t} - 0.0007578535962e^{-1.040000000t} - 0.0003700136500e^{-2.050000000t} - 0.0008971075757e^{-1.020000000t} - 0.007751104760e^{-1.185801404t} - 0.00001499582430e^{-1.106407020t}\}. \quad (56)$$

Table 4 and figure 3 present the system's availability for general repair distribution when $t = 0,10,20, \ldots \ldots \ldots ,100$.

Table 4. Availability against time t in terms of general repair

Time	0	10	20	30	40	50	60	70	80	90	100
Availability	1.0000	0.9374	0.8678	0.8033	0.7437	0.6885	0.6374	0.5900	0.5462	0.5057	0.4681

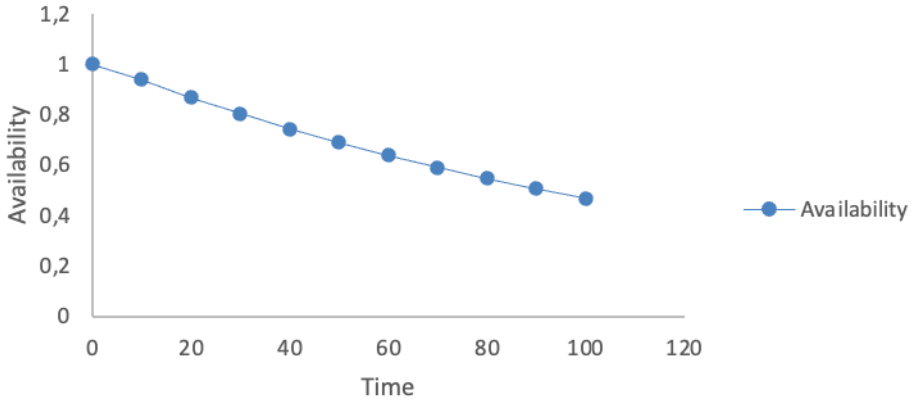


Figure 3. Availability against time in terms of general repair.

RELIABILITY ANALYSIS

By reducing all repair rates to zero i.e., $\emptyset(x) = \emptyset(y) = \mu_0(x) = \mu_0(y) = \mu_0(m) = \mu_0(n) = 0$, and taking $\lambda_1 = 0.01$, $\lambda_2 = 0.02$, $\delta_0 = 0.03$, and $\delta_1 = 0.04$ in equation (54), then using the inverse Laplace transform, the system's reliability can be expressed as:

$$R(t) = \{0.176000000e^{-0.060000000000t} + 0.03600000000e^{-0.040000000000t} - 6e^{-0.100000000000t} + 0.01725000000e^{-0.050000000000t} + 8.245035714e^{-0.090000000000t} - 1.500000000e^{-0.110000000t} + 0.02571428571e^{-0.020000000000t}\}. \quad (57)$$

Reliability can be calculated using equations (57) for different values of time t = 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100, as shown in Table 5 and Figure 4.

Mean Time to Failure (MTTF) Analysis

The expression for MTTF can be obtained by setting all repairs to zero and restricting as s reaches zero:

$$\lim_{s\to 0} \Delta(s) = 3\lambda_1 + 3\lambda_2,$$

$$MTTF = \lim_{s\to 0} \overline{P_{up}}(s) = \frac{1}{\Delta(s)}\left\{1 + \frac{3\lambda_1}{2\lambda_1+3\lambda_2+\delta_0} + \frac{M_1}{\lambda_1++\delta_1} + \frac{3\lambda_2}{3\lambda_1+2\lambda_2+\delta_0} + \frac{M_2}{\lambda_2+\delta_1} + \frac{9\lambda_1}{2} + \frac{9\lambda_2}{2} + \frac{3\lambda_1\delta_0+3\lambda_2\delta_0}{2\lambda_1+2\lambda_2}\right\}. \quad (58)$$

Setting $\lambda_1 = 0.01$, $\lambda_2 = 0.02$, $\delta_0 = 0.03$, $\delta_1 = 0.04$ and varying λ_1, λ_2, δ_0 and δ_1 one as 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09 respectively in equation (58), one may

Table 5. Variation of reliability corresponding to time t

Time	0	10	20	30	40	50	60	70	80	90	100
Reliability	1.0000	0.7978	0.4774	0.2579	0.1340	0.0695	0.0370	0.0206	0.0123	0.0079	0.0054

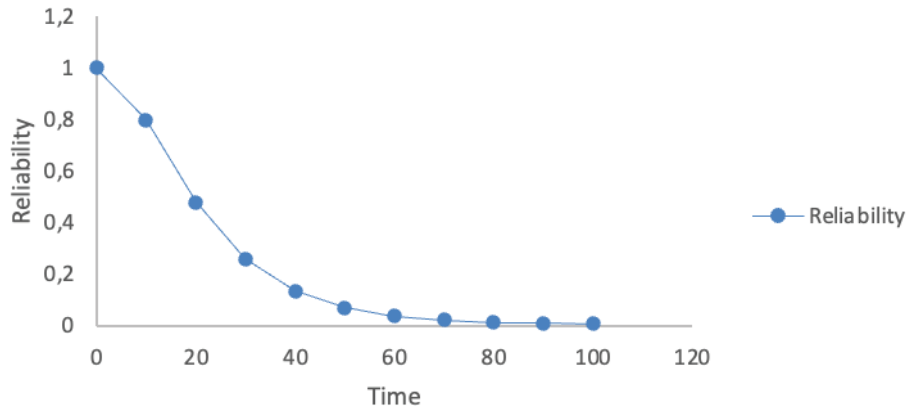


Figure 4. Variation of reliability corresponding to time.

Table 6. MTTF variation in terms of failure rates

Failure Rate	λ_1	λ_2	δ_0	δ_1
0.01	23.4391	31.8526	25.4245	24.1469
0.02	18.6473	23.4391	24.3056	23.7803
0.03	15.8481	19.0529	23.4391	23.5730
0.04	13.9870	16.3130	22.7579	23.4391
0.05	12.6479	14.4229	22.2165	23.3453
0.06	11.6322	13.0340	21.7833	23.2759
0.07	10.8323	11.9676	21.4356	23.2224
0.08	10.1844	11.1217	21.1566	23.1799
0.09	9.6480	10.4337	20.9336	23.1453

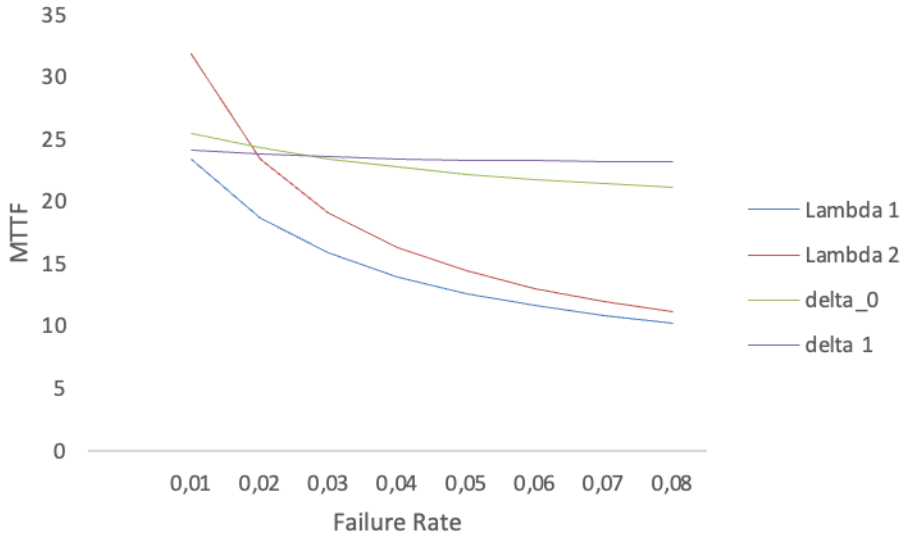


Figure 5. MTTF variation in terms of failure rates.

obtain table 6 and figure 5 which show the variation of MTTF in terms of failure rates.

Sensitivity Analysis

MTTF sensitivity can be determined using the partial derivative of MTTF with respect to failure rates in equation (58). Table 7 and Figure 6 display the MTTF sensitivity for fixed values of failure rates as $\lambda_1 = 0.01$, $\lambda_2 = 0.02$, $\delta_0 = 0.03$, $\delta_1 = 0.04$.

Cost Function Analysis

Case 1: Expression for cost function/Expected profit when the repair follows copula distribution

If the service facility is open at all times, the formula below will determine the estimated profit for the interval [0, t).

$$E_p(t) = K_1 \int_0^t P_{up}(t)dt - K_2 t - K_3 t \tag{59}$$

Where K_1, K_2 and K_3 in the interval [0, t) are the revenue generated, service cost per unit time, and, service cost due to online and offline maintenance, respectively. Equation (60) can be obtained for the same set of parameters in equation (55).

$$\begin{aligned} E_p(t) = K_1\{&0.0001804275832e^{-2.050000000t} - 0.00001940995345e^{-2.718445492t} \\ &+ 0.006582856343e^{-1.185732893t} + 0.00001391920299e^{-1.106406617t} \\ &- 131.2511605e^{-0.007714997787t} + 0.0008764136019e^{-1.020000000t} \\ &+ 0.001365905104e^{-2.060000000t} + 0.0007275825051e^{1.040000000t} \\ &+ 131.2414328\} - (K_2 - K_3)t. \end{aligned} \tag{60}$$

Fixing $K_1 = 1$, $K_2 = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6$, and, $K_3 = 0.1$. Table 8 and figure 7 can be obtained by using different values of the time variable, such as $t = 0, 10, 20, \ldots \ldots \ldots 100$.

Case 2: when the repair follows general distribution, the cost function is expressed as:

Table 7. Sensitivity of the MTTF to changes in failure rate

Failure rate	$\frac{\partial}{\partial\lambda_1}(MTTF)$	$\frac{\partial}{\partial\lambda_2}(MTTF)$	$\frac{\partial}{\partial\delta_0}(MTTF)$	$\frac{\partial}{\partial\delta_1}(MTTF)$
0.01	-654.5725	-1243.4183	-127.5855	-51.0185
0.02	-352.1928	-573.6695	-97.9045	-26.4351
0.03	-223.0796	-336.5049	-76.4815	-16.2583
0.04	-155.5404	-223.4422	-60.5112	-11.0296
0.05	-115.4175	-160.0163	-48.2868	-7.9799
0.06	-89.4430	-120.6018	-38.7212	-6.0437
0.07	-71.5630	-94.3224	-31.0950	-4.7370
0.08	-58.6773	-75.8745	-24.9171	-3.8131
0.09	-49.0550	-62.4030	-19.8423	-3.1358

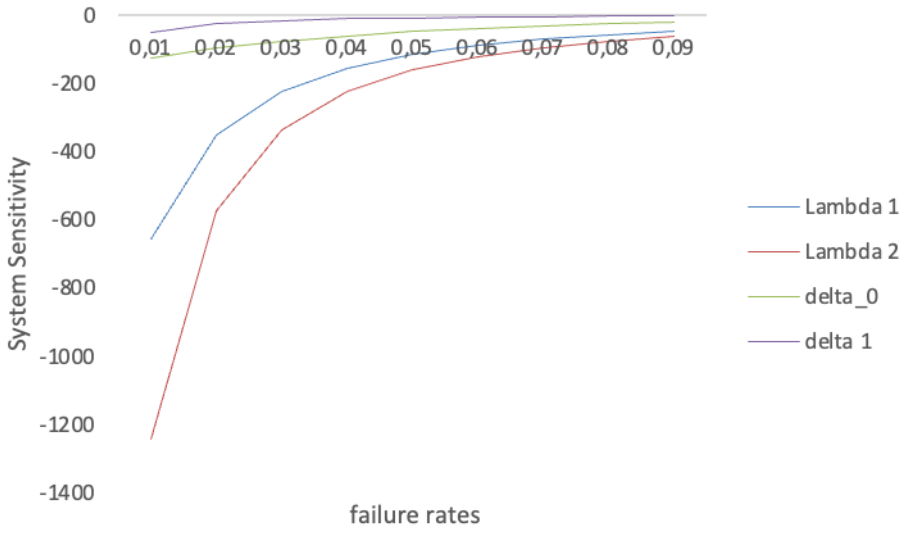


Figure 6. Sensitivity vs failure rate.

Table 8. Cost function/Expected profit versus time for copula repair

Time	E_p(t) $K_2 = 0.1$	E_p(t) $K_2 = 0.2$	E_p(t) $K_2 = 0.3$	E_p(t) $K_2 = 0.4$	E_p(t) $K_2 = 0.6$	E_p(t) $K_2 = 0.6$
0	0	0	0	0	0	0
10	7.7355	6.7355	5.7355	4.7355	3.7355	2.7355
20	14.7571	12.7571	10.7571	8.7571	6.7571	4.7571
30	21.109	18.109	15.109	12.109	9.109	6.109
40	26.8407	22.8407	18.8407	14.8407	10.8407	6.8407
50	32.9984	27.9984	22.9984	17.9984	12.9984	7.5678
60	37.6246	31.6246	25.6246	19.6246	13.6246	7.6246
70	42.7588	35.7588	28.7588	21.7588	14.7588	7.7588
80	48.4375	40.4375	32.4375	24.4375	16.4375	8.4375
90	53.6946	44.6946	35.6946	26.6946	17.6946	8.6946
100	59.5614	49.5614	39.5614	29.5614	19.5614	9.5614

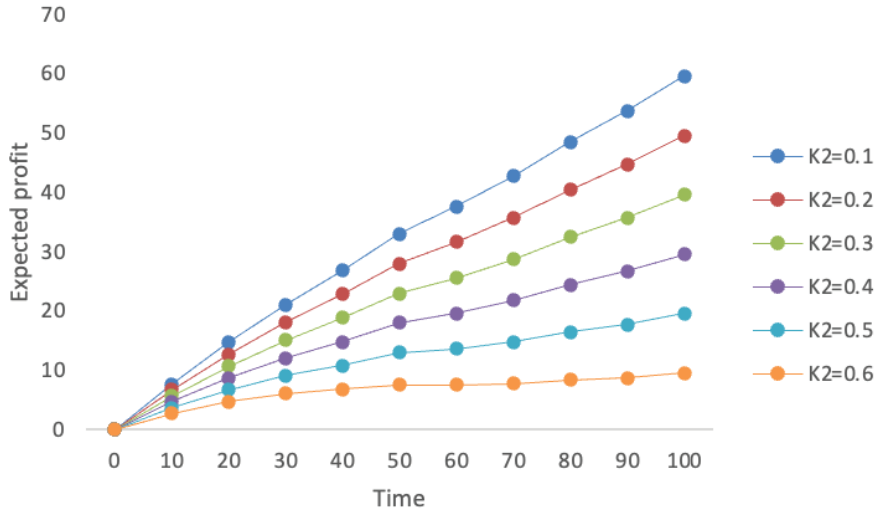


Figure 7. Cost function/Expected profit versus time for copula repair.

Table 9. Cost function/Expected profit versus time for general repair

Time	$E_p(t)$ $K_2 = 0.1$	$E_p(t)$ $K_2 = 0.2$	$E_p(t)$ $K_2 = 0.3$	$E_p(t)$ $K_2 = 0.4$	$E_p(t)$ $K_2 = 0.6$	$E_p(t)$ $K_2 = 0.6$
0	0	0	0	0	0	0
10	7.7348	6.7348	5.7348	4.7348	3.7348	2.7348
20	14.7558	12.7558	10.7558	8.7558	6.7558	4.7558
30	21.1071	18.1071	15.1071	12.1071	9.1071	6.1071
40	26.8384	22.8384	18.8384	14.8384	10.8384	6.8384
50	32.9956	27.9956	22.9956	17.9956	12.9956	7.5678
60	37.6215	31.6215	25.6215	19.6215	13.6215	7.6215
70	42.7555	35.7555	28.7555	21.7555	14.7555	7.7555
80	48.4340	40.4340	32.4340	24.4340	16.4340	8.4340
90	53.6910	44.6910	35.6910	26.6910	17.6910	8.6910
100	59.5576	49.5576	39.5576	29.5576	19.5576	9.5576

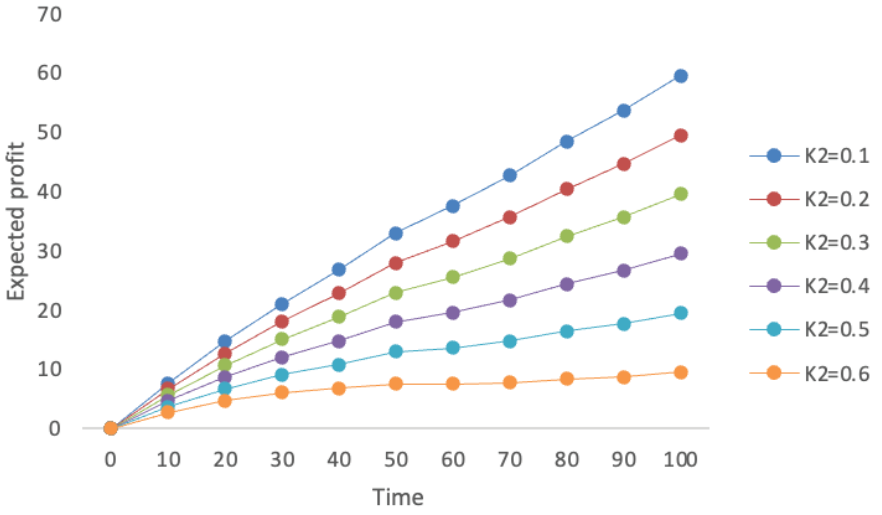


Figure 8. Cost function/Expected profit versus time for copula repair.

$$E_P(t) = K_1\{0.0001804944634e^{-2.050000000t} + 0.006536596039e^{-1.185801404t} + 0.00001355362330e^{-1.106407020t} - 0.00008716725944e^{-1.000077209t} - 131.2510510e^{-0.007714366521t} + 0.0008795172311e^{-1.020000000t} + 0.001366413740e^{-2.060000000t} + 0.0007287053810e^{1.040000000t} + 131.2413316\} - (K_2 - K_3)t. \quad (61)$$

Table 9 and figure 8 are obtained using the same parameters as in equation (60).

RESULTS AND DISCUSSION

For investigating the effects of online and offline preventive maintenance, we calculate the system availability, reliability, mean time to failure (MTTF), MTTF sensitivity, and cost function for the established models numerically. When failure rates are set at various values, figure 2 shows how availability changes over time. The figure depicts how the availability of repairable systems declines over time, but eventually stabilizes at zero after a sufficiently long interval of time. The graphical representation of the model can predict the future behavior of a complex system for a given set of parameters at any time. The values of availability when repair follows a general distribution are similar to the values of availability when repair follows the Gumbel-Hougaard family copula distribution; this can be seen in figure 2 and 3. This is attributable to the system's online and offline preventive maintenance. However, there is no denying that using copula repair increases system's availability over general repair.

Figure 4 shows the difference in reliability over time. When compared to system availability, it can be seen that system reliability plummets. This is due to lack of system repairs. The general consensus is that the less the repairs, the lower the reliability. This illustrates what failure to manage the structure/system entails.

Figure 5 illustrates the differences in the value of the system's mean time to failure (MTTF) for failure rates λ_1, λ_2, δ_0 and δ_1 respectively, fixing other parameters constant. The MTTF values with respect to δ_0 and δ_1 is much higher than the MTTF values corresponding to λ_1 and λ_2. According to this sensitivity study, online and offline preventive maintenance failure rates are more responsible for proper system functioning. Figure 5 illustrates this.

The information on the sensitivity analysis conducted in this paper is shown in figure 6.

Figures 7 and 8 demonstrate the difference in effective benefit over time when the repair follows a copula distribution or a general distribution, respectively, with revenue cost per unit time is set at 1.0, cost of online and offline maintenance set at 0.1, the service cost is varied and the failure rates constant. Figures 7 and 8 show that when the repair follows the copula distribution, the expected gain is higher than when the repair follows the general distribution. Because of online and offline preventive maintenance, the expected benefit when repairing with copula distribution is comparable to the expected profit when repairing with general repair. According to this sensitivity review, the importance of using copula repair would be decreased if online and offline preventive maintenance is mandated. However, copula repair is still recommended for the system's proper operation.

CONCLUSION

The reliability metrics for different failure values and repair rates are critically analyzed in order to measure the performance of the systems under consideration. The system's transient probabilities and reliability metrics such as availability, reliability, mean time to failure and cost are calculated using the Markovian process, Laplace transformation, and supplementary variable approaches. The influence of time and other system characteristics on reliability metrics was simulated using MATLAB. Numerical experiments were used to obtain and validate the fundamental expressions such as availability, reliability, mean time to failure, sensitivity, and cost function. On the basis of the numerical results obtained for a particular case in figures 2-8 and tables 1-7, it is clear that the value, i.e., the importance of copula repairs has decreased with online and offline preventive maintenance. This means by implementing online and offline preventive maintenance, system failure would be greatly reduced, resulting in improved output and revenue generation. It is generally known that system failure will reduce production efficiency and may even result in a tragedy. This research paper will help system engineers, programmers, and maintenance managers design more important systems and maintain them in the best interests of humanity. Future iterations of this work may combine the Gumbel-Hougaard family Copula with the Bivariate Gumbel-Hougaard family Copula.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Niwas R, Garg H. An approach for analyzing the reliability and profit of an industrial system based on the cost-free warranty policy. J Braz Soc Mech Sci Eng 2018;40:265. [CrossRef]

[2] Gahlot M, Singh VV, Ayagi HI, Goel CK. Performance assessment of repairable system in series configuration under different types of failure and repair policies using copula linguistics. Int J Reliab Saf 2018;12:348-363. [CrossRef]

[3] Garg H. An approach for analyzing the reliability of industrial system using fuzzy Kolmogrov's differential equations. Arab J Sci Eng 2016a;40:975-987. [CrossRef]

[4] Garg H. A novel approach for analyzing the reliability of series-parallel system using credibility theory and different types of intuitionistic fuzzy numbers. J Braz Soc Mech Sci Eng 2016a;38:1021-1035. [CrossRef]

[5] Singh VV, Ayagi HI. Stochastic analysis of a complex system under preemptive resume repair policy using Gumbel-Hougaard family of copula. Int J Math Oper 2018;12:273-292. [CrossRef]

[6] Bala, SI, Yusuf I. Markov chain profit modeling and evaluation between two dissimilar systems under two forms of failures. Appl Appl Math 2016;11:489–503.

[7] Sanusi A, Yusuf I, Mamuda BY. Performance evaluation of an industrial configured as series-parallel system. J Math Comput Sci 2020;10:692–712.

[8] Garg H. Multi objective non-linear programming problem for reliability optimization in intuitionistic fuzzy environment. Front Inform Syst 2018;2:197–229. [CrossRef]

[9] Wang J, Xie N, Yang N. Reliability analysis of a two dissimilar unit warm standby repairable system with priority in use. Commun Stat Theory Methods 2019;50:792–814. [CrossRef]

[10] Kumar N, Pant S, Singh SB. Availability and cost analysis of an engineering system involving subsystems in series configuration. Int Qual Reliab Manag 2017;34:879–894. [CrossRef]

[11] Yusuf I, Fagge NJ. Evaluation of some reliability characteristics of a single unit system requiring two types of supporting device for operations. AAM 2017;12:803–812.

[12] Singh VV, Lado A. Cost assessment of complex repairable system consisting of two subsystems in the series configuration using Gumbel-Hougaard family copula. Int J Qual Rel Manag 2019;36:1683-1698. [CrossRef]

[13] Singh VV, Rawal DK. Stochastic analysis of N-unit plant with different types of failure and repair policy using copula distribution. J Rel Stat Stud 2015;8:119-133.

[14] Adhikary DD, Bose GK, Bose D, Mitra S. Maintenance class-based cost-effective preventive maintenance scheduling of coal-fired power plants. Int J Reliab Saf 2013;7:358–371. [CrossRef]

[15] Niwas R, Kadyan MS, Kumar J. Probabilistic analysis of two reliability models of a single-unit system with preventive maintenance beyond warranty and degradation. Eksploatacja I Niezawodnosc Maint Reliab 2015;17:535–543. [CrossRef]

[16] Mujahid SN, Abdur Rahim M. Optimal preventive maintenance warranty policy for repairable products with periodically increasing failure rate. Int J Oper Res 2010;9:227–240. [CrossRef]

[17] Uprety I. Stochastic analysis of a reheating-furnace system subject to preventive maintenance and repair. Int J Oper Res 2012;13:43–52. [CrossRef]

[18] Kadyan MS. Reliability and profit analysis of a single-unit system with preventive maintenance subject to maximum operation time. Eksploatacja i Niezawodnosc Maint Reliab 2013;15:176–181.

[19] Wang S. Integrated model of production planning and imperfect preventive maintenance policy for single machine system. Int J Oper Res 2013;18:140–156. [CrossRef]

[20] Yusuf I. Reliability modeling of a parallel system with a supporting device and two types of preventive maintenance. Int J Oper Res 2016;25:269–287. [CrossRef]

[21] Kumar A, Saini M, Malik SC. A single unit system with preventive maintenance and repair subject to maximum operation and repair times. Int J Appl Math Comput 2014;6:25–36.

[22] Kumar A, Saini M, Devi K. Stochastic modeling of non-identical redundant systems with priority, preventive maintenance, and Weibull failure and repair distributions. Life Cycle Reliab Saf Eng 2018a;7:61–70. [CrossRef]

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1101–1107, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2023.00024

Research Article

Synthesis of silica xeorogel@Mg-Al layered double hydroxide composite for CO_2 capture

Dicle EREN[1], Muge SARI YILMAZ[1,*]

[1]*Faculty of Chemical and Metallurgical Engineering, Yildiz Technical University, İstanbul, Türkiye*

ARTICLE INFO

Article history
Received: 07 November 2022
Revised: 04 January 2023
Accepted: 08 February 2023

Keywords:
CO_2 Capture; Kinetics; Layered Double Hydroxide; Xerogel

ABSTRACT

Today, the increase in the level of carbon dioxide (CO_2) gas in the atmosphere has started to cause concern. Therefore, an appropriate and rapid decrease in CO_2 gas emission levels has become a significant challenge. Capturing CO_2 on a solid surface is proposed due to its ease of application, relatively low energy requirements, and applicability in various processes. This study investigated the preparation of Xerogel@MgAl LDH (X@MAL) composite for CO_2 capture. Firstly, silica-based xerogel was synthesized by the acid and base-catalyzed two-step sol-gel method. Then, the X@MAL composite was prepared by the co-precipitation method. Based on the CO_2 capture analysis, the maximum CO_2 capture capacity of the composite at 25 °C, 75 °C, and 100 °C was 1.90 mmol.g^{-1}, 0.70 mmol.g^{-1}, and 0.40 mmol.g^{-1}, respectively. The kinetic analysis results show that the CO_2 capture of X@MAL can be well-defined by Avrami kinetic model.

Cite this article as: Eren D, Sarı Yılmaz M. Synthesis of silica xeorogel@Mg-Al layered double hydroxide composite for CO_2 capture. Sigma J Eng Nat Sci 2024;42(4):1101–1107.

INTRODUCTION

Emissions of CO_2 into the atmosphere are increasing significantly due to the energy industry, transportation sector, and human activities [1]. The excessive increase in the concentration of greenhouse gases, especially CO_2, is due to the burning of fossil fuels, which causes global warming and other environmental effects [2]. It is extremely important to develop technologies that reduce greenhouse gas emissions [3].

Therefore, the most promising method to mitigate the impact of CO_2 on global climate is CO_2 capture from fossil fuels consumed by power plants. The process economies of such technologies are often not cheap enough to offset the hold costs. Thus, it is highly desirable to develop alternatives that are more energy efficient than conventional separation technologies.

Among these techniques, adsorption is one of the most promising approaches as it can reduce the costs associated with the hold step. In general, high CO_2 capture capacity, high selectivity, low material cost, and stable adsorption capacity after several cycles are the main characteristics of CO_2 adsorbents [3].

Recently, scientists have made dramatic efforts to produce suitable adsorbents capable of meeting the requirements of CO_2 emission sources. Many of the scientific community have worked on the development of various types of porous solid adsorbents based on activated carbon

*Corresponding author.
*E-mail address: mugesari@yildiz.edu.tr

This paper was recommended for publication in revised form by Editor in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey

[4,5], mesoporous silica materials [6,7], zeolites [8], metal-organic frameworks (MOFs) [9], mesocellular foams 10, and layered double hydroxides (LDHs) [11].

LDHs are inorganic materials that are members of the family of two-dimensional clay minerals with multiple metal cations in their inner layers and anions in their interlayers. The structure of LDHs consist of brushite $[Mg(OH)_2]^+$ like positively charged metal hydroxide layers and anions and water molecules that exist as charge balancers between the layers. The potential use of LDHs in high-temperature carbon capture and storage applications is promising [12]. LDHs require less energy for regeneration in CO_2 adsorption and show better stability than some solid adsorbents [13]. They also show rapid adsorption-desorption kinetics, especially in the presence of water. This has made LDHs interesting in pre-combustion CO_2 capture applications [14,15].

Xerogels are used as supporting materials in the preparation of various composites due to their unique properties [16,17]. By uniting LDHs with xerogel, which is found in different forms such as a monolith, powder, fiber, or film, composite materials with superior properties are created.

Xerogels are generally prepared by the sol-gel method under atmospheric pressure. Some properties such as the pore of xerogels can scale, high surface area, low intensity, thermal conductivity, dielectric constant, high optical permeability in visible light, and sound insulation of this material make it unique for several scientific and technological applications [18]. The brittleness, moisture-holding, and instability of silica xerogels in long-term applications can limit the use of these materials. These disadvantages can be eliminated with surface modification, heat treatment, or preparation of composite materials by adding various fillers [19].

To the author's knowledge, only one study has been found in the literature on the combination of xerogel with MgAl LDH. Okada et al. prepared the MgAl LDH and aluminosilicate xerogel composites for CO_2 and NH_3 gas adsorption [20]. The composition of xerogel and the synthesis procedure of the composite differs from our study.

In this study, X@MAL composite was synthesized for the first time for CO_2 capture application. Firstly, silica-based xerogel was synthesized by the acid and base-catalyzed two-step sol-gel method. Then, the X@MAL composite was prepared by the co-precipitation method by using the synthesized xerogel. The obtained material was characterized by X-ray diffraction (XRD) and Fourier transform infrared (FTIR) analysis. The CO_2 capture capacity of the composite was carried out using by gravimetric method at different adsorption temperatures.

EXPERIMENTAL PROCEDURES

Materials and Characterization

The materials used in the xerogel synthesis were supplied by Sigma-Aldrich. For the synthesis of X@MAL composite magnesium nitrate hexahydrate ($Mg(NO_3)_2.6H_2O$, Merck) aluminum nitrate nonahydrate ($Al(NO_3)_3.9H_2O$, Merck), sodium hydroxide (NaOH, Merck) sodium carbonate (Na_2CO_3, Merck) were used.

The characterization analysis of the composite was carried out using a PANalytical X'Pert-Pro XRD diffractometer and Perkin-Elmer Spectrum One FT-IR spectrophotometer. The XRD measurements of xerogel and X@MAL composite were taken in the 2θ range from 5° to 100°.

Infrared spectra were collected using the KBr pellet technique in absorbance mode within the range from 4000 to 450 cm^{-1}.

The CO_2 capture analysis of the composite was studied using Perkin Elmer Pyris Diamond thermogravimetric equipment. In the capture analysis, pure CO_2 (>99.99%) gas was used.

Briefly, approximately 10 mg of sample was put in an alumina crucible, and then, it was heated to 105°C under a nitrogen atmosphere to remove its moisture. Afterward, the ambient temperature was reduced to a specific temperature at a cooling rate of 10°C.min^{-1}. For the capture analysis, the gas was changed to CO_2 at a flow rate of 100 ml per minute for 90 minutes. The weight increase of the sample was used to determine the sample's CO_2 capture capacity.

Synthesis of X@MAL Composite

To synthesize the X@MAL composite, the xerogel was first synthesized using our previous synthesis procedure available in the literature [21].

For the synthesis of the X@MAL composite, two different solutions were prepared. Initially, 0.1 g xerogel was dissolved in distilled water under ultrasonic irradiation. Then, Na_2CO_3 is added to the xerogel solution and it continued the mixing for a while. The obtained solution was named the first solution and its pH of it was adjusted to 10. In order to prepare the second solution, 0.25 g $Mg(NO_3)_2{\cdot}6H_2O$ and 0.18 g $Al(NO_3)_3{\cdot}9H_2O$ were dissolved in the distilled water under ultrasonic irradiation. The obtained second solution is slowly added to the first solution under stirring at room temperature. The resultant solution was continued to stir for 1 hour at room temperature. At the end of the time, the precipitation was collected by centrifugation, and then, it is washed with distilled water and ethanol separately. Finally, it was centrifuged again and dried under a vacuum.

RESULTS AND DISCUSSION

Characterization of Adsorbent

Figure 1 demonstrates the XRD pattern of xerogel and X@MAL composite. The XRD analysis of xerogel showed one broad diffraction peak observed between 20-30° indicating its amorphous structure [22]. In the XRD analysis of the X@MAL composite, the characteristic amorphous peak of xerogel and newly formed (003), (009), (015), and (110) diffraction peaks related to the LDH phase were seen,

indicating that the layered Mg-Al LDH was formed in the xerogel structure. The characteristic (006) peak of Mg-Al LDH overlapped with the broad peak of xerogel, therefore the (006) peak of Mg-Al LDH was not observed in the patterns of X@MAL.

The FT-IR spectrum of xerogel and X@MAL composite were given in Figure 2. In the FT-IR spectrum of xerogel, the characteristic peak at 1087 cm^{-1} corresponded to the asymmetric stretching of Si–O–Si groups [23]. The peaks at 3448 cm^{-1} and 1636 cm^{-1} were related to physically absorbed water or structural –OH groups and adsorbed water molecules, respectively [24]. The Si-OH stretching vibration was observed at 946 cm^{-1} [25]. The peaks at 797 cm^{-1} and 467 cm^{-1} were attributed Si-O-Si symmetric stretching vibrations and bending mode, respectively [26]. Compared with xerogel, the new peaks at around 1498 and 1385 cm^{-1} were observed in the composite, which can be attributed to the vibrations of carbonate species originating from the LDH structure. In addition, differences were observed in the 1000-450 cm^{-1} region due to the stretching of the Al-O and Mg-O metal-oxygen bonds [27–29].

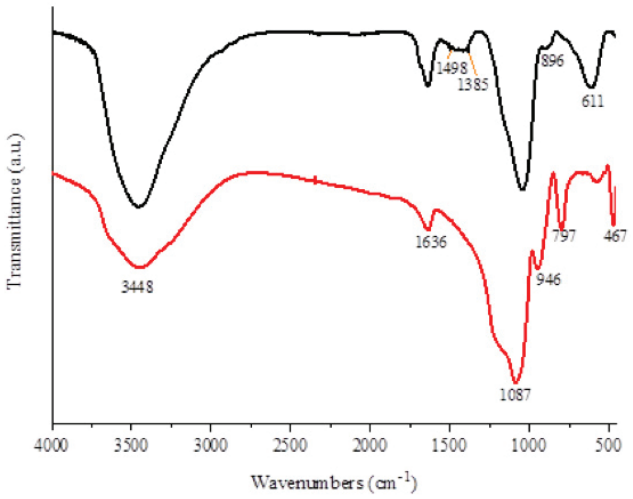


Figure 2. FTIR spectrum of composite (black line) and xerogel (red line).

CO_2 Capture Measurement

The amounts of captured CO_2 at different temperatures (25°C, 75°C, and 100°C) were given in Figure 3. The CO_2 capture capacity of the X@MAL composite decreases with increasing temperature, which is characteristic of physisorption. The CO_2 capture capacity at 75 °C and 100 °C remained constant after 35 minutes, while its capacity at 25 °C remained stable after 80 minutes. The adsorption capacities of the X@MAL composite at different temperatures followed by following trend: 1.90 mmol.g-1 (25°C) > 0.70 mmol.g^{-1} (75°C) > 0.40 mmol.g^{-1} (100°C). At 25°C, the X@MAL exhibited approximately 5 times more capacity than the adsorption capacity at 100°C. Accordingly, it was observed that the CO_2 adsorption capacity of X@MAL decreased with the increase in temperature. This indicates that the adsorption of CO_2 into the composite is an exothermic process.

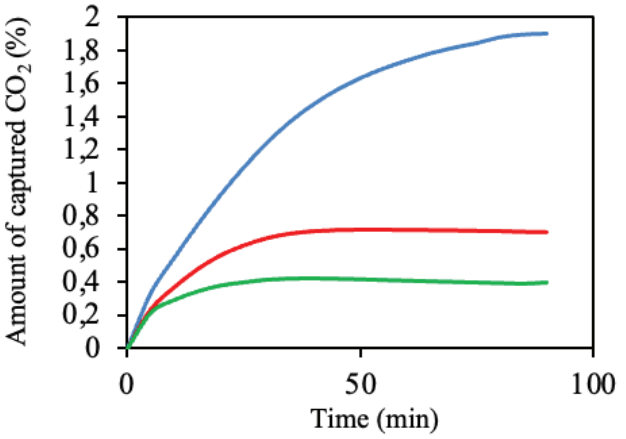


Figure 3. The CO_2 capture capacities of X@MAL composite at different temperatures (Blue line: 25°C, green line: 75°C, and red line: 100°C).

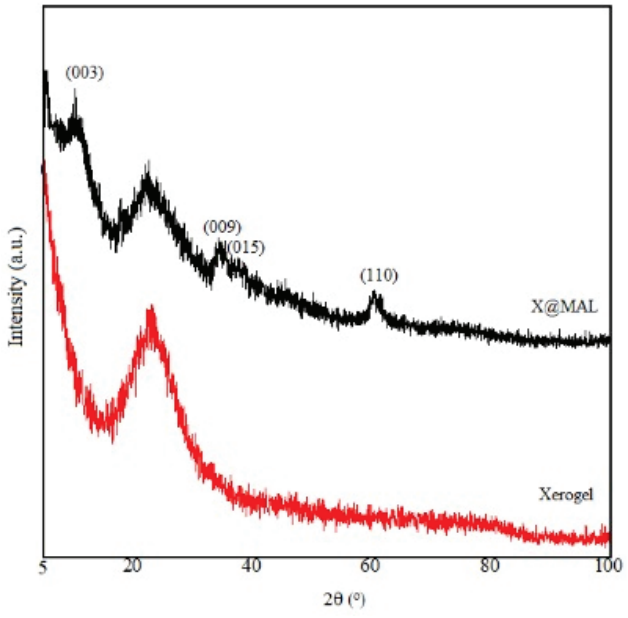


Figure 1. The XRD patterns for silica based xerogel and the X@MAL

Kinetic Analysis

To determine the mechanism of CO_2 adsorption of X@MAL composite, several of the most common kinetic models from the pseudo-first-order (FO), the pseudo-second-order (SO), and the Avrami (Av) model were selected. By using the CO_2 adsorption results at different temperatures, the compatibility of these models with the experimental values will be tested.

FO model was proposed by Lagergren and assumes that the rate of adsorption is proportional to the number of free active sites on the accessible adsorbent surface [30]. This equation can be given by:

$$q_t = q_e \left[1 - exp(-k_f\, t)\right] \quad (1)$$

where q_t (mg/g) and q_e (mg/g) are the amounts of adsorbed CO_2 at time t (min) and equilibrium. k_f (min^{-1}) is the FO rate constant.

SO model was proposed by Ho et al. [31]. The adsorption rate with respect to SO is directly proportional to the square of the number of free active sites on the adsorbent and the kinetic model is expressed by the equation given below.

$$q_t = \frac{q_e^2\, k_s\, t}{1 + q_e\, k_s\, t} \quad (2)$$

where k_s (g/(mg.min) is the SO rate constant.

The Avrami method is a first-order fractional kinetics for particle nucleation and has recently been used to describe CO_2 capture on solid adsorbents [32,33]. This model was described using the following equation:

$$q_t = q_e \left[1 - exp\ exp\ (-(k_a\, t)^{n_a}\right] \quad (3)$$

Where k_a and n_a are the rate constant and kinetic order of Av, respectively.

The error was determined using the following equation to determine the accuracy of the kinetic model:

$$Error(\%) = \sqrt{\frac{\sum_{i=1}^{p}\left[\frac{q_{e(exp)} - q_{e(model)}}{q_{e(exp)}}\right]^2}{p-1}}\ x\ 100\% \quad (4)$$

where $q_{e(exp)}$ and $q_{e(model)}$ are the value of adsorption capacity experimentally and obtained from the fitted model, respectively and p is the number of total experimental data.

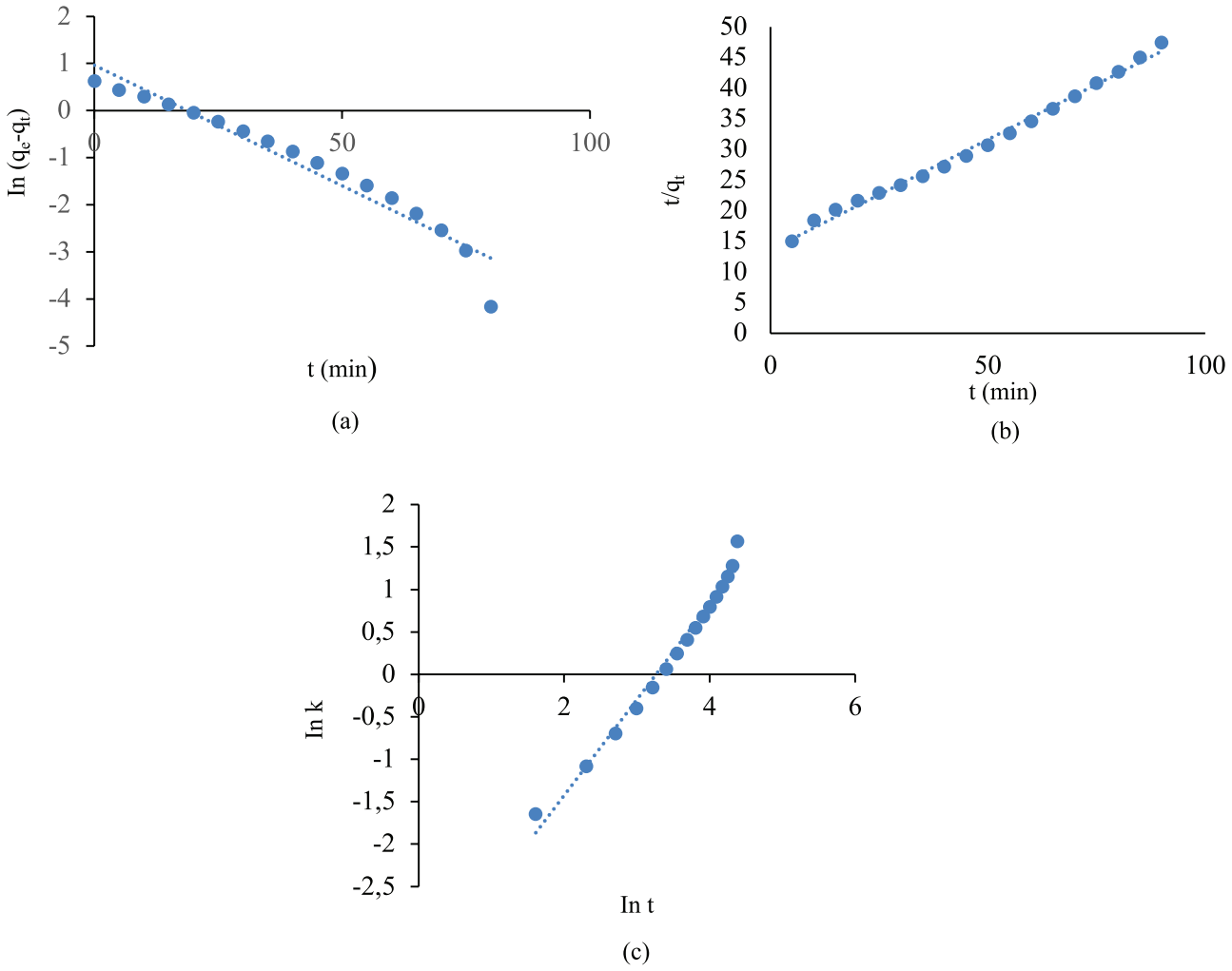


Figure 4. The kinetic model curves of CO_2 capture on X@MAL at 25°C (a) FO, (b) SO, and (c) Av.

All kinetic models were fitted to explain the capture of CO_2 on X@MAL at 25°C (Figure 4). The calculated kinetic parameters of the fitted models are demonstrated in Table 1. It is seen that the SO provides the best fit among the three kinetic models as per the correlation coefficient (R^2) values for X@MAL. However, it was found that the Av kinetic model presented a lower error value than the error values calculated from the FO and SO. In addition, when the fitted data shown in Figure 5 are examined, it is seen that the fitted data from the Av model are very close to the experimental data. Thus, the Av model is applied to predict the CO_2 capture process of X@MAL. The kinetic order of Av (na) is found as 1.128 indicating the multiple kinetic order of the CO_2 capture.

Table 1. The kinetic parameters for CO_2 capture on X@MAL composite at 25°C

Kinetic Model	Parameter	Value
F_O	$q_{e,\,exp}$ (mmol.g^{-1})	1.890
	k_f	0.051
	R^2	0.941
	Error (%)	9.712
S_O	q_e (mmol.g^{-1})	2.624
	k_s	0.010
	R^2	0.994
	Error (%)	11.363
A_v	q_e (mmol.g^{-1})	1.934
	k_a	0.038
	n_a	1.128
	R^2	0.983
	Error (%)	0.600

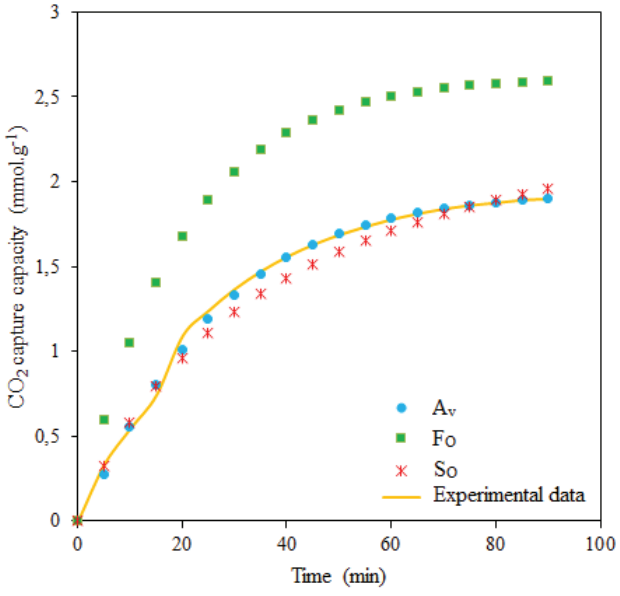


Figure 5. Corresponding fit of X@MAL applying three kinetic models.

Table 2. Comparison of the CO_2 adsorption capacity of various composites

Adsorbent	Capacity (mmol.g^{-1})	References
Silica xerogel-based		
MMS-6	1.90	[3]
PEI functionalized xerogel	1.94	[34]
Polyurethane/ silica xerogels	1.10	[35]
MgAl LDH-based		
MgAl LDH	0.9	[36]
POM/MgAl LDH	0.74	[37]
GO/MgAl LDH	0.54	[38]
X@MAL	1.90	In this study

Comparison with Other Adsorbents

A comparison of the CO_2 uptake capacity of silica xerogel and MgAl LDH-based adsorbents found in the literature are listed in Table 2. Compared with the silica xerogel-based adsorbents, X@MAL showed either close or better adsorption capacity. The CO_2 adsorption performance of X@MAL is superior to MgAl LDH-based adsorbents. Overall, it is seen that the X@MAL composite has a good enough capacity to compete with the adsorbents in the literature.

CONCLUSION

In this study, the X@MAL composite was prepared by the co-precipitation method. The XRD and FTIR characterization results showed that the composite was synthesized successfully. The CO_2 capture experiments at different temperatures indicated that the maximum capture capacity of the composite was 1.90 mmol.g^{-1} at 25°C. Adsorption kinetics demonstrated that the Av model had a better fitting effect and was more suitable for describing the CO_2 capture process.

ACKNOWLEDGMENT

This work was supported by Yildiz Technical University Scientific Research Projects Coordination Unit. Project Number: FLY-2022-4884.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw

data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Loganathan S, Tikmani M, Edubilli S, Mishra A, Ghoshal AK. CO_2 adsorption kinetics on mesoporous silica under wide range of pressure and temperature. Chem Eng J 2014;256:1–8. [CrossRef]

[2] Aquino AS, Vieira MO, Ferreira ASD, Cabrita EJ, Einloft S, de Souza MO. Hybrid ionic liquid-silica xerogels applied in CO_2 capture. Appl Sci 2019;9:2614. [CrossRef]

[3] Witoon T, Tatan N, Rattanavichian P, Chareonpanich M. Preparation of silica xerogel with high silanol content from sodium silicate and its application as CO_2 adsorbent. Ceram Int 2011;37:2297–2303. [CrossRef]

[4] Serafin J, Sreńscek-Nazzal J, Kamińska A, Paszkiewicz O, Michalkiewicz B. Management of surgical mask waste to activated carbons for CO_2 capture. J CO_2 Util 2022;59:101970. [CrossRef]

[5] Ismail IS, Rashidi NA, Yusup S. Production and characterization of bamboo-based activated carbon through single-step H_3PO_4 activation for CO_2 capture. Environ Sci Pollut Res Int 2022;29:12434–12440. [CrossRef]

[6] Sari Yilmaz M. The CO_2 adsorption performance of aminosilane-modified mesoporous silicas. J Therm Anal Calorim 2021;146:2241–2251. [CrossRef]

[7] Sari Yilmaz M, Karakas SB. Low-cost synthesis of organic-ınorganic hybrid MSU-3 from gold mine waste for CO_2 adsorption. Water Air Soil Pollut 2018;229:326. [CrossRef]

[8] Choi HJ, Hong SB. Effect of framework Si/Al ratio on the mechanism of CO_2 adsorption on the small-pore zeolite gismondine. Chem Eng J 2022;433:133800. [CrossRef]

[9] Lei L, Cheng Y, Chen C, Kosari M, Jiang Z, He C. Taming structure and modulating carbon dioxide (CO_2) adsorption isosteric heat of nickel-based metal organic framework (MOF-74(Ni)) for remarkable CO_2 capture. J Colloid Interface Sci 2022;612:132–145. [CrossRef]

[10] Bai F, Liu X, Sani S, Liu Y, Guo W, Sun C. Amine functionalized mesocellular silica foam as highly efficient sorbents for CO_2 capture. Sep Purif Technol. 2022;299:121539. [CrossRef]

[11] Sari Yilmaz M. Synthesis of novel amine modified hollow mesoporous silica@Mg-Al layered double hydroxide composite and its application in CO_2 adsorption. Microporous Mesoporous Mater 2017;245:109–117. [CrossRef]

[12] Garcia-Gallastegui A, Iruretagoyena D, Gouvea V, Mokhtar M, Asiri AM, Basahel SN, et al. Graphene oxide as support for layered double hydroxides: Enhancing the CO_2 adsorption capacity. Chem Mater 2012;24:4531–4539. [CrossRef]

[13] Choi S, Drese JH, Jones CW. Adsorbent materials for carbon dioxide capture from large anthropogenic point sources. ChemSusChem 2009;2:796–854. [CrossRef]

[14] Ding Y, Alpay E. Equilibria and kinetics of CO2 adsorption on hydrotalcite adsorbent. Chem Eng Sci 2000;55:3461–3474. [CrossRef]

[15] Van Selow ER, Cobden PD, Verbraeken PA, Hufton JR, Van Den Brink RW. Carbon capture by sorption-enhanced water-gas shift reaction process using hydrotalcite-based material. Ind Eng Chem Res 2009;48:4184–4193. [CrossRef]

[16] dos Santos-Gómez L, Cuesta N, Cameán I, García-Granda S, García AB, Arenillas A. A promising silicon/carbon xerogel composite for high-rate and high-capacity lithium-ion batteries. Electrochim Acta 2022;426:140790. [CrossRef]

[17] Du G, Seng KH, Guo Z, Liu J, Li W, Jia D, et al. Graphene-V_2O_5·nH_2O xerogel composite cathodes for lithium ion batteries. RSC Adv 2011;1:690. [CrossRef]

[18] Guzel Kaya G, Deveci H. Synergistic effects of silica aerogels/xerogels on properties of polymer composites: A review. J Ind Eng Chem 2020;89:13–27. [CrossRef]

[19] Owens GJ, Singh RK, Foroutan F, Alqaysi M, Han CM, Mahapatra C, et al. Sol-gel based materials for biomedical applications. Prog Mater Sci 2016;77:1–79. [CrossRef]

[20] Okada K, Kaneda A, Kameshima Y, Yasumori A. Acidic and basic gas adsorption properties in composites of layered double hydroxide/aluminosilicate xerogels. Mater Res Bull 2002;37:209–219. [CrossRef]

[21] Sezgin D, Sari Yilmaz M. Xerogel of fast kinetics and high adsorption capacity for cationic dye removal. Sigma J Eng Nat Sci 2024;42:189–197. [CrossRef]

[22] Hu W, Li M, Chen W, Zhang N, Li B, Wang M, et al. Preparation of hydrophobic silica aerogel with kaolin dried at ambient pressure. Colloids Surf A Physicochem Eng Asp 2016;501:83–91. [CrossRef]

[23] Estella J, Echeverría JC, Laguna M, Garrido JJ. Effects of aging and drying conditions on the structural and textural properties of silica gels. Micropor Mesopor Mat 2007;102:274–282. [CrossRef]

[24] Sari Yilmaz M, Dere Özdemir Ö, Pişkin S. Synthesis and characterization of MCM-41 with different methods and adsorption of Sr2+ on MCM-41. Res Chem Intermed 2015;41:199–211. [CrossRef]

[25] Sari Yilmaz M, Piskin S. Evaluation of novel synthesis of ordered SBA-15 mesoporous silica from gold mine tailings slurry by experimental design. J Taiwan Inst Chem Eng 2015;46:176–182. [CrossRef]

[26] Mohammadian M, Kashi TSJ, Erfan M, Soorbaghi FP. Synthesis and characterization of silica aerogel as a promising drug carrier system. J Drug Deliv Sci Technol 2018;44:205–212. [CrossRef]

[27] Barahuie F, Hussein MZ, Arulselvan P, Fakurazi S, Zainal Z. Drug delivery system for an anticancer agent, chlorogenate-Zn/Al-layered double hydroxide nanohybrid synthesised using direct co-precipitation and ion exchange methods. J Solid State Chem 2014;217:31–41. [CrossRef]

[28] dos Santos RMM, Gonçalves RGL, Constantino VRL, Santilli CV, Borges PD, Tronto J, et al. Adsorption of acid yellow 42 dye on calcined layered double hydroxide: Effect of time, concentration, pH and temperature. Appl Clay Sci 2017;140:132–139. [CrossRef]

[29] Auxilio AR, Andrews PC, Junk PC, Spiccia L. The adsorption behavior of C.I. Acid Blue 9 onto calcined Mg-Al layered double hydroxides. Dye Pigment 2009;81:103–112. [CrossRef]

[30] Jia Z, Li Z, Ni T, Li S. Adsorption of low-cost absorption materials based on biomass (Cortaderia selloana flower spikes) for dye removal: kinetics, isotherms and thermodynamic studies. J Mol Liq 2017;229:285–292. [CrossRef]

[31] Ho YS, McKay G. The kinetics of sorption of basic dyes from aqueous solution by sphagnum moss peat. Can J Chem Eng 1998;76:822–827. [CrossRef]

[32] Serna-Guerrero R, Sayari A. Modeling adsorption of CO_2 on amine-functionalized mesoporous silica. 2: Kinetics and breakthrough curves. Chem Eng J 2010;161:182–190. [CrossRef]

[33] Songolzadeh M, Soleimani M, Takht Ravanchi M. Using modified Avrami kinetic and two component isotherm equation for modeling of CO_2/N_2 adsorption over a 13X zeolite bed. J Nat Gas Sci Eng 2015;27:831–841. [CrossRef]

[34] Kaya GG, Deveci H. CO_2 capture using polyethyleneimine functionalized silica xerogels. Konya Müh Bil Derg 2021;9:1109–1118. [CrossRef]

[35] Santos LMD, Bernard FL, Pinto IS, Scholer H, Dias GG, Prado M, Einloft S. Polyurethane/ionic silica xerogel composites for CO_2 capture. Mater Res 2019;22:e20190022. [CrossRef]

[36] Kou X, Guo H, Ayele EG, Li S, Zhao Y, Wang S, Ma X. Adsorption of CO_2 on MgAl-CO_3 LDHs-derived sorbents with 3D nanoflower-like structure. Energ Fuel 2018;32:5313–5320. [CrossRef]

[37] Gunjakar JL, Kim IY, Hwang SJ. Efficient hybrid-type CO_2 Adsorbents of reassembled layered double hydroxide 2D nanosheets with polyoxometalate 0D nanoclusters. Eur J Inorg Chem 2015;7:1198–1202. [CrossRef]

[38] Iruretagoyena D, Shaffer MS, Chadwick D. Layered double oxides supported on graphene oxide for CO_2 adsorption: Effect of support and residual sodium. Ind Eng Chem Res 2015;54:6781–6792. [CrossRef]

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1108–1115, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00091

Research Article

Bayesian estimation of inverse weibull distribution scale parameter under the different loss functions

Esin KÖKSAL BABACAN[1,*]

[1]*Department of Statistics, Faculty of Science, Ankara University, Ankara, 06100, Türkiye*

ARTICLE INFO

Article history
Received: 29 November 2022
Revised: 12 March 2023
Accepted: 26 June 2023

Keywords:
Bayesian Estimation; Inverse Weibull Distribution, Loss Function

ABSTRACT

In this paper, the Bayesian estimators for the Inverse Weibull Distribution (IWD) scale parameter are derived when the shape parameter of distribution is known. The Bayesian estimators for the parameter are obtained by using the Gamma prior under the different types of loss functions such as square error loss function (Self), Entropy loss function (Elf), Precautionary loss function (Plf), Linear exponential loss function (Linexlf) and nonlinear exponential loss function (Nlinexlf). A classical maximum likelihood estimator (mle) for the parameter is also derived. To compare the efficiency of the parameter estimation methods, a simulation study is carried out. The comparison is based on mean square error.

Cite this article as: Köksal Babacan E. Bayesian estimation of inverse weibull distribution scale parameter under the different loss functions. Sigma J Eng Nat Sci 2024;42(4):1108–1115.

INTRODUCTION

The Weibull distribution is frequently used in reliability engineering, especially for analyzing lifetime data. The probability density function (pdf) of the Weibull is a uni-modal or decreasing function. Also, the Weibull hazard function depends on its shape parameter. Depending on the parameter's value, the hazard function decreases or increases. If the data have a non-monotone hazard function, the Weibull distribution is not considered the appropriate model, for example, lung and breast cancer patients' mortalities [1-3]. In these circumstances, the problem is to find an appropriate distribution for the analysis of such data sets. Kundu and Howlader (2010) remarked that the IWD is an appropriate model for these data sets [4]. The IWD might be considered a suitable model when the study concludes that the pdf of the data can be unimodal [3].

Keller and Kamath (1982) used the IWD to investigate the failures of mechanical components subject to deterioration [5]. Calabria and Pulcini (1994) studied parameter estimations of IWD based on classical and Bayesian methods [6]. In the Bayesian aspect, they used informative priors. Some crucial theoretical properties of the IWD are given in [5]. Kundu and Howlader (2010) considered Bayesian inference for IWD type II censored data in their study [4]. Helu and Samawi (2015) studied progressively the first failure censoring data in their work and used Lindey's methods to derive a Bayesian estimator for IWD parameters [7]. Bi and Gui (2017) studied the stress-strength reliability of IWD [8]. Nasar and Kaser (2017) described frequentist and Bayesian estimation for the parameters of the IWD based

***Corresponding author.**
*E-mail address: ekoksal@science.ankara.edu.tr

This paper was recommended for publication in revised form by Editor in-Chief Ahmet Selim Dalkilic

on an adaptive type-II progressive hybrid censoring scheme [9]. To obtain Bayesian estimation, they used the Lindley approximation. Singh and Tripathi considered the parameter estimation of an IWD when it is known that samples are progressive type-I interval censored [10]. They proposed an EM algorithm to obtain maximum likelihood estimates and mid-point estimates. They obtained Bayes estimates under the square error loss function. Under the entropy loss function, Jana and Bera (2022) developed Bayes estimators of the IWD parameters [11]. They, also investigated the reliability of multi-component stress-strength model using classical and Bayesian approaches.

Suppose that Y is a random variable from the Weibull distribution. The pdf of Y is given as follows:

$$f(y;\lambda,\alpha) = \alpha\lambda y^{\alpha-1}e^{-\lambda y^{\alpha}}, y > 0;\ \lambda,\alpha > 0$$

where α is a shape and λ is a scale parameter of the distribution. If we take the transformation of Y with X = 1/Y, then X has the IWD, and the pdf of X is derived as follows:

$$f(x) = \alpha\lambda\left(x^{-(\alpha+1)}\right)\left(e^{-\lambda x^{-\alpha}}\right),\ x > 0;\ \alpha,\lambda > 0 \quad (1)$$

The cumulative distribution function (cdf) of X is written as

$$F(x) = e^{-\lambda x^{-\alpha}},\ x > 0. \quad (2)$$

The expected value and the variance of the IWD are

$$E(X) = \lambda^{\frac{1}{\alpha}}\Gamma\left(1-\frac{1}{\alpha}\right) \quad (3)$$

and

$$Var(X) = \lambda^{\frac{2}{\alpha}}\left(\Gamma\left(1-\frac{2}{\alpha}\right) - \left(\Gamma\left(1-\frac{1}{\alpha}\right)\right)^2\right). \quad (4)$$

respectively, where Γ is the Gamma function.

The reliability function of X,

$$R(x) = 1 - \left(e^{-\lambda x^{-\alpha}}\right), x \geq 0, \alpha, \lambda > 0 \quad (5)$$

and the hazard function

$$H(x) = \frac{\lambda(\alpha^{-\alpha-1})\left(e^{-\lambda x^{-\alpha}}\right)}{\left(1-e^{-\lambda x^{-\alpha}}\right)}, x \geq 0, \alpha, \lambda > 0 \quad (6)$$

is given. IWD has a scale parameter (α) and a shape parameter (λ). The λ is equal to the slope of the regression model, which is obtained via the graphical method. So it's known as a slope. When the slope takes a value between zero and one, failure rates increase. If α = 1, this distribution is called the Inverse Exponential distribution, and if α = 2, then the Inverse Rayleigh distribution. If the λ is a constant and the α increases, the kurtosis of the distribution increases, and parallel to this, the height decreases. If the value of the scale parameter decreases, the height of the curve increases, and this distribution is sharper. Density curves for different values of λ and α are shown in Figure 1 and Figure 2, respectively.

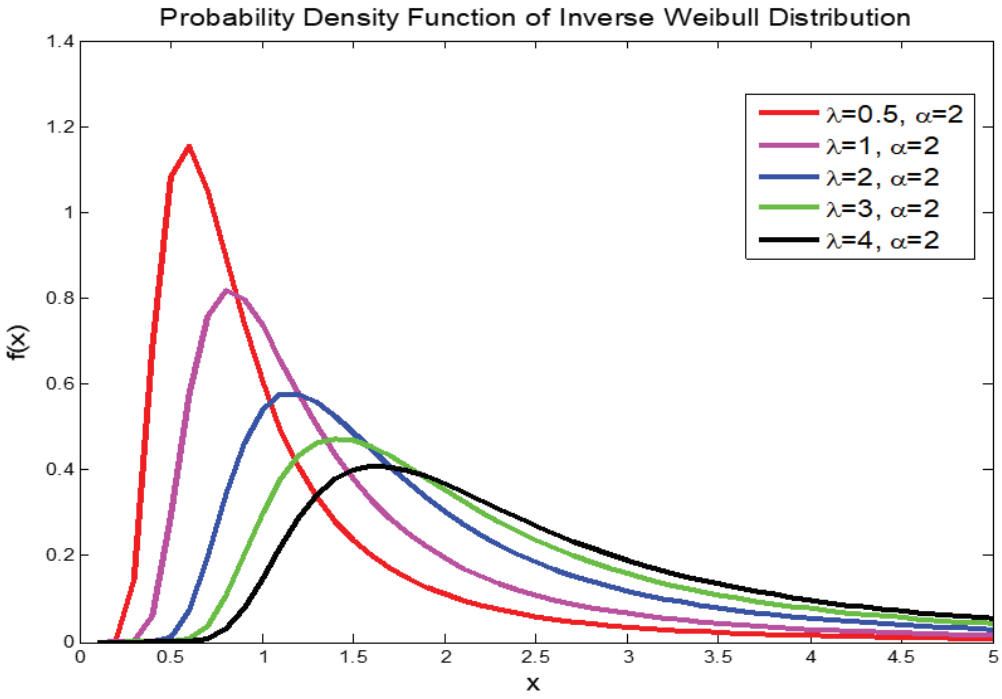


Figure 1. IWD density curves with various values of λ when $\alpha = 2$.

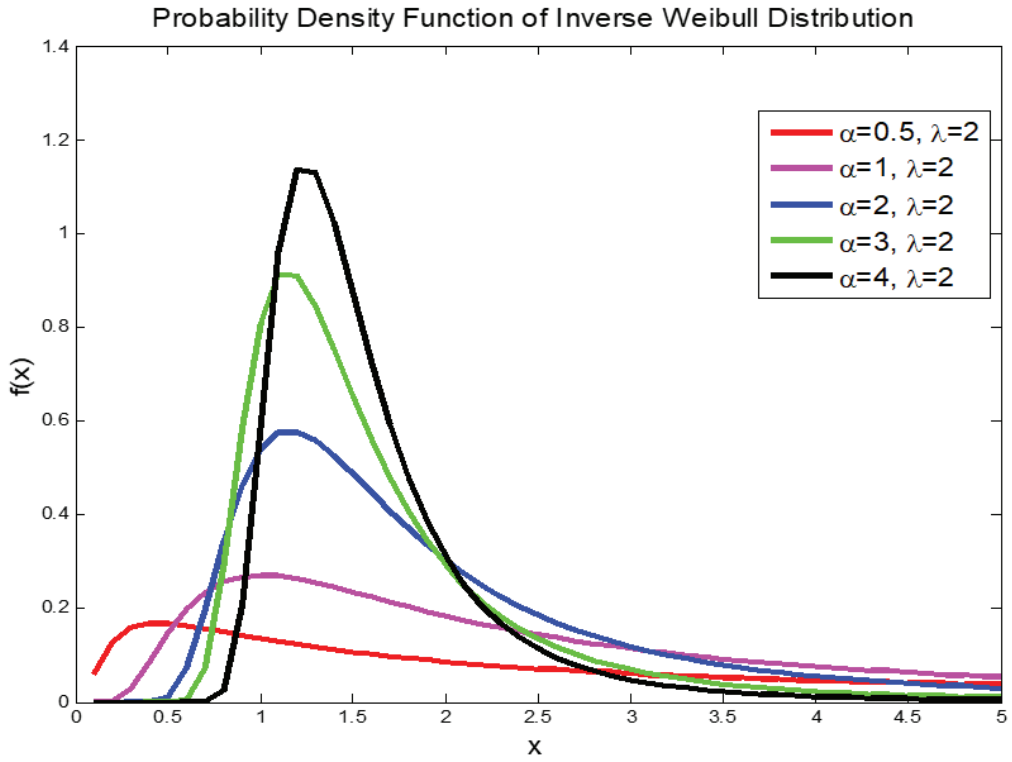


Figure 2. IWD density curves with various values α when $\lambda = 2$.

The IWD might be considered an alternative to the Log-Normal and the Gamma distributions, especially in life testing and reliability engineering. This study deals with the estimation of the scale parameter of the IWD. The estimation is considered in both classical and Bayesian aspects.

The rest of the study is as follows: In Section 2, the mle of λ is derived when the α is known. In Section 3, the Bayesian estimators with the Gamma prior under the Self, Elf, Plf, Linexlf, and Nlinexlf are discussed for the λ in IWD. In Section 4, the mle and the Bayes estimators for λ are compared based on the mean squared error criteria. In Section 5, the main observations of the simulation results are given. Concluding remarks are presented at the end of the paper.

MLE for the Scale Parameter of IWD

Mle is one of the most widely used techniques of estimation in statistics. In this section, for IWD, the mle for λ is derived when α is known. Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from the IWD when α is known. The likelihood is given as

$$L(\lambda|\alpha,x) = \alpha^n \lambda^n \prod_{i=1}^{n} x_i^{-(\alpha+1)} e^{-\lambda(\sum_{i=1}^{n} x_i^{-\alpha})} \quad (7)$$

then, the log-likelihood function (l) corresponding to (7) can be obtained as

$$l(\lambda|\alpha,x) = n(\log\alpha) + n(\log\lambda) - (\alpha+1)(\sum_{i=1}^{n}\log(x_i)) - \lambda(\sum_{i=1}^{n}(x_i^{-\alpha})) \quad (8)$$

By taking the first derivative of l relating to the parameter λ and setting the equation to zero, the following estimating equation can be derived:

$$\frac{dl(\lambda|\alpha,x)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i^{-\alpha} = 0. \quad (9)$$

Then mle for the λ

$$\hat{\lambda}_{mle} = \frac{n}{\sum_{i=1}^{n} x_i^{-\alpha}} \quad (10)$$

is obtained.

Bayesian Estimator for the Scale Parameter of IWD Under The Different Loss Functions

The selection of appropriate priors and appropriate loss functions are the most important issues in Bayesian inference. The symmetric square loss function is frequently used in Bayesian theory. In this loss function, positive and negative prediction errors are given equal weight. It may not be appropriate to use a symmetric loss function for a different valuation of estimation errors [6], [12], [13], [14]. This paper deals with comparing estimation results for different loss functions in IWD. Parsian and Kinmari (2002) and Misra and van der Meulen (2003) studied asymmetric Linexlf, and they estimated the Normal distribution location parameter θ [15], [16]. Nematollahi and Shariati (2009) estimated the scale parameter of the Gamma distribution

under the Elf [17]. Azimi et al. (2012) studied the comparison of Bayesian estimation methods under different loss functions for a progressive censored Rayleigh distribution [18].

Suppose that X_i, $i = 1, \ldots, n$ is an iid random sample from the IWD, and the shape parameter α is known. The likelihood is given as

$$L(\lambda|x,\alpha) = \prod_{i=1}^{n} f\ (x;\ \alpha,\lambda) \tag{11}$$

$$\mathrm{L}= \prod_{i=1}^{n}(\alpha\lambda)\left(x_i^{-(\alpha+1)}\right) e^{-\lambda x_i^{-\alpha}} = \alpha^n\lambda^n e^{-\lambda T}\prod_{i=1}^{n}\left(x_i^{-(\alpha+1)}\right) \tag{12}$$

where $\mathrm{T}= \sum_{i=1}^{n}(x_i^{-\alpha})$.

In this study, to compute the Bayesian estimate of the IWD scale parameter λ, a Gamma prior and five different loss functions are used. Choosing a prior distribution for the unknown model parameter is essential in Bayesian theory. Being a natural conjugate prior, in this study, the Gamma distribution is considered a prior for the scale parameter λ in IWD with hyperparameters a and b. Then it has the following density function:

$$\pi(\lambda) = \frac{b^a\lambda^{a-1}}{\Gamma(a)}e^{-b\lambda}\ , \lambda > 0, a, b > 0. \tag{13}$$

For the λ, the posterior distribution is computed as follows:

$$\pi\ (\lambda\ |x) = \frac{L(\lambda|x,\alpha)\pi(\lambda)}{\int_0^\infty L(\lambda|x,\alpha)\pi(\lambda)d\lambda} \tag{14}$$

$$\pi(\lambda|x) = \frac{\lambda^{n+a-1}e^{-\lambda(T+b)}}{\int_0^\infty \alpha^{n+a-1}e^{-\alpha(T+b)}d\lambda} = \frac{(T+b)^{n+a}}{\Gamma(n+a)}\lambda^{n+a-1}e^{-\lambda(T+b)}. \tag{15}$$

This distribution is appropriate for the Gamma distribution with parameters $(n + a)$ and $(T + b)$.

Bayesian Estimator under the Self

In Bayesian estimation, a frequently used loss function is the symmetric Self. The Self is given as

$$L(\lambda,\hat{\lambda}) = \left(\lambda - \hat{\lambda}\right)^2. \tag{16}$$

The Self gives equal weight to both overestimation and underestimation. If the Self is used as the loss function in Bayesian inference, the estimator is the expected value of the posterior distribution. Then, for the λ, it is given as

$$\hat{\lambda}_{Self} = E(\lambda|x) \tag{17}$$

then the Bayes estimator

$$\hat{\lambda}_{Self} = \frac{(n+a)}{(T+b)} \tag{18}$$

is obtained.

Bayesian Estimator under the Elf

The Elf is a useful asymmetric loss function that was pointed out by [6]. Let $f(x, \lambda)$represent the pdf of the random variable X, and λ is the parameter. If $\hat{\lambda}$ is an estimator for λ, then general Elf is defined by

$$L(\lambda,\hat{\lambda}) = \left(\left(\frac{\hat{\lambda}}{\lambda}\right)^{c_1} - c_1 ln\frac{\hat{\lambda}}{\lambda} - 1\right) \tag{19}$$

[3]. In Elf, if $c_1 > 0$, then an overestimation error is more important than an underestimation error. If $c_1 < 0$, then an underestimation error is more important than an overestimation error. Under the general Elf, the Bayesian estimator of λ is given as

$$\hat{\lambda}_{Gelf}\ (X) = \left(E(\ \lambda^{-c_1}\ |X)\right)^{-c_1} \tag{20}$$

when $E_\lambda(.)$ exists and is finite. If $c_1 = -1$ then the Bayesian estimator under the Elf is the same as under the Self. If in the Elf, c_1 is taken as 1 then for the λ, the Bayesian estimator becomes

$$\hat{\lambda}_{Elf}\ (X) = \left(E\left(\frac{1}{\lambda}\middle|X\right)\right)^{-1}. \tag{21}$$

The expectation is computed as follows:

$$E\left(\frac{1}{\lambda}\middle|X = x\right) = \int_0^\infty \frac{1}{\lambda}\frac{(T+b)^{n+a}}{\Gamma(n+a)}\lambda^{n+a-1}e^{-\lambda(T+b)}d\lambda = \frac{T+b}{n+a-1} \tag{22}$$

Then the estimator for the λ under the Elf

$$\hat{\lambda}_{Elf}\ (X = x) = \left(\frac{T+b}{n+a-1}\right)^{-1} = \frac{n+a-1}{T+b} \tag{23}$$

is obtained.

Bayesian Estimator under the Linexlf

The Linexlf for the parameter λ is expressed as the following:

$$L\left(\lambda,\hat{\lambda}\right) = \left(e^{c\left(\hat{\lambda}-\lambda\right)} - 1 - c\left(\hat{\lambda} - \lambda\right)\right) \tag{24}$$

where $\hat{\lambda}$ is an estimate of λ and $c \neq 0$ [19]. This loss function is asymmetric. Many authors have discussed the Linex loss function [15],[20], [21], [22]. Under the Linexlf, the Bayesian estimator for λ is obtained following

$$\hat{\lambda}_{Linex} = \frac{-1}{c}\log\left(E\left[e^{-c\lambda}\ \middle|X\right]\right) \tag{25}$$

[23]. For IWD, the $\hat{\lambda}_{Linexlf}$ is

$$\begin{aligned} E\left(e^{-c\lambda}\middle|x\right) &= \int_0^\infty \frac{(T+b)^{n+a}}{\Gamma(n+a)}\ e^{-c\lambda}\lambda^{n+a-1}e^{-\lambda(T+b)}d\lambda \\ &= \int_0^\infty e^{-\lambda(c+T+b)}\frac{(T+b)^{n+a}}{\Gamma(n+a)}\lambda^{n+a-1}\ d\lambda \end{aligned} \tag{26}$$

$$E\left(e^{-c\lambda}|x\right)=\frac{(T+b)^{n+a}}{\Gamma(n+a)}\int_0^{\infty}e^{-t}\left(\frac{t}{c+T+b}\right)^{n+a-1}d\lambda \\ =\left(\frac{(T+b)^{n+a}}{\Gamma(n+a)}\right)\left(\frac{1}{(c+T+b)^{n+a-1}}\right)\int_0^{\infty}(e^{-t})(t^{n+a-1})d\lambda \quad (27)$$

$$\int_0^{\infty}(e^{-t})(t^{n+a-1})d\lambda=\Gamma(n+a) \\ E\left(e^{-c\lambda}|x\right)=\frac{(T+b)^{n+a}}{(c+T+b)^{n+a-1}} \quad (28)$$

$$\hat{\lambda}_{Linexlf}=-\frac{1}{c}\log\left(E\left(e^{-c\lambda}|x\right)\right) \\ \hat{\lambda}_{Linexlf}=-\frac{1}{c}\log\left(\frac{(T+b)^{n+a}}{(c+T+b)^{n+a-1}}\right) \quad (29)$$

computed.

Bayesian Estimator under the Plf

The Plf is introduced in [14] as follows

$$L\left(\lambda,\hat{\lambda}\right)=\frac{\left(\hat{\lambda}-\lambda\right)^2}{\hat{\lambda}}. \quad (30)$$

This function is asymmetric. By solving the equation that follows, the Bayesian estimator under the Plf is obtained as

$$\hat{\lambda}_{plf}^2=E(\lambda^2|X). \quad (31)$$

Then

$$E(\lambda^2|x)=\int_0^{\infty}\lambda^2\frac{(T+b)^{n+a}}{\Gamma(n+a)}\lambda^{n+a-1}e^{-\lambda(T+b)}d\lambda \\ =\int_0^{\infty}\lambda^{n+a+1}\frac{(T+b)^{n+a}}{\Gamma(n+a)}e^{-\lambda(T+b)}d\lambda \quad (32)$$

$$E(\lambda^2|x)=\frac{\Gamma(n+a+2)}{\Gamma(n+a)(T+b)^2} \quad (33)$$

is computed. Finally, under the Plf, the Bayesian estimator of λ can be obtained as

$$\hat{\lambda}_{plf}=\frac{\sqrt{(n+a+1)(n+a)}}{(T+b)}. \quad (34)$$

Bayesian Estimator under the Nlinexlf

The Nlinex loss function is given as

$$L_{Nlinexlf}\left(\hat{\lambda},\lambda\right)=k\left[e^{cD}-cD^2-cD-1\right],k>0,c>0. \quad (35)$$

In Equation (35) D= $\hat{\lambda}$ -λ represents estimation error. For lack of generality, one can assume k = 1. The estimator under Nlinexlf $\hat{\lambda}_{Nlinexlf}$ is obtained as follows:

$$\hat{\lambda}_{Nlinexlf}=-\frac{\left[\ln E\left(e^{-c\lambda}\right)-2E\left(\lambda\right)\right]}{(c+2)} \quad (36)$$

[24]. Then at first,

$$E\left(e^{-c\lambda}\right)=\int_0^{\infty}e^{-c\lambda}\,\pi(\lambda|x)d\lambda \quad (37)$$

is computed. From (15), it is known that the posterior distribution is the Gamma, then

$$E((\lambda|x))=\frac{(n+a)}{(T+b)} \quad (38)$$

and

$$E\left(e^{-c\lambda}|x\right)=\frac{(T+b)^{n+a}}{(c+T+b)^{n+a-1}}. \quad (39)$$

Then, the Bayes estimator under Nlinexlf is as follows:

$$\hat{\lambda}_{Nlinexlf}=-\frac{\left[\ln\left(\frac{(T+b)^{n+a}}{(c+T+b)^{n+a-1}}\right)-2(n+a)(T+b)\right.}{(c+2)} \quad (40)$$

SIMULATION STUDY

In a simulation study, different random samples from the IWD with the sample sizes n=10,30,50,70,90,110 are generated in Matlab. A simulation study is carried out 1000 times for the parameter values α = 2 and λ=2,3,4 where the prior hyperparameters are chosen as a = 2, b = 2. This is iterated 1000 times, and the λ is estimated using each of the methods given in the previous sections. The mean squared error (MSE) is used as a criterion to compare the efficiency of the methods. The MSE is calculated as follows:

$$MSE=\frac{\sum_{i=1}^{1000}\left(\boldsymbol{\lambda_i}-\hat{\boldsymbol{\lambda}}_i\right)^2}{\mathbf{1000}}$$

The results are given in Table 1.

RESULTS AND DISCUSSION

The main observations of the results from Table 1 are summarized below:

1) From Table 1, it can be said that the Bayes estimate using the Gamma prior under the Plf provides the smallest Mse values in most cases as compared to the other loss functions and the classical mle. Especially for small sample sizes, Bayes estimates give better results than classical mle, except for the linexlf and the Nonlinexlf.
2) The Linexlf and the Nlinexlf give the worst conclusions in all cases.
3) The Bayesian estimates under linexlf and nonlinexlf are sensitive to the values of the corresponding shape parameter c.

Table 1. Estimation results for the scale parameter λ

			λ=2		λ=3		λ=4	
		α = 2	$\hat{\lambda}$	MSE	$\hat{\lambda}$	MSE	$\hat{\lambda}$	MSE
Mle		***n* = 10**	2.20644	0.62068	3.32654	1.53461	4.45114	2.80048
Self			1.57090	0.30698	2.04200	1.07607	2.40437	2.71816
Elf			1.42129	0.43547	1.84753	1.45778	2.17539	3.47012
Plf			1.64401	0.26129	2.13703	0.91810	2.51626	2.39000
Linexlf	***c* = -10**		1.13154	1.58873	0.30863	7.43790	0.04522	16.44679
	***c* = 5**		0.66794	1.82347	0.95072	4.24738	1.14415	8.19917
	***c* = 10**		0.66837	1.79457	0.85270	4.63050	0.97563	9.16387
Nlinexlf	***c* = -10**		1.02170	2.43911	0.12471	10.18573	0.65762	21.91668
	***c* = 5**		0.92593	1.22029	1.26252	3.09159	1.50422	6.30024
	***c* = 10**		0.81880	1.427641	1.050918	3.832341	1.213758	7.794792
Mle		***n* = 30**	2.07868	0.15796	3.09597	0.35062	4.13027	0.61825
Self			1.84925	0.11318	2.59618	0.32256	3.27339	0.75132
Elf			1.78862	0.12930	2.51106	0.38826	3.16607	0.90439
Plf			1.87932	0.10799	2.63840	0.29548	3.32662	0.68412
Linexlf	***c* = -10**		3.16380	2.46902	6.31791	18.95462	7.49366	23.26038
	***c* = 5**		0.99492	1.07413	1.59093	2.07454	8.54872	38.44228
	***c* = 10**		1.11109	0.82798	1.56196	2.11680	1.91857	4.38625
Nlinexlf	***c* = -10**		3.49244	3.79986	7.24834	30.01407	2.07873	3.79572
	***c* = 5**		1.23902	0.65013	1.87814	1.36562	2.42006	2.63001
	***c* = 10**		1.23412	0.63158	1.73433	1.66474	2.14437	3.51741
Mle		***n* = 50**	2.04451	0.08798	3.06739	0.20508	4.08864	0.36833
Self			1.90612	0.07218	2.75382	0.18680	3.53995	0.40785
Elf			1.86838	0.078214	2.699294	0.211681	3.469854	0.469568
Plf			1.92490	0.07026	2.78095	0.17668	3.57482	0.38086
Linexlf	***c* = -10**		2.69097	0.63340	4.26068	2.30543	6.55848	11.14111
	***c* = 5**		1.05226	0.95238	1.79978	1.52877	2.43634	2.56340
	***c* = 10**		1.25340	0.59388	1.85760	1.36063	2.35641	2.77175
Nlinexlf	***c* = -10**		2.88718	0.97274	4.63739	3.62416	7.31312	17.64494
	***c* = 5**		1.29622	0.55202	2.07236	0.95889	2.75166	1.69695
	***c* = 10**		1.36218	0.44724	2.00697	1.05143	2.55366	2.17880
Mle		***n* = 70**	2.02797	0.05776	3.04745	0.14047	4.04641	0.24060
Self			1.92921	0.05087	2.82030	0.13190	3.64807	0.27840
Elf			1.90184	0.05420	2.78030	0.14507	3.59633	0.31314
Plf			1.94284	0.04978	0.12654	0.12654	3.67385	0.26311
Linexlf	***c* = -10**		2.58868	0.42674	3.89315	1.06752	5.41010	2.67484
	***c* = 5**		1.06035	0.92584	1.88711	1.31728	2.60159	2.06323
	***c* = 10**		1.31867	0.49517	2.01267	1.02877	2.59702	2.03842
Nlinexlf	***c* = -10**		2.75354	0.65812	4.16136	1.67429	5.85061	4.30444
	***c* = 5**		1.30860	0.52177	2.15374	0.80062	2.90059	1.32891
	***c* = 10**		1.42043	0.36915	2.14727	0.78773	2.77220	1.58935
Mle		***n* = 90**	2.02507	0.04677	3.02781	0.10255	4.04501	0.18797
Self			1.94772	0.04171	2.85082	0.10119	3.72882	0.20581
Elf			1.92620	0.04357	2.81932	0.10984	3.68762	0.22695
Plf			1.95846	0.04114	2.86653	0.09762	3.74936	0.19656
Linexlf	***c* = -10**		2.55689	0.36793	3.74179	0.71043	5.09393	1.57450
	***c* = 5**		1.05983	0.92237	1.92404	1.22483	2.71174	1.75989
	***c* = 10**		1.35875	0.44017	2.10215	0.85470	2.76497	1.59458
Nlinexlf	***c* = -10**		2.70919	0.56603	3.96453	1.11530	5.43521	2.51891
	***c* = 5**		1.31351	0.50986	2.18883	0.72840	3.0023	1.10430
	***c* = 10**		1.45691	0.32548	2.22693	0.65075	2.92561	21.23268
Mle		***n* = 110**	2.01365	0.037582	3.02873	0.08849	4.04084	0.15622
Self			1.95074	0.03501	2.88234	0.08498	3.77911	0.16572
Elf			1.93309	0.03647	0.09052	0.09052	3.74491	0.17989
Plf			1.95955	0.03452	2.89535	0.08273	3.79617	0.15953
Linexlf	***c* = -10**		2.53318	0.32799	3.68057	0.58632	4.93338	1.13273
	***c* = 5**		1.04366	0.94787	1.95337	1.15882	2.77927	1.58430
	***c* = 10**		1.37505	0.41641	2.17095	0.73494	2.88021	1.32190
Nlinexlf	***c* = -10**		2.67878	0.50751	3.88013	0.91301	5.22194	1.79988
	***c* = 5**		1.30283	0.51913	2.21879	0.67584	3.06494	0.97472
	***c* = 10**		1.47100	0.30676	2.28951	0.55600	3.03003	1.01606

4) The results show that estimators of the different methods are closer to each other as the sample size increases, except for Linexlf and Nlinexlf.
5) Moreover, it is seen that when the sample size increases, the MSE decreases significantly.

CONCLUSION

This study deals with the investigation of the mle and the Bayesian estimator for the scale parameter of IWD when the shape parameter is known. For Bayesian inference, five different loss functions are used respectively: Self, Plf, Elf, Linexlf, and Nlinexlf. In the simulation study, the Bayesian estimates and the mle are computed. To compare the results of the estimations, the MSE values are calculated. The results show that the Bayesian method of estimation for the Gamma prior under the Self, Plf, and Elf gives better results than the mle method. Also, the Bayesian estimators that are obtained under the Plf have the smallest MSE as compared with the Bayesian estimators that are obtained under the other loss functions. Also, Bayesian estimators, which are obtained under Linexlf and Nlinexlf, are worse than the other loss functions and the mle.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Langlands AO, Pocock SJ, Kerr GR, Gore SM. Long-term survival of patients with breast cancer: A study of the curability of the disease. Br Med J 1979;2:1247–1251. [CrossRef]
[2] Bennett S. Log-logistic regression models for survival data. J R Stat Soc 1983;32:165–171. [CrossRef]
[3] Kumar SS, Umesh S, Kumar D. Bayesian estimation of parameters of inverse Weibull distribution. J Appl Stat 2013;40:1597–1607. [CrossRef]
[4] Kundu D, Howlader H. Bayesian ınference and prediction of the ınverse weibull distribution for type-II censored data. Comput Stat Data Anal 2010;54:1547–1558. [CrossRef]
[5] Keller AZ, Kamath ARR. Alternative reliability models for mechanical systems. In proceeding of the 3rd International Conference on Reliability and Maintainability; 1982. pp. 411–415.
[6] Calabria R, Pulcini G. Bayes 2-sample prediction for the inverse Weibull distribution. Commun Stat Theory Methods 1994;23:1811–1824. [CrossRef]
[7] Helu A, Samawi H. The inverse weibull distribution as a failure model under various loss functions and based on progressive first-failure censored data. Quality Technol Quant Manage 2015;12:517–535. [CrossRef]
[8] Bi Q, Gui W. Bayesian and classical estimation of stress-strength reliability for inverse weibull lifetime models. Algorithms 2017;10:71. [CrossRef]
[9] Nassar M, Abo-Kasem EE. Estimation of the inverse Weibull parameters under adaptive type-II progressive hybrid censoring scheme. J Comput Appl Math 2017;315:228–239. [CrossRef]
[10] Singh S, Tripathi YM. Estimating the parameters of an inverse Weibull distribution under progressive type-I interval censoring. Stat Papers 2018;59:21–56. [CrossRef]
[11] Jana N, Bera S. Estimation of parameters of inverse Weibull distribution and application to multi-component stress-strength model. J Appl Stat 2022;49:169-194. [CrossRef]
[12] Basu AP, Ebrahimi N. Bayesian approach to life testing and reliability estimation using asymmetric loss function. J Stat Plan Inference 1992;29:21–31. [CrossRef]
[13] Berger JO. Statistical decision theory, foundation, concepts and method. New York, NY, USA: Springer; 1985. [CrossRef]
[14] Norstrom JG. The use of precautionary loss functions in risk analysis. IEEE Trans Reliab vol. 1996;45:400–403. [CrossRef]
[15] Parsian A, Kirmani SNU. Estimation Under LINEX Loss Function. In: Ullah A, Wan ATK, Chaturvedi A, Dekker M, editors. Handbook of applied econometrics and statistical inference. Boca Raton, FL, USA: CRC Press; 2020. pp. 53–76. [CrossRef]
[16] Misra N, Meulen EC. On estimating the mean of the selected normal population under the LINEX loss function. Metrika 2003;58:173–183. [CrossRef]
[17] Nematollahi N, Motamed-Shariati F. Estimation of the scale parameter of the selected gamma population under the entropy loss function, communications in statistics. Theory Methods 2009;38;208–221. [CrossRef]
[18] Azimi R, Yaghmaei F, Azimi D. Comparison of bayesian estimation methods for rayleigh progressive censored data under the different asymmetric loss function. Int J Appl Math Res 2012;1:452–461. [CrossRef]

[19] Ahmad K, Ahmad SP, Ahmed A. Classical and bayesian approach in estimation of scale parameter of ınverse weibull distribution. Math Theory Model 2015;5.

[20] Calabria R, Pulcini G. Point estimation under asymmetric loss functions for left-truncated exponential samples. Commun Stat Aeory Methods 1996;25:585–600. [CrossRef]

[21] Gencer G, Saraçoğlu B. Comparison of approximate Bayes estimators under different loss functions for parameters of odd Weibull distribution. J Selçuk Univ Nat Appl Sci 2016;5:18–32.

[22] Khatun N, Matin MA. A study on LINEX loss function with different estimating methods. Open J Stat 2020;10:52–63. [CrossRef]

[23] Sultan KS. Bayesian estimates based on record values from the inverse Weibull lifetime model. Qual Technol Quant Manage 2008;5:363–374. [CrossRef]

[24] Islam SAFM, Roy MK, Ali MM. A non-linear exponential (NLINEX) loss function in bayesian analysis. J Korean Data Inform Sci Soc 2004;15:899–910.

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1116–1132, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00092

Research Article

RAMD analysis of mixed standby serial manufacturing system

Abdulkarim MUAZU IGGI[1,*], Ibrahim YUSUF[2]

[1]*Department of Mathematics, Federal College of Education, Kano, 562261, Nigeria*
[2]*Department of Mathematical Sciences, Bayero University, Kano, 700006, Nigeria*

ARTICLE INFO

Article history
Received: 23 December 2022
Revised: 09 February 2023
Accepted: 26 February 2023

Keywords:
Availability; Exponential, Lindley; Exponentiated Weibull; Reliability; Textile

ABSTRACT

This study aimed to increase textile manufacturing system dependability, reliability, maintainability, availability, and metrics like MTBF and MTTF by boosting RAMD. The textile system under investigation is a serial system consisting of five subsystems, which are; subsystem A is weaving section, subsystem B is the dry clean section, subsystem C is the cross cut section, subsystem D is the side seam section and subsystem E is the cleaning section. Each of the subsystem consist of main unit, warm standby unit and cold standby unit. For design and prediction, the Markovian birth-death method is employed to assemble the system governing the differential difference equation from the state-to-state transition diagram. The rates of repair and failure of each subsystem are exponentially distributed and statistically independent. For several subsystems of the system, the findings for RAMD, all of which are crucial to system performance, have been acquired and shown in figures and tables. Furthermore, the results of this study reveal that the highest system performance and dependability may be achieved when the overall system failure rate is low. The findings of this research are thought to be valuable for analyzing performance and determining the best system design and feasible maintenance strategies that may be used in the future to improve system performance, strength, effectiveness, production output as well as revenue mobilization.

Cite this article as: Muazu Iggi A, Yusuf I. RAMD analysis of mixed standby serial manufacturing system. Sigma J Eng Nat Sci 2024;42(4):1116–1132.

INTRODUCTION

RAMD is a logistical technique for assessing the strength, effectiveness, and performance of equipment at various levels. It ensures system safety and operation problems and identifies which of the system's units, components, or subsystems require adequate maintenance. RAMD (reliability, availability, maintainability, and dependability) management is critical to a company's success. These four measures of system strength, effectiveness, and performance can be used to forecast system speed, product quality, and volume production output.

Researchers have used a variety of approaches to assess reliability measures in the literature. RAMD analysis was used by [1] to generate a mathematical model for assessing the effectiveness of serial mechanisms in a sugar plant's refining system. [2] proposed a reliability and availability assessment of the skim industry powder business. The

*Corresponding author.
*E-mail address: amiggi1977@gmail.com, iyusuf.mth@buk.edu.ng

This paper was recommended for publication in revised form by Editor in Chief Ahmet Selim Dalkılıç

Published by Yıldız Technical University Press, İstanbul, Turkey

Markovian process is used to evaluate measures such as maintainability, reliability, dependability, and availability in determining its capability and reliability. [3] concentrate on increasing the profit of engineering systems with serial subsystems by improving system performance indicators like availability and reliability. [4] investigate the reciprocating unit's system availability, maintainability, and dependability in the oil and gas industries in order to improve the unit's operating performance. [5] uses particle swamp optimization and fuzzy techniques to assess industrial reliability, maintainability, and availability. [6] investigated the efficiency of the forming industry by assessing system maintainability, dependability, and availability. [7] developed Markov models for RAM performance estimation of circulation system of water. [8] discuss the RAM evaluation of Load Haul Dumpers.

Available studies either neglects or overlooks the importance of warm and cold standby in strengthening system reliability, availability, mean time to failure, and MTBF. Most previous studies focused solely on system availability and effectiveness evaluation, paying little attention to the influence of warm and cold standby units on reliability, availability, mean time to failure, and generated revenue. More advanced designs with mixed standby units should indeed be established to reduce the likelihood of a complete breakdown, expenditures, overall reliability, availability, mean time to failure, and revenue generated (profit).

The aforementioned literature review presented in Table 1 above reveals that the RAMD evaluation of some industrial and manufacturing system having mixture of warm and cold standby units when failure and repair rates as Lindley and Exponentiated Weibull distributed has not been explored so far. Motivated by the aforementioned studies in Table 1 above, the objective of this work is to perform RAMD analysis of textile system with mixed standby unit when failure rates follows Lindley and Exponentiated Weibull distribution. As a result, this study considers a textile manufacturing system that consists of five distinct subsystems equipped as a series-parallel system, each consisting of a combination of primary units, warm standby units, and cold standby units. The system's effectiveness is investigated via first order differential difference equations. Availability as one of the performance measures of system strength and effectiveness have been computed for each configuration. The present work will perform RAMD analysis of textile system with mixed standby unit when failure rates follows Lindley and Exponentiated Weibull distribution.

The following are the paper's contributions:

- ✓ To formulated novel models of RAMD analysis of textile manufacturing system considering models; main,

Table 1. Some related research on availability, maintainability, reliability and dependability of some complex systems

Reference	System	Standby used	Exponential distribution	Lindley distribution	Exponentiated Weibull distribution	Reliability	Availability	Maintainability	Dependability
[9]	Sewage treatment plant	N/A	yes	no	no	yes	yes	yes	yes
[10]	Series-parallel	Cold	yes	no	no	yes	yes	yes	yes
[11]	Cement	Cold	yes	no	no	yes	yes	yes	no
[12]	Steam turbine power plant	N/A	yes	no	no	yes	yes	yes	No
[13]	Water treatment plant	N/A	yes	no	no	yes	yes	yes	Yes
[14]	Tube-well	N/A	yes	no	no	yes	yes	yes	Yes
[15]	Sugar Plant	N/A	yes	no	no	yes	yes	yes	yes
[16]	microprocessor	N/A	yes	no	no	yes	yes	yes	yes
[17]	sugar manufacturing plant	N/A	no	no	no	yes	no	yes	no
[18]	hot standby database systems	N/A	yes	no	no	yes	yes	yes	no
[19]	Automotive manufacturing	N/A	yes	no	no	yes	yes	yes	yes
[20]	power generating unit of sewage treatment plant	N/A	yes	no	no	yes	no	yes	no
Proposed study	Textile confection plant	mixed	yes	yes	yes	yes	yes	yes	yes

warm and cold standby units. Warm standby unit reduce energy use and recovery period because a standby unit is partly energized and subjected to maximum stress while the primary unit is up and running and completely powered and functional after the primary unit stops working.

- ✓ Developing the explicit expressions for the availability, reliability, mean time between failure, maintainability, mean time to failure and dependability for each subsystem.
- ✓ To see the performance of the system through ramd models under exponential, Lindley and exponentiated Weibull distributions.

The following is how this paper is structured. The framework for this study is described in Section 2. Section 3 discusses the methods and materials used. Section 4 is dedicated to the modelling approach. Section 5 presents the simulation studies and consequences discussion, and Section 6 concludes the paper.

DESCRIPTION OF THE SYSTEM AND NOTATIONS

Description of The System

The textile system under investigation is a serial system consisting of five subsystems, which are; weaving section, dry clean section, cross cut section, side seam section and cleaning section. Each of the subsystem consist of main unit, warm standby unit and cold standby unit as shown in Table 2. Warm standby unit are introduced in enhancing the performance of the system. Warm standby units have the capacity to reduce energy use and recovery period because a standby unit is partly energized and subjected to maximum stress while the primary unit is up and running and completely powered and functional after the primary unit stops working. When one of the primary units fails, the warm standby resumes to work with minimal service interruption. Sequel to this, system with warm or mixed standby units have gained the attention of different researchers. To cite few, [21] analysed the cost benefit of warm standby retrial systems with imperfect coverage. Analysis of reliability and availability of a redundant k-out-of-n warm standby system in the presence of common cause failure has been presented in [22]. Evaluation of reliability and performance of power system having warm standby unit is given in [23]. [24] focus on profit optimization of a warm standby non identical system in normal and abnormal environment. [25] analysed reliability of warm standby serial system with switching mechanism and uncertain lifetimes. [26] presented reliability simulation of warm standby two component system having switching and back switching failures. [27] focus on economic analysis of warm standby system attended by single server. [28] analysed the profit of warm standby system attended by single server with priority. [29] analysed the performance of warm standby machine repair problem with servers' vacation, impatient and controlling F-policy.

The system can be in perfect or initial state when new. At the failure of one of the primary unit, a warm standby unit will shift to take over the failed unit while the cold standby unit will take the position of warm standby unit. This failure is called the partial failure. When all the primary and warm standby failed, the system is down. This called complete failure.

Subsystem A (Weaving)

Any machine that weaves yarn into fabric is referred to as a weaving machine. They are used to render upholstery fabric, silk, and ornate carpets. They come in shuttle, circular, and narrow fabric options.

Subsystem B (Dry Clean): A dry cleaning machine is any sanitizing device that uses a solvent other than water to tidy clothing and textiles. Although liquid is still used in dry cleaning, clothes are submerged in a water-free liquid solvent and other detergent, which is the most commonly used solvent.

Subsystem C (Cross Cut): A cross cutter machine is an equipment that cuts both hard and soft wood.

Subsystem D (Side Seam): A seam is a method of joining a number of pieces of garment, typically with thread to form stitches. Seams can be hand-stitched or machine-stitched. A seam is a line that connects pieces of fabric and other materials in a garment.

Subsystem E (Cleaning): Cleaning is the mechanical removal of loosely bound fibers, such as brushing, sueding, or grinding. Cleaning processes that are solvent-free are workable alternatives to the traditional solvent-based regular cleaning. They reduce waste generation and remove potential risks caused by the use and application of toxic, ozone-depleting, and frequently flammable solvents. Sanding, grinding, polishing, brushing / sueding, cropping, and shearing are examples of cleaning operations.

Table 2. System Configuration

Machine/Subsystem	Primary Unit	Warm Standby Unit	Cold Standby Unit	Total
Weaving (A)	4	1	1	6
Dry Clean (B)	5	2	1	8
Cross Cut (C)	2	2	2	6
Side Seam (D)	3	2	1	6
Cleaning (E)	4	2	1	7

Notations

q: time variable

λ_1 / λ_2 / λ_3 / λ_4 / λ_5: main unit failure rate in weaving subsystem, dry clean subsystem, cross cot subsystem, side seam subsystem and cleaning subsystem.

α_1 / α_2 / α_3 / α_4 / α_5: warm standby unit failure rate in weaving subsystem, dry clean subsystem, cross cot subsystem, side seam subsystem and cleaning subsystem.

μ_1 / μ_2 / μ_3 / μ_4 / μ_5: warm standby unit failure rate in weaving subsystem, dry clean subsystem, cross cot subsystem, side seam subsystem and cleaning subsystem.

$\vartheta_k(q)$: probability that the system is in state S_k at time q.

MATERIALS AND METHODS

Reliability Models

The chance that a system/machine will be up and running throughout a period of time q is defined as reliability. Thus, reliability $R(q) = P_r\{Q > q\}$, where Q is the time when the system is down and not running with $R(q) \geq 0$, $R(q) = 1$. (For a full description, see Ebeling (2000)). Thus,

$$R(q) = \int_q^{\infty} f(q_0)\, dq_0 \tag{1}$$

and

$$R(q) = e^{-mq} \tag{2}$$

$$R(q) = \left(\frac{1+m+mq}{1+m}\right) e^{-mq} \tag{3}$$

$$R(t) = 1 - \left(1 - e^{-(\lambda t)^{\gamma}}\right)^{\alpha} \tag{4}$$

for exponentially, Lindley and exponentiated Weibull distributed rate of failure respectively.

$$A(q) = limA(Q) = \frac{MTBF}{MTBF+MTTR}. \tag{5}$$

Maintainability

$$M(q) = P(Q \leq q) = 1 - e^{\left(\frac{-q}{MTTR}\right)} = 1 - e^{-\mu q}. \tag{6}$$

where μ is the constant system's repair rate.

Dependability

Dependability is a metric given by

$$D_{min} = 1 - \left(\frac{1}{h-1}\right)\left(e^{-log(h)/h-1} - e^{-hlog(h)/h-1}\right). \tag{7}$$

where

$$h = \frac{\mu}{\theta} = \frac{MTBF}{MTTR}. \tag{8}$$

Mean Time Between Failure

The average time between the failures is known as MTBF. It's usually expressed in hours. As the MTBF increases, so does the system's reliability. The MTBF is given by

$$MTBF = \int_0^{\infty} R(q)\, dq = \int_0^{\infty} e^{-mq}\, dq = \frac{1}{m}. \tag{9}$$

Mean Time to Repair

The reciprocal of the system repair rate is specified as MTTR given by

$$MTTR = \mu^{-1} \tag{10}$$

where μ is the system's repair rate.

FORMULATION OF MATHEMATICAL MODELS FOR RAMD

In this section, Chapman Kolmogorov differential equations for each subsystem have been constructed using the Markov birth-death process for mathematical modeling of textile manufacturing system. Table 3 displays various subsystem failure and repair rates. Table 4 below gives the description of the state of each subsystem.

Table 4. Transition rate table for Subsystem A

	S_0	S_1	S_2	S_3
S_0	0	$4\lambda_1 + \alpha_1$	0	0
S_1	μ_1	0	$4\lambda_1 + \alpha_1$	0
S_2	0	$2\mu_1$	0	$4\lambda_1$
S_3	0	0	$3\mu_1$	0

Table 3. Failure and repair rate

Machine/Subsystem	Failure rate (λ) Operational Units	Failure rate (α) Warm standby Units	Repair rate (μ)
Weaving (A)	0.015	0.015	0.35
Dry Clean (B)	0.025	0.016	0.20
Cross Cut (C)	0.010	0.014	0.15
Side Seam (D)	0.035	0.017	0.40
Cleaning (E)	0.050	0.013	0.55

RAMD Analysis for Subsystem A (Weaving unit)

This section consists of four primary operation unit (main unit), one warm standby unit and one cold standby unit. When one of the primary units failed, the warm standby unit switch to operation as primary unit and the cold standby unit switch to the position of warm standby unit. Through Table 4 below, the Chapman-Kolmogrov differential difference equations (11)-(14) are derived using Markovian birth-death process.

Where S_0 is the perfect state, S_1, S_2 are partial failure states and S_3 is the complete failure state.

$$\frac{d}{dq}\vartheta_0(q) = -(4\lambda_1 + \alpha_1)\vartheta_0(q) + \mu_1\vartheta_1(q) \tag{11}$$

$$\frac{d}{dq}\vartheta_1(q) = -(4\lambda_1 + \alpha_1 + \mu_1)\vartheta_1(q) + (4\lambda_1 + \alpha_1)\vartheta_0(q) + 2\mu_1\vartheta_2(q) \tag{12}$$

$$\frac{d}{dq}\vartheta_2(q) = -(4\lambda_1 + 2\mu_1)\vartheta_1(q) + (4\lambda_1 + \alpha_1)\vartheta_1(q) + 3\mu_1\vartheta_3(q) \tag{13}$$

$$\frac{d}{dq}\vartheta_3(q) = -3\vartheta_3(q) + 4\lambda_1\vartheta_2(q) \tag{14}$$

The normalizing condition for this problem is

$$\vartheta_0(q) + \vartheta_1(q) + \vartheta_2(q) + \vartheta_3(q) = 1 \tag{15}$$

Availability of subsystem A is

$$A_{S1} = \vartheta_0(q) + \vartheta_1(q) + \vartheta_2(q) \tag{16}$$

Setting (11) to (14) to zero as $q \to \infty$ in steady state, availability of subsystem A in (16) is now

$$A_{S1}(\infty) = \frac{1 + m_1 + \frac{m_1^2}{2}}{1 + m_1 + \frac{m_1^2}{2} + \frac{2\lambda_1 m_1^2}{3\mu_1}} \tag{17}$$

Where $m_1 = \left(\frac{4\lambda_1 + \alpha_1}{\mu_1}\right)$

The Corresponding reliability, maintainability, dependability and MTBF, MTTR for main and warm standby unit of subsystem A are

$$R_{S1}(q) = \exp^{-\lambda_1 q} \tag{18}$$

$$R_{S1}(q) = \exp^{-\alpha_1 q} \tag{19}$$

$$M_{S1} = 1 - \exp^{-\mu_1 q} \tag{20}$$

Mean time between failure (MTBF= $\lambda_1^{-1} = 66.6667\,h$ for main unit

Mean time between failure (MTBF)= $\alpha_1^{-1} = 66.6667\,h$ for warm standby unit

Mean time to repair (MTTR)= $\mu_1^{-1} = 2.8571\,h$

Dependability ratio d= $\frac{\mu_1}{\lambda_1} = 23.3345$ for main unit

Dependability ratio d= $\frac{\mu_1}{\alpha_1} = 23.3345$ for warm standby unit

$$D_{\min}(s_1) = 1 - \left(\frac{1}{d-1}\right)\left(\exp^{-\frac{\ln d}{d-1}} - \exp^{\frac{d\ln d}{d-1}}\right) = 0.9595$$

for main and warm standby unit

RAMD Analysis for Subsystem B (Dry Clean section)

This section consist of five primary unit, two warm standby and one cold standby unit. Similar to the method described in section 4.1 above, from Table 5 the differential difference equations in (21)-(25) are derived using Markovian birth-death process.

Where S_0 is the perfect state, S_1, S_2, S_3 are partial failure states and S_4 is the complete failure state

Table 5. Transition rate table for Subsystem B

	S_0	S_1	S_2	S_3	S_4
S_0	0	$5\lambda_2 + 2\alpha_2$	0	0	0
S_1	μ_2	0	$5\lambda_2 + 2\alpha_2$	0	0
S_2	0	$2\mu_2$	0	$5\lambda_2 + \alpha_2$	0
S_3	0	0	$3\mu_2$	0	$5\lambda_2$
S_4	0	0	0	$4\mu_2$	0

$$\frac{d}{dq}\vartheta_0(q) = -(5\lambda_2 + \alpha_2)\vartheta_0(q) + \mu_2\vartheta_1(q) \tag{21}$$

$$\frac{d}{dq}\vartheta_1(q) = -(5\lambda_2 + 2\alpha_2 + \mu_2)\vartheta_1(q) + (5\lambda_2 + 2\alpha_2)\vartheta_0(q) + 2\mu_2\vartheta_2(q) \tag{22}$$

$$\frac{d}{dq}\vartheta_2(q) = -(5\lambda_2 + \alpha_2 + 2\mu_2)\vartheta_2(q) + (5\lambda_2 + 2\alpha_2)\vartheta_1(q) + 3\mu_2\vartheta_3(q) \tag{23}$$

$$\frac{d}{dq}\vartheta_3(q) = -(5\lambda_2 + 3\mu_2)\vartheta_3(q) + (5\lambda_2 + \alpha_2)\vartheta_2(q) + 4\mu_2\vartheta_4(q) \tag{24}$$

$$\frac{d}{dq}\vartheta_4(q) = -4\mu_2\vartheta_4(q) + 5\lambda_2\vartheta_3(q) \tag{25}$$

Table 6. Transition rate table for Subsystem C

	S_0	S_1	S_2	S_3	S_4	S_5
S_0	0	$2\lambda_3 + 2\alpha_3$	0	0	0	0
S_1	μ_3	0	$2\lambda_3 + 2\alpha_3$	0	0	0
S_2	0	$2\mu_3$	0	$2\lambda_3 + 2\alpha_3$	0	0
S_3	0	0	$3\mu_3$	0	$2\lambda_3 + \alpha_3$	0
S_4	0	0	0	$4\mu_3$	0	$2\lambda_3$
S_5	0	0	0	0	$5\mu_3$	0

The normalizing condition for this problem is

$$\vartheta_0(q) + \vartheta_1(q) + \vartheta_2(q) + \vartheta_3(q) + \vartheta_4(q) = 1 \tag{26}$$

Availability of subsystem B is

$$A_{S2} = \vartheta_0(q) + \vartheta_1(q) + \vartheta_2(q) + \vartheta_3(q) \tag{27}$$

Setting (21) to (25) to zero as $q \to \infty$ in steady state, availability of subsystem B in (27) is now

$$A_{S2} = \frac{1 + m_2 + \frac{m_2^2}{2} + \frac{(5\lambda_2 + \alpha_2)m_2^2}{6\mu_2}}{1 + m_2 + \frac{m_2^2}{2} + \frac{(5\lambda_2 + \alpha_2)m_2^2}{6\mu_2} + \frac{5\lambda_2(5\lambda_2 + \alpha_2)m_2^2}{24\mu_2^2}} \tag{28}$$

Where $m_2 = \left(\frac{5\lambda_2 + 2\alpha_2}{\mu_2}\right)$

The Corresponding reliability, maintainability, dependability and MTBF, MTTR for main and warm standby unit of subsystem B are

$$R_{S2}(q) = \exp^{-\lambda_2 q} \tag{29}$$

$$R_{S2}(q) = \exp^{-\alpha_2 q} \tag{30}$$

$$M_{S2} = 1 - \exp^{-\mu_2 q} \tag{31}$$

Mean time between failure (MTBF)= $\lambda_2^{-1} = 40\,h$ for main unit

Mean time between failure (MTBF)= $\alpha_2^{-1} = 62.5\,h$ for warm standby unit

Mean time to repair (MTTR)= $\mu_2^{-1} = 5h$

Dependability ratio d= $\frac{\mu_2}{\lambda_2} = 8$ for main unit

$$D_{\min}(s_2) = 1 - \left(\frac{1}{d-1}\right)\left(\exp^{-\frac{\ln d}{d-1}} - \exp^{\frac{d\ln d}{d-1}}\right) = 0.8877$$

for main unit

Dependability ratio d= $\frac{\mu_2}{\alpha_2} = 12.5$ for warm standby unit

$$D_{\min}(s_2) = 1 - \left(\frac{1}{d-1}\right)\left(\exp^{-\frac{\ln d}{d-1}} - \exp^{\frac{d\ln d}{d-1}}\right) = 0.9357$$

for main and warm standby unit

RAMD Analysis for Subsystem C (Cross Cut Unit)

The cross-cut section consists of two primary operation unit, two warm standby unit and two cold standby unit. Using the method described in section 4.1 above, the Chapman-Kolmogrov differential difference equations (32)-(37) are derived using Markovian birth-death process from Table 6 below:

Where S_0 is the perfect state, S_1, S_2, S_3, S_4 are partial failure states and S_5 is the complete failure state

$$\frac{d}{dq}\vartheta_0(q) = -(2\lambda_3 + 2\alpha_3)\vartheta_0(q) + \mu_3\vartheta_1(q) \tag{32}$$

$$\frac{d}{dq}\vartheta_1(q) = -(2\lambda_3 + 2\alpha_3 + \mu_3)\vartheta_1(q) + (2\lambda_3 + 2\alpha_3)\vartheta_0(q) + 2\mu_3\vartheta_2(q) \tag{33}$$

$$\frac{d}{dq}\vartheta_2(q) = -(2\lambda_3 + 2\alpha_3 + 2\mu_3)\vartheta_2(q) + (2\lambda_3 + 2\alpha_3)\vartheta_1(q) + 3\mu_3\vartheta_3(q) \tag{34}$$

$$\frac{d}{dq}\vartheta_3(q) = -(2\lambda_3 + \alpha_3 + 3\mu_3)\vartheta_3(q) + (2\lambda_3 + 2\alpha_3)\vartheta_2(q) + 4\mu_3\vartheta_4(q) \tag{35}$$

$$\frac{d}{dq}\vartheta_4(q) = -(2\mu_3 + 4\mu_3)\vartheta_4(q) + (2\lambda_3 + \alpha_3)\vartheta_3(q) + 5\mu_3\vartheta_5(q) \tag{36}$$

$$\frac{d}{dq}\vartheta_5(q) = -5\mu_3\vartheta_5(q) + 2\lambda_3\vartheta_4(q) \tag{37}$$

The normalizing condition for this problem is

$$\vartheta_0(t) + \vartheta_1(t) + \vartheta_2(t) + \vartheta_3(t) + \vartheta_4(t) + \vartheta_5(t) = 1 \tag{38}$$

Availability of subsystem C is

$$A_{S3} = \vartheta_0(t) + \vartheta_1(t) + \vartheta_2(t) + \vartheta_3(t) + \vartheta_4(t) \quad (39)$$

Setting (32) to (37) to zero as $q \to \infty$ in steady state, availability of subsystem C in (39) is now

$$A_{S3} = \frac{1 + \frac{m_3}{\mu_3} + \frac{m_3^2}{2\mu_3^2}\vartheta + \frac{m_3^3}{6\mu_3^3} + \frac{(2\lambda_3 + \alpha_3)m_3^3}{24\mu_3^4}}{1 + \frac{m_3}{\mu_3} + \frac{m_3^2}{2\mu_3^2} + \frac{m_3^3}{6\mu_3^3} + \frac{(2\lambda_3 + \alpha_3)m_3^3}{24\mu_3^4} + \frac{2\lambda_3(2\lambda_3 + \alpha_3)m_3^3}{120\mu_3^5}} \quad (40)$$

Where $m_3 = (2\lambda_3 + 2\alpha_3)$

The Corresponding reliability, maintainability, dependability and MTBF, MTTR for main and warm standby unit of subsystem C are

$$R_{S3}(q) = \exp^{-\lambda_3 q} \quad (41)$$

$$R_{S3}(q) = \exp^{-\alpha_3 q} \quad (42)$$

$$M_{S3} = 1 - \exp^{-\mu_3 q} \quad (43)$$

Mean time between failure (MTBF)= $\lambda_3^{-1} = 100\,h$ for main unit

Mean time between failure (MTBF)= $\alpha_3^{-1} = 71.4286\,h$ for warm standby unit

Mean time to repair (MTTR)= $\mu_3^{-1} = 5h$

Dependability ratio d= $\frac{\mu_3}{\lambda_3} = 14.9999$ for main unit

$$D_{\min}(s_3) = 1 - \left(\frac{1}{d-1}\right)\left(\exp^{-\frac{\ln d}{d-1}} - \exp^{\frac{d\ln d}{d-1}}\right) = 0.9451$$

for main unit

Dependability ratio d= $\frac{\mu_3}{\alpha_3} = 14.2857$ for warm standby unit

$$D_{\min}(s_3) = 1 - \left(\frac{1}{d-1}\right)\left(\exp^{-\frac{\ln d}{d-1}} - \exp^{\frac{d\ln d}{d-1}}\right) = 0.9427$$

for main and warm standby unit

RAMD Analysis for Subsystem D (Side Seam)

The side seam section consists of three primary operation unit, two warm standby unit and one cold standby unit. Using the method described in section 4.1 above, the Chapman-Kolmogrov differential difference equations (44)-(48) are derived using Markovian birth-death process from Table 7 below.

Where S_0 is the perfect state, S_1, S_2, S_3 are partial failure states and S_4 is the complete failure state

$$\frac{d}{dq}\vartheta_0(q) = -(3\lambda_4 + 2\alpha_4)\vartheta_0(q) + \mu_4\vartheta_1(q) \quad (44)$$

$$\frac{d}{dq}\vartheta_1(q) = -(3\lambda_4 + 2\alpha_4 + \mu_4)\vartheta_1(q) + (3\lambda_4 + 2\alpha_4)\vartheta_0(q) + 2\mu_4\vartheta_2(q) \quad (45)$$

$$\frac{d}{dq}\vartheta_2(q) = -(3\lambda_4 + \alpha_4 + 2\mu_4)\vartheta_2(q) + (3\lambda_4 + 2\alpha_4)\vartheta_1(q) + 3\mu_4\vartheta_3(q) \quad (46)$$

$$\frac{d}{dq}\vartheta_3(q) = -(3\lambda_4 + 3\mu_4)\vartheta_3(q) + (3\lambda_4 + \alpha_4)\vartheta_2(q) + 4\mu_4\vartheta_4(q) \quad (47)$$

$$\frac{d}{dq}\vartheta_4(q) = -4\mu_4\vartheta_4(q) + 3\lambda_4\vartheta_3(q) \quad (48)$$

The normalizing condition for this problem is

$$\vartheta_0(q) + \vartheta_1(q) + \vartheta_2(q) + \vartheta_3(q) + \vartheta_4(q) = 1 \quad (49)$$

Availability of subsystem D is

$$A_{S4} = \vartheta_0(q) + \vartheta_1(q) + \vartheta_2(q) + \vartheta_3(q) \quad (50)$$

Setting (44) to (48) to zero as $q \to \infty$ in steady state, availability of subsystem D in (50) is now

$$A_{S4} = \frac{1 + \frac{(2\lambda_2 + 2\alpha_2)}{\mu_2} + \frac{m_4^2}{2\mu_2^2} + \frac{(3\lambda_4 + \alpha_4)m_4^2}{6\mu_2^3}}{1 + \frac{(2\lambda_2 + 2\alpha_2)}{\mu_2} + \frac{m_4^2}{2\mu_2^2} + \frac{(3\lambda_4 + \alpha_4)m_4^2}{6\mu_2^3} + \frac{3\lambda_4(3\lambda_4 + \alpha_4)m_4^2}{24\mu_2^4}} \quad (51)$$
$$m_4 = (3\lambda_4 + 2\alpha_4)$$

The Corresponding reliability, maintainability, dependability and MTBF, MTTR for main and warm standby unit of subsystem D are

$$R_{S4}(q) = \exp^{-\lambda_4 q} \quad (52)$$

$$R_{S4}(q) = \exp^{-\alpha_4 q} \quad (53)$$

$$M_{S4} = 1 - \exp^{-\mu_4 q} \quad (54)$$

Mean time between failure (MTBF)= $\lambda_4^{-1} = 28.5714\,h$ for main unit

Table 7. Transition rate table for Subsystem D

	S_0	S_1	S_2	S_3	S_4
S_0	0	$3\lambda_4 + 2\alpha_4$	0	0	0
S_1	μ_4	0	$3\lambda_4 + 2\alpha_4$	0	0
S_2	0	$2\mu_4$	0	$3\lambda_4 + \alpha_4$	0
S_3	0	0	$3\mu_4$	0	$3\lambda_4$
S_4	0	0	0	$4\mu_4$	0

Mean time between failure (MTBF)= $\alpha_4^{-1} = 58.8235h$ for warm standby unit

Mean time to repair (MTTR)= $\mu_4^{-1} = 2.5h$

Dependability ratio d= $\frac{\mu_4}{\lambda_4} = 11.4286$ for main unit

$$D_{\min}(s_4) = 1-\left(\frac{1}{d-1}\right)\left(\exp^{-\frac{\ln d}{d-1}} - \exp^{\frac{d\ln d}{d-1}}\right) = 0.9307$$

for main unit

Dependability ratio d= $\frac{\mu_4}{\alpha_4} = 23.5294$ for warm standby unit

$$D_{\min}(s_4) = 1-\left(\frac{1}{d-1}\right)\left(\exp^{-\frac{\ln d}{d-1}} - \exp^{\frac{d\ln d}{d-1}}\right) = 0.9630$$

for main and warm standby unit

RAMD Analysis for Subsystem E (Cleaning)

The cleaning section consists of four primary operation unit, two warm standby unit and one cold standby unit. Using the method described in section 4.1 above, the Chapman-Kolmogrov differential difference equations (55)-(59) are derived using Markovian birth-death process from Table 8 below.

Table 8. Transition rate table for Subsystem E

	S_0	S_1	S_2	S_3	S_4
S_0	0	$4\lambda_5 + 2\alpha_5$	0	0	0
S_1	μ_5	0	$4\lambda_5 + 2\alpha_5$	0	0
S_2	0	$2\mu_5$	0	$4\lambda_5 + \alpha_5$	0
S_3	0	0	$3\mu_5$	0	$4\lambda_5$
S_4	0	0	0	$4\mu_5$	0

Where S_0 is the perfect state, S_1, S_2, S_3 are partial failure states and S_4 is the complete failure state

$$\frac{d}{dq}\vartheta_0(q) = -(4\lambda_5 + 2\alpha_5)\vartheta_0(q) + \mu_5\vartheta_1(q) \quad (55)$$

$$\frac{d}{dq}\vartheta_1(q) = -(4\lambda_5 + 2\alpha_5 + \mu_5)\vartheta_1(q) + (4\lambda_5 + 2\alpha_5)\vartheta_0(q) + 2\mu_5\vartheta_2(q) \quad (56)$$

$$\frac{d}{dq}\vartheta_2(q) = -(4\lambda_5 + \alpha_5 + 2\mu_5)\vartheta_2(q) + (4\lambda_5 + 2\alpha_5)\vartheta_1(q) + 3\mu_5\vartheta_3(q) \quad (57)$$

$$\frac{d}{dq}\vartheta_3(q) = -(4\lambda_5 + 3\mu_5)\vartheta_3(q) + (4\lambda_5 + \alpha_5)\vartheta_2(q) + 4\mu_5\vartheta_4(q) \quad (58)$$

$$\frac{d}{dq}\vartheta_4(q) = -4\mu_5\vartheta_4(q) + 4\lambda_5\vartheta_3(q) \quad (59)$$

The normalizing condition for this problem is

$$\vartheta_0(q) + \vartheta_1(q) + \vartheta_2(q) + \vartheta_3(q) + \vartheta_4(q) = 1 \quad (60)$$

Availability of subsystem E is

$$A_{S5} = \vartheta_0(q) + \vartheta_1(q) + \vartheta_2(q) + \vartheta_3(q) \quad (61)$$

Setting (55) to (59) to zero as $q \to 0$ in steady state, availability of subsystem E in (61) is now

$$A_{S5} = \frac{1 + \frac{m_5}{\mu_5} + \frac{m_5^2}{2\mu_5^2} + \frac{(4\lambda_5 + \alpha_5)m_5^2}{6\mu_5^3}}{1 + \frac{m_5}{\mu_5} + \frac{m_5^2}{2\mu_5^2} + \frac{(4\lambda_5 + \alpha_5)m_5^2}{6\mu_5^3} + \frac{\lambda_5(4\lambda_5 + \alpha_5)m_5^2}{24\mu_5^4}} \quad (62)$$

$$m_5 = (4\lambda_5 + 2\alpha_5)$$

The Corresponding reliability, maintainability, dependability and MTBF, MTTR for main and warm standby unit of subsystem D are

$$R_{S4}(q) = \exp^{-\lambda_5 q} \quad (62)$$

$$R_{S4}(q) = \exp^{-\alpha_5 q} \quad (63)$$

$$M_{S5} = 1 - \exp^{-\mu_5 q} \quad (64)$$

Mean time between failure (MTBF)= $\lambda_5^{-1} = 20h$ for main unit

Mean time between failure (MTBF)= $\alpha_5^{-1} = 76.9230h$ for warm standby unit

Mean time to repair (MTTR)= $\mu_5^{-1} = 1.8182h$

Dependability ratio d= $\frac{\mu_5}{\lambda_5} = 10.9285$ for main unit

$$D_{\min}(s_5) = 1-\left(\frac{1}{d-1}\right)\left(\exp^{-\frac{\ln d}{d-1}} - \exp^{\frac{d\ln d}{d-1}}\right) = 0.9285$$

for main unit

Dependability ratio d= $\frac{\mu_5}{\alpha_5} = 42.3072$ for warm standby uni

$$D_{\min}(s_5) = 1-\left(\frac{1}{d-1}\right)\left(\exp^{-\frac{\ln d}{d-1}} - \exp^{\frac{d\ln d}{d-1}}\right) = 0.9784$$

for main and warm standby unit

NUMERICAL SIMULATIONS AND DISCUSSION

Numerical simulations of reliability, availability, maintainability, and dependability are discussed in this section.

Reliability Using Exponential Distribution

Reliability Using Lindley Distribution

Reliability Using Exponentiated Weibull Distribution

This section discusses the numerical simulations in order to obtain understanding of how the strength, efficacy, and performance of the model under review are evaluated at various levels. Here, we employ the exponential, Lindley, and exponentiated Weibull distributions as three alternative distributions to first choose the optimum distribution that will improve system reliability. On the basis of this, the performance of the model is evaluated.

Table 9 and Figure 1 displayed the results of availability of individual subsystems and the entire system with respect to failure rates. From the table and figure, it is noted that availability of individual subsystems and the entire system decreases with increase in failure rate. It is clear from the table and figure that the availability of the system is lower than the availability of the individual subsystems. This can

Table 9. Variation in Availability of system due to with respect to availability of individual subsystem

Failure rate	Availability					System
	Subsystem A	Subsystem B	Subsystem C	Subsystem D	Subsystem E	
0.00	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.02	1.0000	1.0000	0.9998	1.0000	1.0000	0.9998
0.04	0.9999	0.9999	0.9994	1.0000	0.9999	0.9991
0.06	0.9998	0.9998	0.9989	1.0000	0.9999	0.9984
0.08	0.9997	0.9996	0.9982	0.9999	0.9998	0.9972

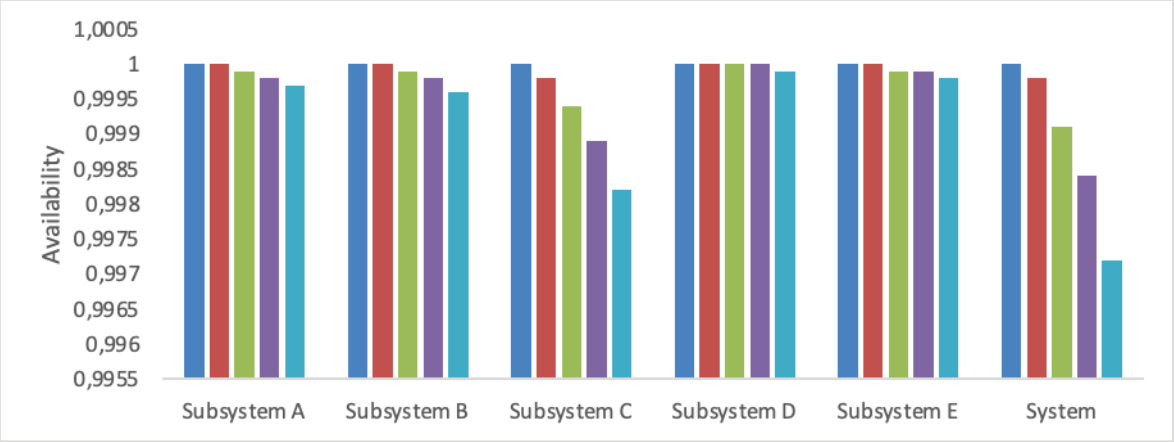


Figure 1. Availability of the system and individual subsystems.

Table 10. Variation in reliability of system due to changes in Exponential failure rate of subsystems for main unit

Time	Reliability of Subsystem A $\lambda_1 = 0.015$	Reliability of Subsystem B $\lambda_2 = 0.025$	Reliability of Subsystem C $\lambda_3 = 0.010$	Reliability of Subsystem D $\lambda_4 = 0.035$	Reliability of Subsystem E $\lambda_5 = 0.050$	System Reliability
0	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.00000000
20	0.74081822	0.60653066	0.81873075	0.49658530	0.36787944	0.06720551
40	0.54881164	0.36787944	0.67032005	0.24659696	0.13533528	0.00451658
60	0.40656966	0.22313016	0.54881164	0.12245642	0.04978706	0.00030353
80	0.30119421	0.13533528	0.44932896	0.06081006	0.01831563	0.00002039
100	0.22313016	0.082084999	0.36787944	0.03019738	0.00673794	0.00000137

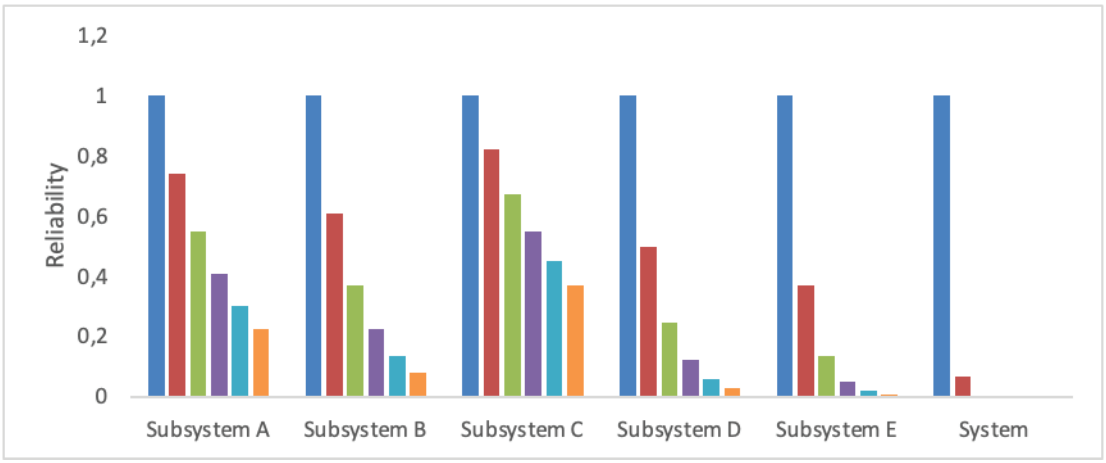


Figure 2. Variation in reliability of system due to changes in Exponential failure rate of subsystems for main unit.

Table 11. Variation in reliability of system due to changes in Exponential failure rate of subsystems for warm standby unit

Time	Reliability of Subsystem A $\alpha_1 = 0.015$	Reliability of Subsystem B $\alpha_2 = 0.016$	Reliability of Subsystem C $\alpha_3 = 0.014$	Reliability of Subsystem D $\alpha_4 = 0.017$	Reliability of Subsystem E $\alpha_5 = 0.013$	System Reliability
0	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000
20	0.74081822	0.72614903	0.75578374	0.7117703	0.77105158	0.22313016
40	0.54881163	0.52729242	0.57120906	0.50661699	0.59452054	0.04978706
60	0.40656965	0.38289288	0.43171052	0.36059494	0.45840601	0.01110899
80	0.30119421	0.27803730	0.32627979	0.25666077	0.35345468	0.00247875
100	0.22313016	0.2018965	0.24659696	0.18268352	0.27253179	0.00055308

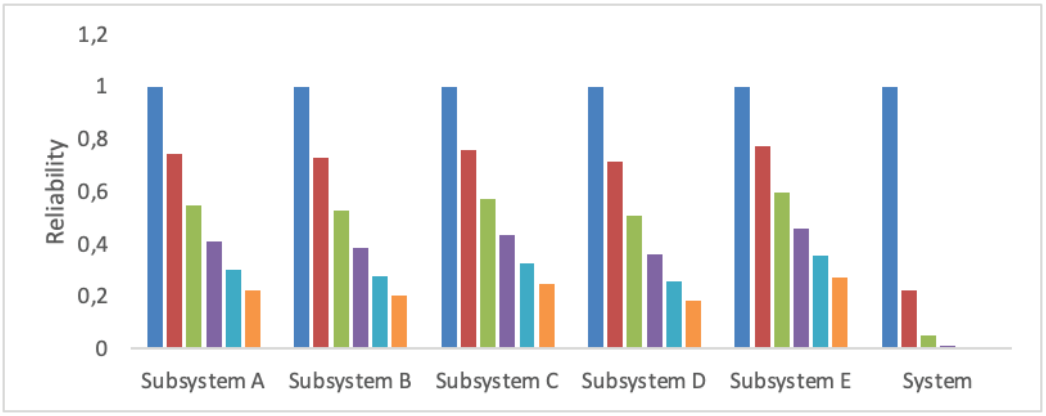


Figure 3. Variation in reliability of system due to changes in Exponential failure rate of subsystems for warm standby unit.

lead to decrease in production which will in turn culminated in less revenue mobilization. To avert this problem adequate preventive maintenance before such as regular inspection, oiling, greasing etc should be invoke to avoid system failure. From the table and figure, it is worthwhile to notice that subsystem C has the least availability. Therefore, maintenance priority should be set aside to subsystem C in order to improve its availability.

Table 10 and Figure 2 and table 11 and Figure 3 presents the results of reliability of the individual subsystems and the system when the failure rate of the main and warm standby unit follows exponential distribution. The table and figure show that reliability decreases drastically with passage of

Table 12. Variation in reliability of system due to changes in Lindley failure rate of subsystems for main unit

Time	Reliability of Subsystem A $\lambda_1 = 0.015$	Reliability of Subsystem B $\lambda_2 = 0.025$	Reliability of Subsystem C $\lambda_3 = 0.010$	Reliability of Subsystem D $\lambda_4 = 0.035$	Reliability of Subsystem E $\lambda_5 = 0.050$	System Reliability
0	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.0000000
20	0.95977927	0.90239927	0.98085565	0.83244010	0.71824081	0.14697050
40	0.87323231	0.72678621	0.93579333	0.58015807	0.39311677	0.00590089
60	0.76707478	0.54966210	0.87483835	0.37091875	0.19203583	0.00014415
80	0.65728589	0.39940413	0.80523309	0.22532037	0.08808950	0.00000267
100	0.55287917	0.28229231	0.73211651	0.13231414	0.03882340	0.0000004

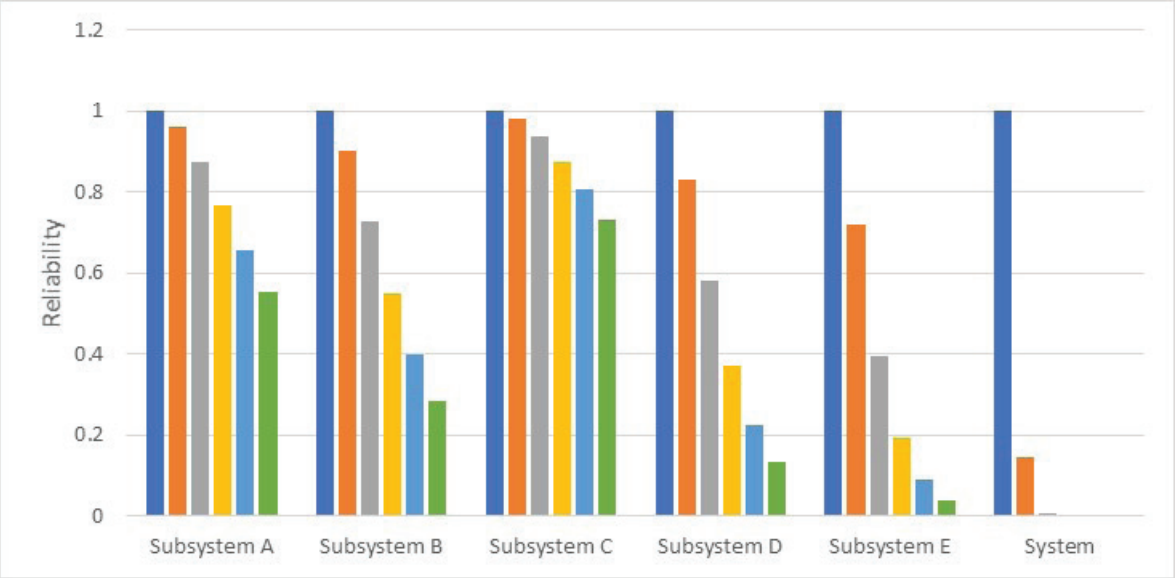


Figure 4. Variation in reliability of system due to changes in Lindley failure rate of subsystems for main unit.

Table 13. Variation in reliability of system due to changes in Lindley failure rate of subsystems for Warm standby Unit

Time	Reliability of Subsystem A $\alpha_1 = 0.015$	Reliability of Subsystem B $\alpha_2 = 0.016$	Reliability of Subsystem C $\alpha_3 = 0.014$	Reliability of Subsystem D $\alpha_4 = 0.017$	Reliability of Subsystem E $\alpha_5 = 0.013$	System Reliability
0	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000
20	0.95977927	0.95485739	0.96448142	0.94972696	0.96895228	0.50827697
40	0.87323231	0.85944513	0.88666969	0.84535795	0.89970384	0.03210535
60	0.76707478	0.74468143	0.78934054	0.72225358	0.81137411	0.00096330
80	0.65728589	0.62832051	0.68666773	0.59988462	0.71633017	0.00002017
100	0.55287917	0.51984379	0.58706614	0.48805421	0.62227644	0.0000004

time from 0 to 100. From the table and figure it can be seen that reliability of the system is less than the reliability of each subsystem. Subsystem E has the least reliability among the subsystems from the Table 10 and Figure 2 when the failure rate of the main unit obeys exponential distribution while subsystem D has the least reliability from Table 11 and Figure 3 when the failure rate of the warm standby unit obeys exponential distribution.

From Table 12 and Figure 4 and Table 13 and Figure 5 for reliability analysis of the individual subsystems and the system when the failure rate of the main and warm standby unit obeys Lindley distribution. It is observed from the

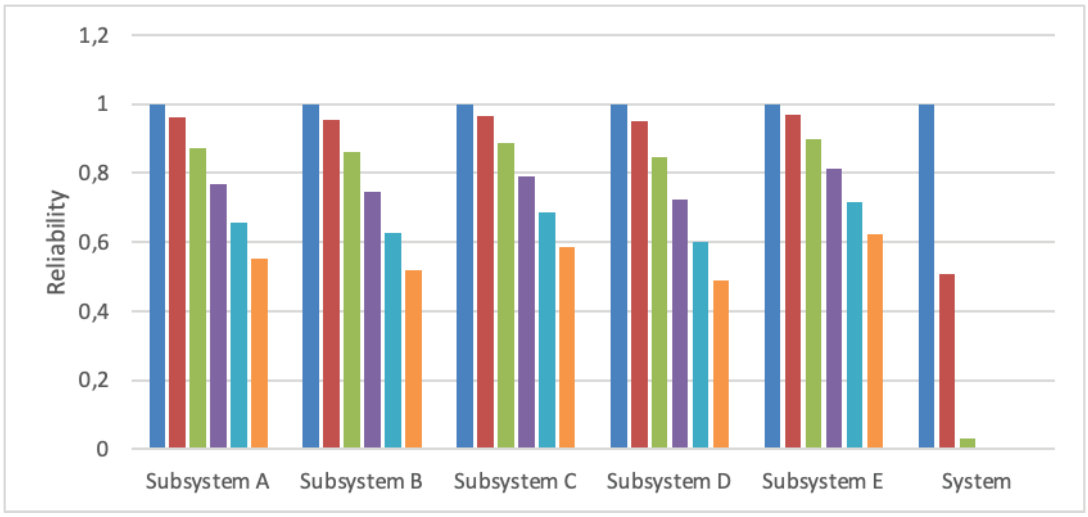


Figure 5. Variation in reliability of system due to changes in Lindley failure rate of subsystems for warm standby unit.

Table 14. Variation in reliability of system due to changes in Exponentiated Weibull failure rate of subsystems for main unit

Time	Reliability of Subsystem A $\lambda_1 = 0.015$	Reliability of Subsystem B $\lambda_2 = 0.025$	Reliability of Subsystem C $\lambda_3 = 0.010$	Reliability of Subsystem D $\lambda_4 = 0.035$	Reliability of Subsystem E $\lambda_5 = 0.050$	System Reliability
0	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.0000000
20	0.99660863	0.99094408	0.89131112	0.982932729	0.93282480	0.83679084
40	0.98721301	0.96714146	0.69676141	0.94035841	0.79642906	0.55758066
60	0.97286409	0.93282480	0.51167047	0.88238311	0.64784043	0.32632241
80	0.95447233	0.89131112	0.36303083	0.81613833	0.51167047	0.17562407
100	0.93282480	0.84518187	0.25235492	0.74657364	0.39647325	0.08918513

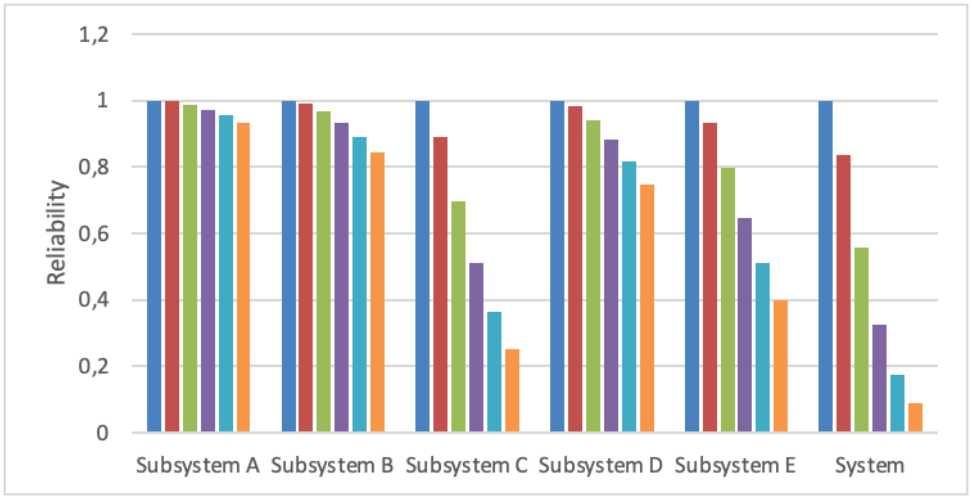


Figure 6. Variation in reliability of system due to changes in Exponentiated Weibull failure rate of subsystems for main unit.

Table 15. Variation in reliability of system due to changes in Exponentiated Weibull failure rate of subsystems for warm standby unit

Time	Reliability of Subsystem A $\alpha_1 = 0.015$	Reliability of Subsystem B $\alpha_2 = 0.016$	Reliability of Subsystem C $\alpha_3 = 0.014$	Reliability of Subsystem D $\alpha_4 = 0.017$	Reliability of Subsystem E $\alpha_5 = 0.013$	System Reliability
0	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000
20	0.99660863	0.99615661	0.99703401	0.99567831	0.99743243	0.98302549
40	0.98721301	0.98556478	0.98877338	0.98383100	0.99024355	0.93724977
60	0.97286409	0.96948230	0.97608456	0.96594586	0.97913685	0.87071433
80	0.95447233	0.94898814	0.95972558	0.94328660	0.96473380	0.79108357
100	0.93282480	0.92500565	0.94035841	0.91692365	0.94758262	0.70499843

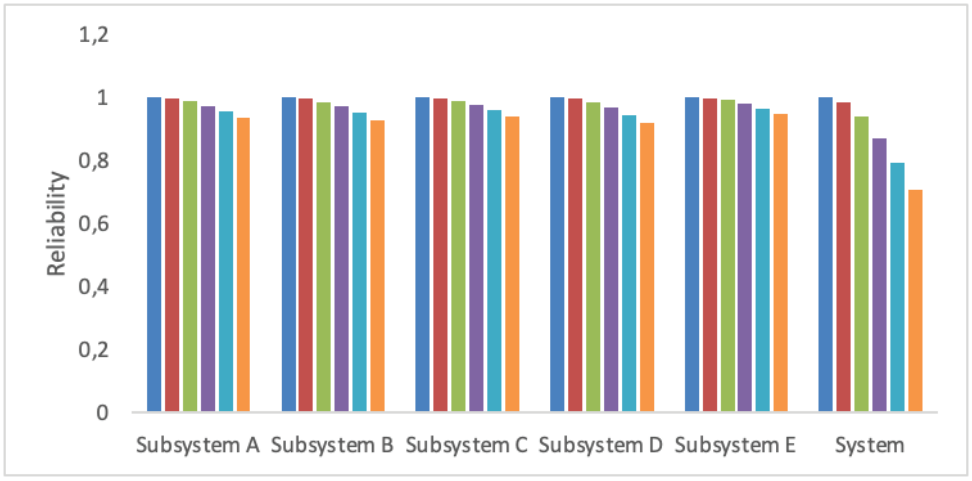


Figure 7. Variation in reliability of system due to changes in Exponentiated Weibull failure rate of subsystems for warm standby unit.

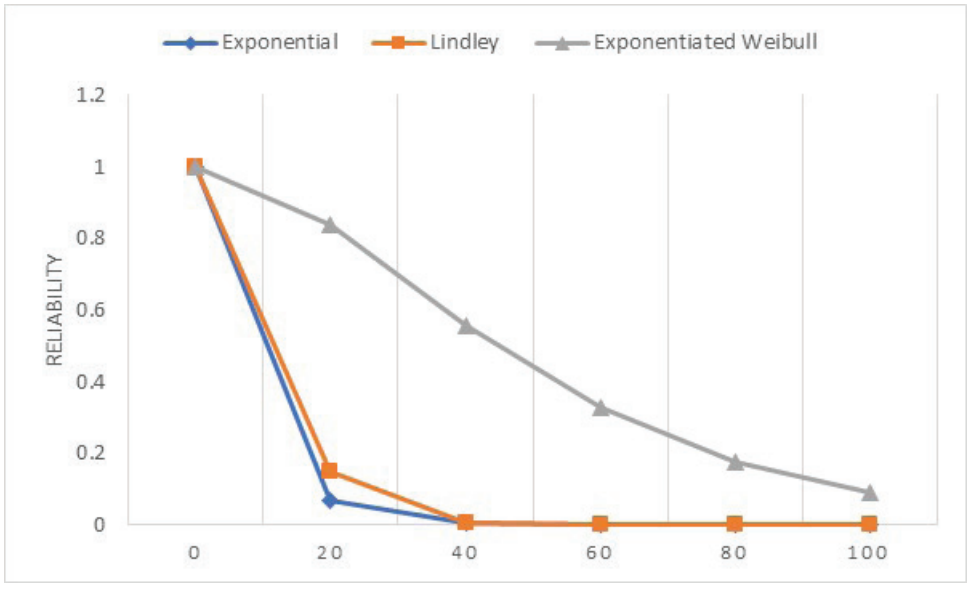


Figure 8. Reliability for main unit failure against time for different distributions.

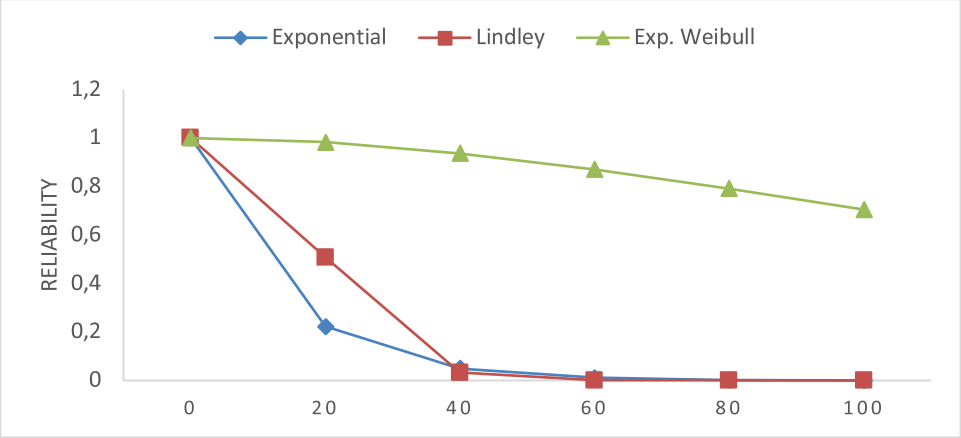


Figure 9. Reliability for warm standby unit failure against time for different distributions.

Table 16. Variation in maintainability of system due to with respect to of individual subsystem

Time	Maintainability of Subsystem A $\mu_1 = 0.35$	Maintainability of Subsystem B $\mu_2 = 0.20$	Maintainability of Subsystem C $\mu_3 = 0.15$	Maintainability of Subsystem D $\mu_4 = 0.40$	Maintainability of Subsystem E $\mu_5 = 0.55$	System Maintainability
0	0. 000000000	0. 000000000	0. 000000000	0. 000000000	0. 000000000	0. 000000000
20	0.9990881180	0.9816843611	0.9502129316	0.9996645374	0.9999832983	0.9316303655
40	0.9999991685	0.9996645374	0.9975212478	0.9999998875	0.9999999997	0.9971856751
60	0.9999999992	0.9999938558	0.9998765902	1.000000000	1.000000000	0.9998704460
80	1.000000000	0.9999998875	0.9999938558	1.000000000	1.000000000	0.9999937433
100	1.000000000	0.9999999979	0.9999996941	1.000000000	1.000000000	0.9999996920

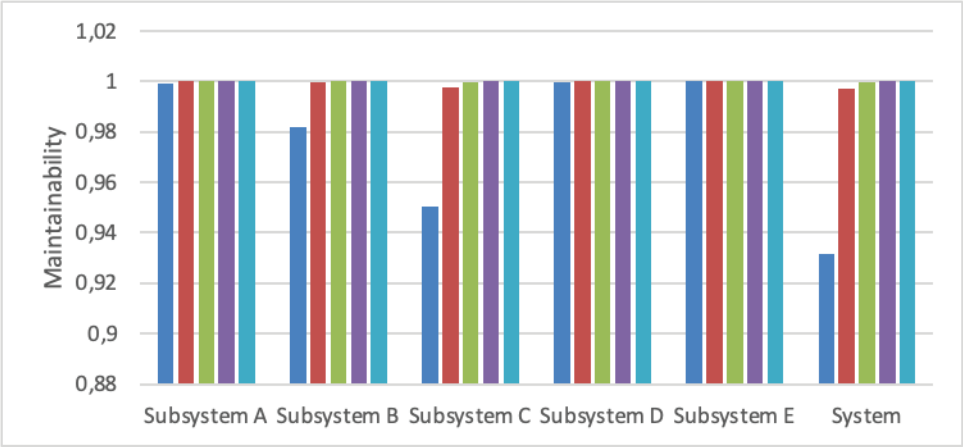


Figure 10. Variation in maintainability of system and subsystems.

tables and figures that reliability decreases slightly with passage of time from 0 to 100 in which reliability of the system is less than the reliability of each subsystem. It is evident from the tables and figures that subsystem E has the least reliability among the subsystems when the failure rate of the main obeys Lindley distribution and subsystem D for warm standby unit obeys Lindley distribution.

On other hand, when the failure follows exponentiated Weibull distribution for both main and warm standby unit From Table 14 and Figure 6 and Table 15 and Figure 7 for reliability analysis of the individual subsystems and the

Table 17. Ramd indices

Indices	Subsystem A	Subsystem B	Subsystem C	Subsystem D	Subsystem E
Reliability Main	$\exp^{-0.015q}$	$\exp^{-0.025q}$	$\exp^{-0.010q}$	$\exp^{-0.035q}$	$\exp^{-0.050q}$
Reliability Warm	$\exp^{-0.015q}$	$\exp^{-0.016q}$	$\exp^{-0.014q}$	$\exp^{-0.017q}$	$\exp^{-0.013q}$
Reliability Main	$\frac{1.015+0.015q}{1.015}e^{-0.015q}$	$\frac{1.025+0.025q}{1.025}e^{-0.025q}$	$\frac{1.010+0.010q}{1.010}e^{-0.010q}$	$\frac{1.035+0.035q}{1.035}e^{-0.035q}$	$\frac{1.050+0.050q}{1.050}e^{-0.050q}$
Reliability Warm	$\frac{1.015+0.015q}{1.015}e^{-0.015q}$	$\frac{1.016+0.016q}{1.016}e^{-0.016q}$	$\frac{1.014+0.014q}{1.014}e^{-0.014q}$	$\frac{1.017+0.017q}{1.017}e^{-0.017q}$	$\frac{1.013+0.013q}{1.013}e^{-0.013q}$
Maintainability	$1-\exp^{-0.35q}$	$1-\exp^{-0.20q}$	$1-\exp^{-0.15q}$	$1-\exp^{-0.40q}$	$1-\exp^{-0.55q}$
Reliability Main	$1-\left(1-e^{-(0.015q)^{0.2}}\right)^2$	$1-\left(1-e^{-(0.025q)^{0.2}}\right)^2$	$1-\left(1-e^{-(0.010q)^{0.2}}\right)^2$	$1-\left(1-e^{-(0.035q)^{0.2}}\right)^2$	$1-\left(1-e^{-(0.050q)^{0.2}}\right)^2$
Reliability Warm	$1-\left(1-e^{-(0.015q)^{0.2}}\right)^2$	$1-\left(1-e^{-(0.016q)^{0.2}}\right)^2$	$1-\left(1-e^{-(0.014q)^{0.2}}\right)^2$	$1-\left(1-e^{-(0.017q)^{0.2}}\right)^2$	$1-\left(1-e^{-(0.013q)^{0.2}}\right)^2$
Availability	0.9996	0.9995	0.9978	0.9999	0.9997
Dependability Main	0.9595	0.8877	0.9451	0.9307	0.9784
Dependability Warm	0.9595	0.9357	0.9427	0.9630	0.9784
MTTR	2.8571	5	5	2.5	1.8182
MTBF Main	66.6667	40	100	28.5714	20
MTBF Warm	66.6667	62.5	71.4286	58.8235	76.9230
Dependability ratio Main	23.3345	8	14.9999	11.4286	10.9285
Dependability ratio Warm	2345.33	12.5	14.2857	23.5294	42.3072

system it is clear that reliability decreases slightly with passage of time from 0 to 100 in which reliability of the system is less than the reliability of each subsystem. It is evident from the tables and figures that subsystem C for main unit has the least reliability among the subsystems and subsystem D is the least when the failure rate of warm standby unit obeys exponentiated Weibull distribution.

Exponentiated Weibull distribution, in contrast, has a higher system reliability than the other two distributions for both main unit and warm standby units. This is seen in Figure 8 and 9, Table 14 and Figure 6 and Table 15 and Figure 7. The variation in system reliability caused by variations in the exponentiated Weibull failure rate of subsystems for main units is depicted in table 13 and figure 6. From this table 13 and its corresponding figure 6, we can see that the system reliability's equivalent values for main unit at time $t = 40$ are $Rel._{subsystem\ A} = 0.98721301$, $Rel._{subsystem\ B} = 0.96714146$, $Rel._{subsystem\ C} = 0.69676141$, $Rel._{subsystem\ D} = 0.94035841$, and $Rel._{subsystem\ D} = 0.79642906$. In time $t = 40$, there is $Main._{system} = 0.32632241$ chance of successfully completing maintenance and repairs, and $Main._{subsystem\ A} = 0.99999916$, $Main._{subsystem\ B} = 0.99966453$, $Main._{subsystem\ C} = 0.99752124$, $Main._{subsystem\ D} = 0.99999988$ and $Main._{subsystem\ E} = 0.999999999$. The system is 0.33632241 times reliable at $t = 60$ due to a form decline. This is brought on by the low reliability value of subsystem C. This demonstrates that subsystem C is the main unit's key subsystem. The value of availability is another indicator of how important subsystem C is to the main unit..

Table 9-15 and Figure 1-7 show the variation in system reliability caused by changes in the exponential, Lindley and exponentiated Weibull failure rate of the main and warm standby unit's subsystems. Subsystems with the lowest reliability value among the other subsystems need adequate attention of the management for proper maintenance in order to avoid system breakdown and subsequent loss of production and revenue as the tables and figures make sufficient evident. This demonstrates that critical subsystems are the most important and delicate part of the system and needs careful consideration.

CONCLUSION

In this study, the metrics of RAMD for both weaving, dry clean, cross cut, side seam and cleaning section of the textile are analyzed to assess the performance of the textile manufacturing system. Expressions associated with metrics for weaving, dry clean, cross cut, side seam and cleaning section have been derived and numerical experiments are performed. The assumed values for failure and repair rates for each subsystem are given in table 1. Table 16 lists all RAMD measurements, while tables 3 and 4 capture the variation in reliability and maintainability over time, respectively. Tables 9, 10, 11, 13 and 14 indicate the impact of different failure rates on subsystems and system reliability and figures 2-7 that side seam is the most important and delicate component of the system. The models/results described in this work, if modified, will allow management to stop poor reliability assessments and decision-making, which will cause high expenditures. Moreover, the accepted framework for the model under consideration's inspection and maintenance could be proposed and incorporated to satisfy the client and lower failure rates. These are the findings of the current investigation. This work can be enlarged to include both offline and online routine maintenance at both partial and total failure states. This study will be carried out in the future.

REFERENCES

[1] Aggarwal AK, Kumar S, Singh V. Performance modeling of the serial processes in refining system of a sugar plant using RAMD analysis. Int J Syst Assur Eng Manag 2017;8:1910-1922. [CrossRef]

[2] Aggarwal AK, Kumar S, Singh V. Reliability and availability analysis of the serial processes in skim milk powder system of a dairy plant: a case study. Int J Ind Syst 2016;22:36–62. [CrossRef]

[3] Kumar A, Pant S, Singh SB. Availability and cost analysis of an engineering system involving subsystems in series configuration. Int J Qual Reliabil Manag 2017;34:879-894. [CrossRef]

[4] Corvaro F, Giacchetta G, Marchetti B, Recanati M. Reliability, availability, maintainability (RAM) study on reciprocating compressors API 618. Petroleum 2017;3:266–272. [CrossRef]

[5] Garg H. Reliability, Availability and Maintainability analysis of industrial system using PSO and fuzzy methodology. MAPAN-J Metrol Soc India 2014;29:115–129. [CrossRef]

[6] Velmurugan K, Venkumar P, Sudhakarapandian R. Reliability availability maintainability analysis in forming industry. Int J Eng Adv Technol 2019;9:822-828.

[7] Jagtap HP, Bewoor AK, Kumar R, Ahmadi MH, Assad MEH, Sharifpur M. RAM analysis and availability optimization of thermal power plant water circulation system using PSO. Energy Rep 2021;7:1133-1153. [CrossRef]

[8] Jakkula B, Mandela G, Chivukula S. Reliability, availability and maintainability (RAM) investigation of Load Haul Dumpers (LHDs): a case study. Int J Syst Assur Eng Manag 2022;13:504-515. [CrossRef]

[9] Goyal D, Kumari A, Saini M, Joshi H. Reliability, maintainability and sensitivity analysis of physical processing unit of sewage treatment plant. SN Appl Sci 2019;1:1507. [CrossRef]

[10] Danjuma MU, Yusuf B, Yusuf I. Reliability, availability, maintainability, and dependability analysis of cold standby series-parallel system. J Comput Cogn Eng 2022;1-8. [CrossRef]

[11] Choudhary D, Tripathi M, Shankar R. Reliability, availability and maintainability analysis of a cement plant: a case study. Int J Qual Reliab Manag 2019;36:298–313. [CrossRef]

[12] Gupta N, Kumar A, Saini M. Reliability and maintainability investigation of generator in steam turbine power plant using RAMD analysis. J Phys Conf Ser 2021;1714:012009. [CrossRef]

[13] Kumar A, Singh R, Saini M, Dahiya O. Reliability, availability and maintainability analysis to improve the operational performance of soft water treatment and supply plant. J Eng Sci Technol Rev 2020;13:183-192. [CrossRef]

[14] Kumari A, Saini M, Patil RB, Al-Dahidi S, Mellal MA. Reliability, availability, maintainability, and dependability analysis of tube-wells integrated with underground pipelines in agricultural fields. Adv Mech Eng 2022;14:1-17. [CrossRef]

[15] Saini M, Kumar A. Performance analysis of evaporation system in sugar industry using RAMD analysis. J Braz Soc Mech Sci Eng 2019;41:4. [CrossRef]

[16] Saini M, Kumar A, Shankar VG. A study of microprocessor systems using RAMD approach. Life Cycle Reliab Saf Eng 2020;9:181–194. [CrossRef]

[17] Saini M, Kumar A, Sinwar D. Parameter estimation, reliability and maintainability analysis of sugar manufacturing plant. Int J Syst Assur Eng Manag 2022;13:231-249. [CrossRef]

[18] Saini M, Yadav J, Kumar A. Reliability, availability and maintainability analysis of hot standby database systems. Int J Syst Assur Eng Manag 2022;13:2458-2471. [CrossRef]

[19] Soltanali H, Garmabaki AH, Thaduri A. Sustainable production process: an application of reliability, availability, and maintainability methodologies in automotive manufacturing. Proc IMechE Part O: J Risk Reliabil 2019;233:682-697. [CrossRef]

[20] Kumar A, Goyal D, Saini M. Reliability and maintainability analysis of power generating unit of sewage treatment plant. Int J Stat Reliab Eng 2020;7:41-48.

[21] Yen TC, Wang KH. Cost benefit analysis of four

retrial systems with warm standby units and imperfect coverage. Reliab Eng Syst Saf 2020;202. [CrossRef]

[22] El-Ghamry E, Muse AH, Aldallal R, Mohamed MS. Availability and reliability analysis of a k-out-of-n warm standby system with common-cause failure and fuzzy failure and repair rates. Math Probl Eng 2022;2022:1–11. [CrossRef]

[23] Jia H, Liu D, Li Y, Ding Y, Liu M, Peng R. Reliability evaluation of power systems with multi-state warm standby and multi-state performance sharing mechanism. Reliab Eng Syst Saf 2020;204:107139. [CrossRef]

[24] Kumar A, Malik SC, Pawar D. Profit analysis of a warm standby non-identical units system with single server subject to priority. Int J Future Revolut Comput Sci Commun Eng 2018;4:108-112.

[25] Liu ZC, Hu LM, Liu SJ, Wang YY. Reliability analysis of a warm standby series-parallel system with different switches and bi-uncertain lifetimes. Iran J Fuzzy Syst 2021;18:187–202.

[26] Tenekedjiev K, Cooley S, Mednikarov B, Fan G, Nikolova N. Reliability simulation of two component warm-standby system with repair, switching, and back-switching failures under three aging assumptions. Mathematics. 2021;9:2547. [CrossRef]

[27] Kumar A, Pawar D, Malik SC. Profit analysis of a warm standby non-identical unit system with single server performing in normal/abnormal environment. Life Cycle Reliab Saf Eng 2019;8:219-226. [CrossRef]

[28] Kumar A, Pawar D, Malik SC. Economic analysis of a warm standby system with single server. Int J Math Stat Invent 2018;6:1-6.

[29] Kumari U, Sharma DC. Performance analysis of a warm standby machine repair problem with servers vacation, impatient and controlling F-policy. MESA 2021;12:1–21.

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1133–1147, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00093

Research Article

Blocking harmful images with a deep learning based next generation firewall

Kenan BAYSAL[1,*], Deniz TAŞKIN[2]

[1]*Hayrabolu Vocational School, Tekirdag Namik Kemal University, Tekirdag, 59400, Türkiye*
[2]*Department of Computer Engineering, Trakya University, Edirne, 22030, Türkiye*

ARTICLE INFO

Article history
Received: 06 January 2023
Revised: 16 February 2023
Accepted: 12 March 2023

Keywords:
Convolutional neural network; Deep learning; Firewall

ABSTRACT

There are various blocking and filtering algorithms for protection against harmful contents on the Internet. However, it is impossible to classify particularly the visual contents according to their genres and block them through traditional methods. In order to block the harmful visual contents, such as various advertisements and social media posts, we need to review and classify them as per their contents. Deep learning method is today's most efficient method to review the visual contents.

In this study, only the harmful images were blocked without completely blocking the entire website. Alcoholic drinks were selected as the harmful content data set. For this purpose, a training was provided with 4.6 million images by using CNN (Convolutional Neural Networks) and GoogLeNet architecture. At the end of this training, 97.6469% of accuracy was achieved. F1 score was calculated as 87.75526188% at the end of the test conducted with 154501 images.

The images were determined through the network traffic via mitmproxy and classified as harmful or harmless thanks to the trained model, and the filtering process was successfully completed.

Cite this article as: Baysal K, Taşkın D. Blocking harmful images with a deep learning based next generation firewall. Sigma J Eng Nat Sci 2024;42(4):1133–1147.

INTRODUCTION

Artificial neural networks aim to create machines that can decide on their own with living mathematical models inspired by biological neural tissue samples. The idea of creating a mathematically similar model to the biological structure of the brain in a computer environment was suggested by McCulloch et al. [1] They showed that logical expressions can be created with the artificial nerve cell they created using logical electrical circuits. It is accepted that the first artificial nerve cell was created with this study. It was concluded that more than one artificial nerve cell should be used to create a useful structure, and the foundation of artificial neural networks was laid out.

Nowadays, various studies are carried out in many fields by making use of artificial neural networks. Demir and Karaboga have studied on successful and unsuccessful

***Corresponding author.**
*E-mail address: kbaysal@nku.edu.tr

This paper was recommended for publication in revised form by Editor in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey

students mathematical achievements [2]. According to their studies results, Jordan Neural Networks (JNN) has been more successful to predict students achievements than the others. Rajkovic et al have used ANNs in their study to optimize parameters of producing biodiesel from sunflower oil [3]. Zettler et al have used ANN combined with fuzzy logic for pressure sensitive grouting on rock surfaces [4]. They used ANNs on Transient Pressure Analysis (TPA) section to control grouting process. Ma et al have used ANNs to early detection and diagnosis of chronical kidney disease [5]. They used Heterogeneous Modified Artificial Neural Network (HMANN) method, which is reduced the noise and helps a clear identification on kidney images, for training the model. Sabilla et al have worked on an electronic nose on their studies [6]. They have used ANNs for estimate gas concentration. They used fresh air, raw mango, ripe mango and rotten mango parameters to train the model. According to their results model has good results on prediction. Esen et al. have used ANNs modelling a solar assisted heat pump systems [7]. And they have made performance analysis by using ANNs. Esen et al also studied to predicting performance of a ground-source heat pump systems using adaptive-network based fuzzy inference systems (ANFIS) [8]. According to their conclusion, ANFIS can be used to predict performance for such ground-source heat pump systems. Efe and Alganci have used machine learning methods to classify the land covers from satellite images [9]. They classified the images as, artificial fields, forests, water bodies, bare and semi-natural areas, agricultural fields and urban texture. They used these classified data to detect urban expansion in the areas they studied.

The Aim of the Study

Due to the high-speed development of electronics and computer technologies from the second half of the 1900s, the 2000s are known as the "Information Age" or the "Internet Age". With the rapid development and evolution of the concept of social media within the last decade, the use of Internet had observable and permanent impacts on human relations. Due to the Internet's inherently global structure, it is observed that activities, such as consumption, communication, and interaction, influence large masses in different ways regardless of space and time. Visual images constitute the majority of this interaction. According to 2021 report of Wearesocial.com website, there are 4.20 billion active social media users among 4.66 billion active Internet users throughout the world. According to the same data, people spend approximately 6 hours and 54 minutes daily on devices connected to the Internet. In Turkey, people spend approximately 2 hours and 57 minutes daily on social media [10].

Cited from Lehu and Bressoud, Yaraş et al., stated that with the decreasing impact of traditional media, the advertisers began to use product placement technique as a marketing method [11]. Yaraş et al., stated that the companies prefer product placement techniques in order to find a way around advertisement restrictions for weapons, alcoholic drinks and tobacco products and influence particularly the young population [11].

A considerable part of the contents posted on social media includes product promotions and product placement. Although some contents, such as violence and sexuality, are blocked by the social media platforms, alcoholic drinks, tobacco use, or firearms are not subject to this filtering. The websites other than the social media contain more images that may affect the cognitive activities of the youth, such as violence and sexuality.

Especially the children are more vulnerable than the adults against the inappropriate contents of the websites. İplikçi and Batu conducted a content analysis study on advertisements on children's websites in Turkey. The results of this study showed that there were 18+ adult contents within these advertisements, and some other contents that might have a negative impact on children's mental health. Cited from Kapferer, İplikçi et al., stated that children under 5 years of age do not have any knowledge about the functions of advertisement images, and 26% of children consider advertisements as reliable sources of information [12].

There are numerous standards and regulations regarding visual media contents. In his study about media ethics and codes of conduct for children, Uzun, reviewed the ethical rules of BBC, however, he reported that considering the commercial structures of these web-casting organizations, they could not be expected to make any regulations regarding regulation and enhancement of these ethical rules. In this case, one may conclude that blocking and banning an entire website may fix this problem [13].

In his study on the impacts of visual media on children's mental health, Kanbur, stated that in today's world, completely prohibiting the children and the young from using visual media tools is not an effective solution. These restrictions may cause rejection and peer pressure among the social circle of the children. This may cause pressure on children [14].

It would not be a realistic approach to expect an enhancement and regulation on existing conditions from content creators, social media companies, and other digital media organizations regarding these contents. Therefore, in order to avoid these harmful contents, the users need filtering software programs. The filtering processes through the existing firewall system are performed as creating certain rules and word-based or address blocking. This system cannot detect whether there are inappropriate contents within an image. Thus, content-based filtering is not possible through traditional methods. Filtering process occurs in accordance with a predetermined domain list or a certain URL. This means complete blocking of a website. However, completely banning these channels, such as social media, which are home to various contents and environments, and which have a commercial value in terms of advertisement revenues, would lead to various problems. Therefore, content-based filtering is required.

Within this context, the purpose of this study is to detect contents through a smart algorithm, and to ensure safe Internet use without complete prohibition of the Internet.

Deep Learning

The concept of deep learning has been formed with contribution of various disciplines in the background as a subfield of machine learning. The first actually operating algorithm for deep learning is accepted as the study of Ivakhnenko and Lapa [15]. Although deep learning has come to a deadlock in the 90s due to hardware equipment inadequacy, with the increasing GPU performance, it was begun to be used in many fields within the last decade. Successful results were obtained via deep learning, such as computer games, speech and emotion recognition, natural language processing, and autonomous vehicles [16-19].

Deep learning is considered as a field of research with highest potential that will have a big impact for years to come [20]. CAFFE (Convolutional Architecture for Fast Feature Embedding), Tensorflow, Torch, and Keras are among the widely-used deep learning library stacks. Thanks to these open-source and easily-accessible library stacks, successful results were obtained in visual recognition and classification [21].

Deep learning uses the layer structure of artificial neural networks. It mimics the learning processes of human brain with mathematical models and processes the data and creates semantic concepts [22]. This system is called deep learning because the layers are interconnected, and the system has a multilayer and deep structure. It is accepted as a subfield of machine learning. Unlike machine learning, numerous hidden layers are used in artificial neural network layer structure of deep learning. The training is performed with high amount of data without deducing any features.

The algorithm, which is trained by using existing data sets, creates its own rules through multilayer artificial neural networks and can filter out the required information. Unlike traditional methods, training of deep learning algorithm requires large amounts of data sets. As the size of the data set increases, the working accuracy and efficiency of the trained algorithm increases [23]. As shown in Figure 1, the size of the neural network increases with the size of the data, and this has a positive impact on performance. Although a small neural network gives better results than traditional classifiers on a small amount of data, the large-scale neural networks provide higher accuracy performance on large amount of data [24].

Deep learning methods might not be always a best option. For example, Yiğit et al. have conducted studies in the financial field and compared traditional methods with deep learning methods to predict market price movements. In their study, they chose BIST30, BIST50 and BIST100 data as the data set. According to their results ARIMA method, which is mentioned as traditional method on their studies, has been better results then LSTM and GRU, which are deep learning methods. They referred that, deep learning methods has much better results for massive data sets, but ARIMA is one-step ahead forecasting on univariate datasets [25].

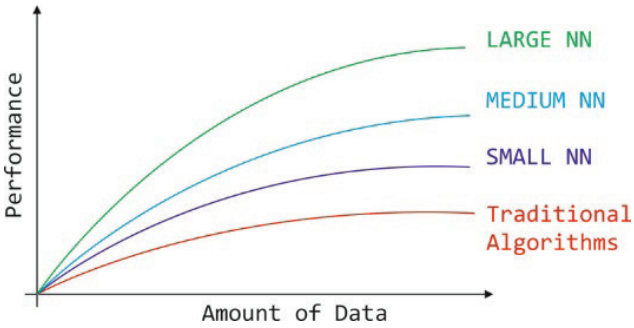


Figure 1. The relationship between data size and performance of deep neural networks [24].

Convolutional Neural Networks

Convolutional neural networks (CNN) are Artificial Neural Network (ANN) architectures developed based on the image processing method of the human eye. It is used for processes, such as filtering out, classification, or object recognition of shapes or scenes searched within the images. It learns about low-level characteristics of an object in order to describe it and learns about separating that object from the others. For example, in order to describe a cat, all sub-features that makes up a cat are gradually deduced. The large number of parameters calculated in traditional artificial neural network architectures lead to higher process power.

Expansion of data sets necessitates deduction of features particularly in processes, such as visual recognition. The image, which is the input value in CNN structure, are subjected to various filters in the convolution layer and the features are deducted, and the network is trained over these features. These filters may be various vertical and horizontal side filters, which will ensure side capturing in the image.

In their study, LeCun et al. became successful in recognizing hand-written numbers by using CNN architecture [26]. This study is considered as a pioneering work in CNN. After that, there has been a huge gap until the 2010 for studies on CNN. In due course, with the cell phone cameras becoming widespread, a boom of visual materials and tagged data has occurred in social media and other web environments, and the required effloresce has reached the desired level with GPU power. The deep learning model created by using CNN architecture has won the 2012 ImageNet image recognition competition, and CNN has been proven to be an efficient architecture [27]. In the long view, CNN architecture has become a popular method for image recognition and classification.

CNN training process is a multilayer structure. Throughout this training, the values obtained at the end

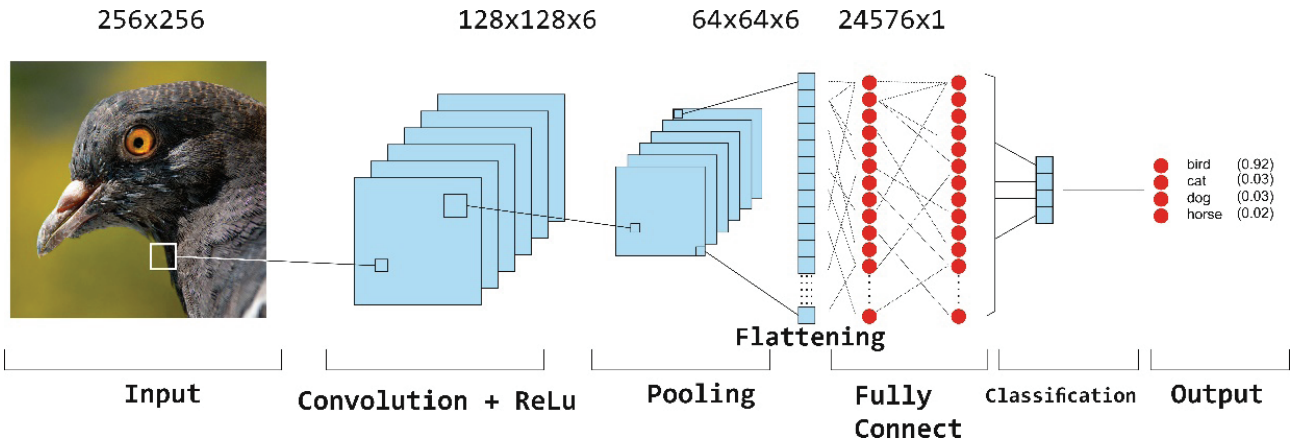


Figure 2. CNN Layer Structure.

of convolutional processes applied through various filters are subjected to activation functions. The backpropagation occurs and the weights are updated with the loss value obtained at the end of a training tour performed on neural networks. The training process continues until the best possible result is obtained.

The layers of CNN can be listed as the input layer, convolution layer (convolution + ReLU), pooling layer, full connection layer, classification layer, and output layer. The rankings of these layers are given in Figure 2.

The input layer is the first layer of the network, and the raw data is given to the network at this layer. If the size of the input data is too large at this stage, the time needed for training and the need for memory would increase. On the contrary, if the size of the input data is too small, the depth of the network would decrease since details would get lost, and the features that may be deduced would be insufficient. The size of the input data determines the depth of the network, the number of filters to be applied determines the width of the network.

MATERIALS AND METHODS

This study aims to develop a system to filter only the images instead of blocking an entire web site URL, where potentially harmful images may be present. A website may not always contain only harmful images. For example, a blog may prepare a content including these harmful images in only a few of its posts. The remaining contents may include images that may be helpful, rather than including harmful images. Social media is another example of this case. Social media platforms, such as Instagram, Facebook, Twitter, etc. may include helpful groups, lists, and posts of friends and family, but they may also include posts with images that can easily get through the filters of these sites. Completely blocking these websites, which are a part of social life in terms of communication, could escalate this problem to another level. In addition, URL-based blocking contains various challenges within itself. The blocked websites may get through this blocking with another address. Continuously tracking these changes and updating this blocked list can be considered as another workload. The filters of these websites can be flexible in terms of violence, sexuality, and substance abuse. In this study, the factors encouraging alcohol use were discussed, and a deep learning model was trained, which can recognize the images including alcoholic drinks and environments containing alcoholic drinks.

The process steps of this study Are given in Figure 3. This study consists of six main sections as provision of images for the data set, organization and tagging of the data set, selection of hyper-parameters, model performance evaluation, and firewall application.

The main harmful content selected for the filtering process was alcoholic beverages. The alcoholic drinks are gathered under 43 different titles solely according to their raw materials under the alcoholic drinks title of Wikipedia website. The same website lists fermented drinks under 64 titles and distilled drinks under 78 titles [28]. Considering each category has its own brands and packaging style, the size of the data set has to be extremely large. In addition, in case the color and packages of these drinks resemble the packages and color of nonalcoholic drinks, data for comparison with nonalcoholic beverages should be given in higher amounts in the data set. In addition to this variety, since there is a pattern resemblance between some alcoholic drinks and their real-life images, the real-life images were also included within the data set. At the image evaluation stage, the existence of alcoholic drinks within an image has also been considered rather than only blocking the packages of alcoholic drink images.

Data Set

Since there are no public data sets for alcoholic drinks at a size fit for this purpose, the images used in model training were obtained through the Internet via a web crawler designed in Python environment. The designed crawler structure worked continuously on an address database

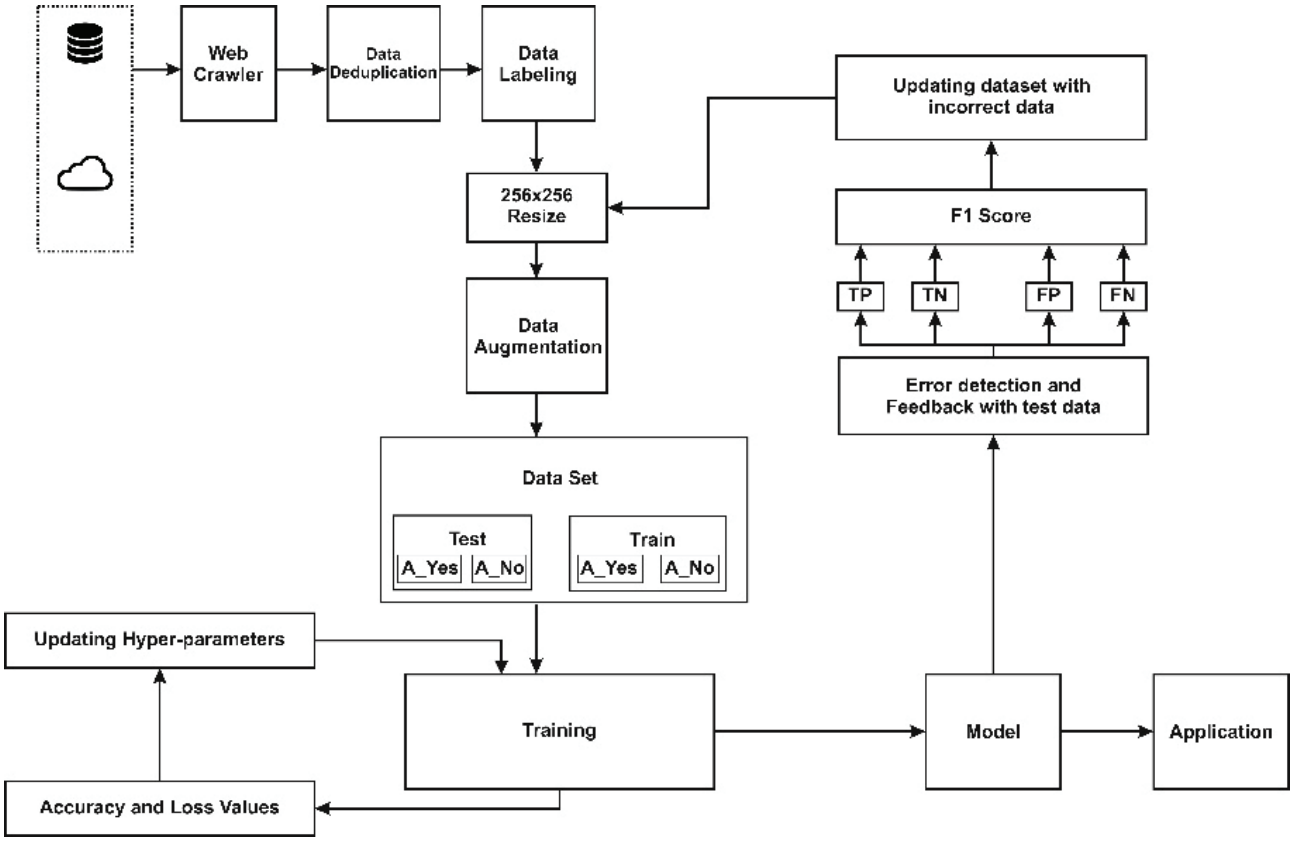


Figure 3. Process Steps.

containing various web addresses for 7 months, and approximately 3 million images were obtained.

Data sets were tagged in two main categories. The tags were determined as "alkol_var" which means "alcohol_yes" and "alkol_yok" which means "alcohol_no" according to whether there are alcoholic drinks within the image for filtering. Figure 4 shows a small example of the data set containing images tagged "alkol_var" and "alkol_yok".

While preparing this data set, not only the images containing alcoholic drinks in the foreground. Many different environments and scenes including exchange of toasts, crowded night clubs and entertainment venues that serve alcoholic drinks, and dinner tables including alcoholic drinks were tagged as "alkol_var".

Data Deduplication

In case an image is found more than once in the data set, the trained model may perform false learning towards a certain direction or begin to memorize. Therefore, all images were compared through Mean Squared Error (MSE) function. MSE is a risk function, and it is also used to measure estimation performance of machine learning models. It compares the desired value and the obtained value and presents a result. It always produces a positive value [29]. Thus, it is suitable for comparing similarities between images. This method was used to find other copies of multiple images during data set formation. The images with less than 5% of differences were eliminated.

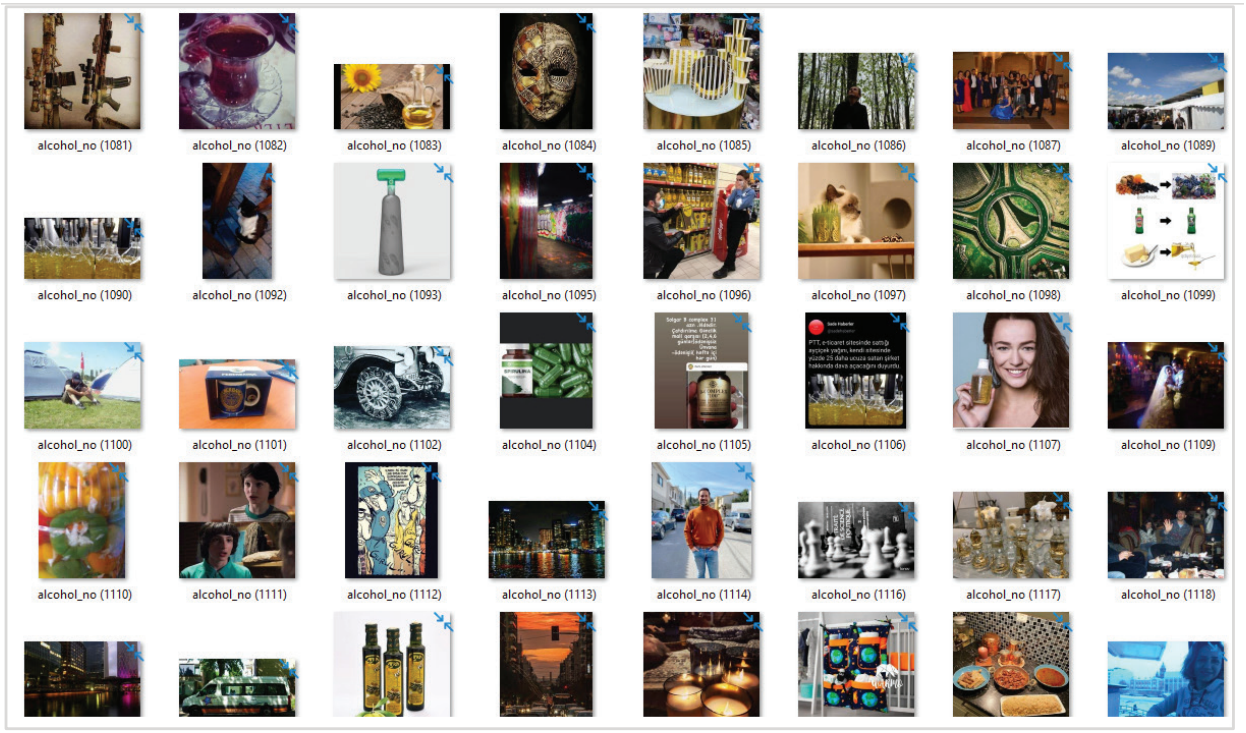
$$MSE = \sum_{n}^{i=1}(y_i - f(x_i))^2 \quad (1)\ [29]$$

Data Augmentation

Deep artificial neural networks require massive sizes of data set for the learning process. Limited amount of image data can be augmented and enhanced through various algorithms. Geometric and photometric methods are used to augment visual data [30]. Geometric augmentation is the process of creating a different image to be used in CNN learning process by changing the direction and position of the initial form of an image. These processes may be reversing, cropping, scaling, and rotating. Photometric augmentation is the process of augmentation that is applied through falsifying the chrominance channels of the image [31].

Augmentor library in python environment was used for data augmentation process. Figure 5 shows images obtained through data augmentation from an image.

Augmentor is a Python software suite that is used for data generation for machine learning problems [32]. It is possible to select probability parameters for each process. Table 1 shows the processes and parameters. Each image in



(a) alcohol_no examples

(b) alcohol_yes examples

Figure 4. Data Set Examples.

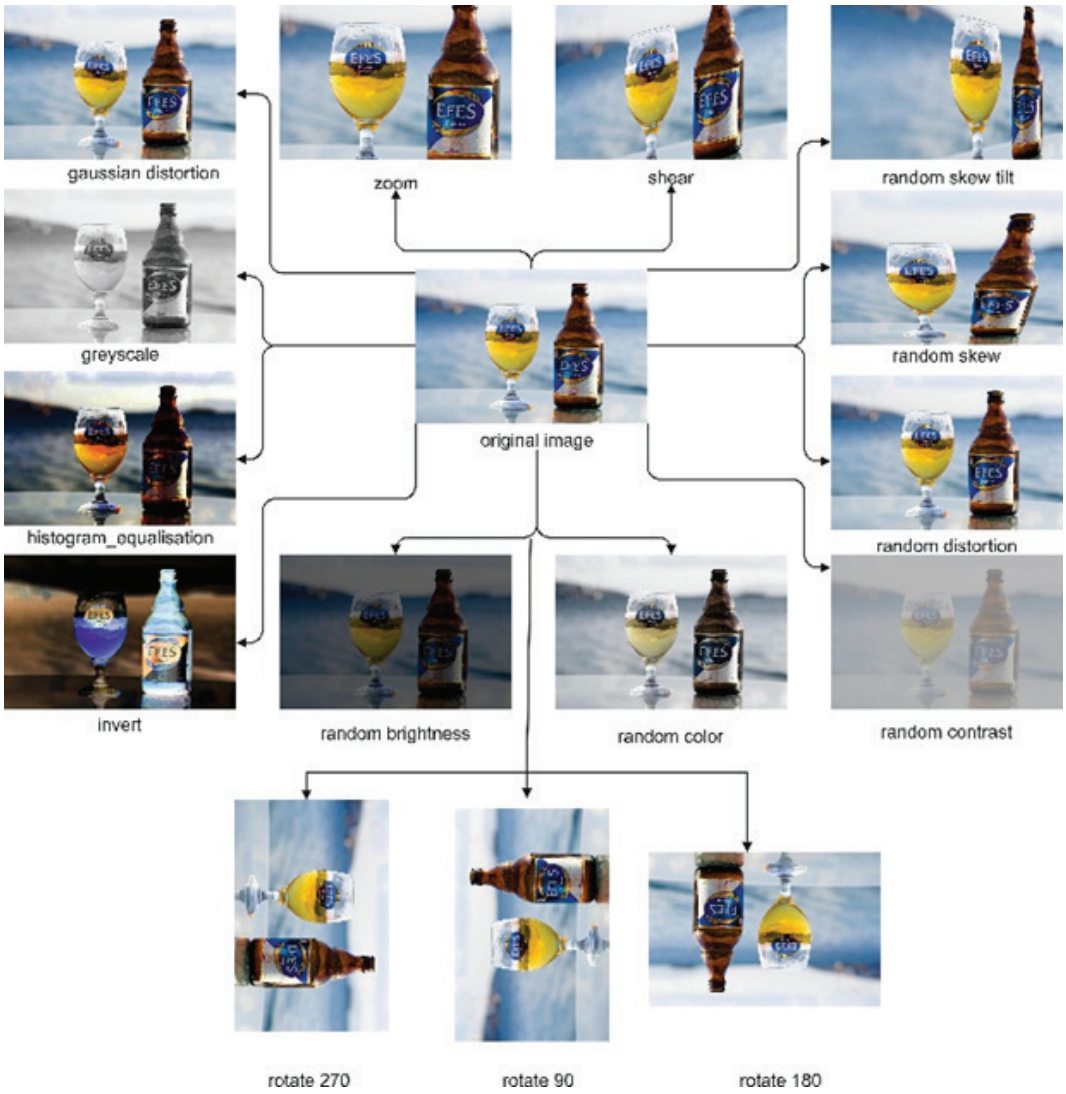


Figure 5. Data Augmentation Examples.

Table 1. The processes and parameters applied from the augmentor library

Process	Parameters
gaussian_distortion	probability=0.9, grid_width=8, grid_height=5, magnitude=5, corner="bell", method="in", mex=0.5, mey=0.5, sdx=0.05, sdy=0.05)
greyscale	probability=0.9
histogram_equalisation	probability=0.9
invert	probability=0.9
random_brightness	probability=0.9, min_factor=0.2, max_factor=0.6
random_color	probability=0.9, min_factor=0.1, max_factor=0.8
random_contrast	probability=0.9, min_factor=0.1, max_factor=0.8
random_distortion	probability=0.9, grid_width=5, grid_height=6, magnitude=4
shear	probability=0.9, max_shear_left=15, max_shear_right=15
skew	probability=0.9, magnitude=1
skew_tilt	probability=0.9, magnitude=1
zoom	probability=0.9, min_factor=1.1, max_factor=1.8

Table 2. The number of data augmentation processes and total amount of data for data sets for db_01, db_02, and db_03

	db_t	i_s	S_v	r_n	*Total Data*
db_01	279643	12	10000	3	1598572
db_02	305254	12	70000	3	4581016
db_03	311370	12	70000	3	4605480

Figure 6. Sectional ratios of data set's training, test, and validation sections.

Table 3. Data sets and data amounts of test, train, and validation

Data Set	*Total Data*	*train*	*test*	*validation*
db_00	80000	48000	16000	16000
db_01	1598572	959144	319714	319714
db_02	4581016	2748610	916203	916203
db_03	4605480	2763288	921096	921096

the data set has been multiplied using the procedures given in Table 1. In this way, the data set has been augmented.

In Equation (2), dbt represents the number of individual data, is represents the applied processes, sv represents the amount of synthetic data produced, and rn represents the number of rotations per data.

$$Total\ Data = (db_t + i_s.s_v)(1 + r_n) \quad (2)$$

Table 2. shows the evaluation of data augmentation process in Augmentor library and in terms of the amount of data produced. db_00 data set, which was used for parameter selection, was excluded from the data augmentation process.

The specified amount of sv gives the number of data randomly put into process from the individual data set per process. At the end of each process, a data as many as the amount of sv are produced. Since it was refrained from falling into redundant data repetition, the researchers tried to limit this number. 90, 180, and 270-degree rotation processes were performed separately on the data obtained at the end of the overall process.

Since CNN's were used in training, the data set was arranged accordingly. Each data set was divided into sections as train, test, and validation. Each section was divided into two as to its "alkol_var" (alcohol_yes) or "alkol_yok" (alcohol_no) status. As shown in Figure 6, the data set was arranged as 20% test, 60% train, and 20% validation.

Lightning Memory-Mapped Database (Lmdb) structure was used when creating the database. Table 3 shows the total data of these data sets, data amounts of test, train and validation.

Training Environment

Three different systems were used for training and application. At the beginning of the training process and while determining the hyper-parameters, the first system with an NVIDIA GeForce 1050Ti display card was used. In the following process, since this equipment was no longer sufficient due to increasing size of the data set, another system with an Nvidia GeForce 1080Ti display card was used. The firewall was applied on a laptop with Ubuntu 20.04 operating system, Intel Core i3 5505U processor, and Intel Graphics 5500 display card.

The training was performed by using Nvidia Digits software interface. Since the training status was visualized through Digits, the training status could be instantly tracked through visual graphics on a regular basis, and the training process could be ceased at the desired level when an undesirable result was observed.

CAFFE, which is an open-source deep learning application platform developed by Berkeley University, was preferred as the training platform. CAFFE also supports other deep learning architectures with different structures, such as image classification and segmentation [33].

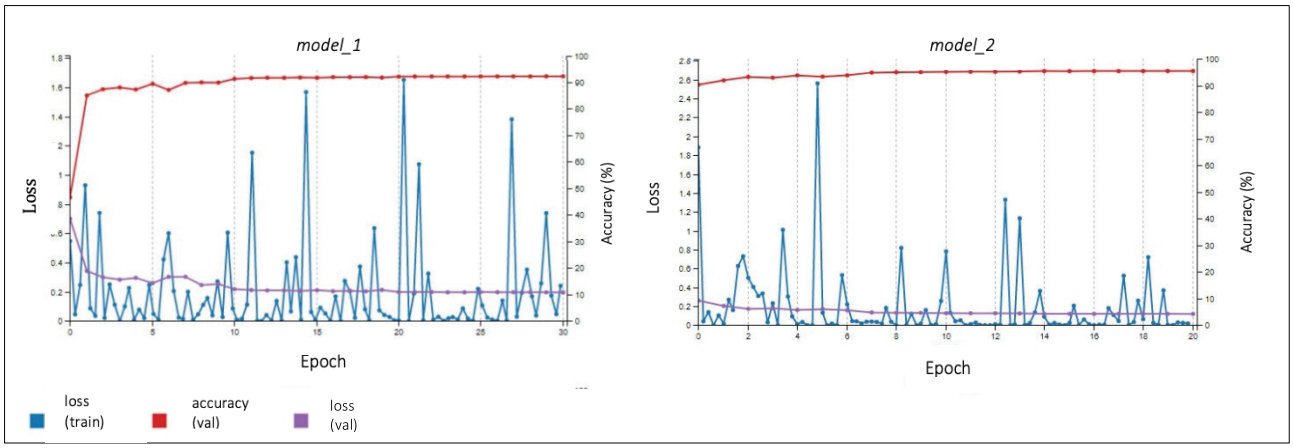


Figure 7. Model_1 and Model_2 training graphic.

Training

After determining the hyper-parameters, the training process was approached in three stages. The data set was trained for 8 days and 21 hours with db_01, model_1, 30 epoch by using NVIDIA GeForce 1050Ti display card. The data set was expanded with data, which were detected to be incorrect according to the model application test result, and db_02 was created. The model_2 was trained for 8 days and 21 hours by taking the model_1 training weights as the initial value with 20 epoch by using NVIDIA GeForce 1050Ti display card. Db_03 was created with incorrect estimate data according to the test result of model_2, and model_3 was trained in 30 epoch within 2 days and 17 hours with NVIDIA GeForce 1080Ti. Model_3 was used in firewall application.

Training graphics of model_1 and model_2 are given in Figure 7. The data sets were improved according to their error statuses, and the data set was expanded and retrained as model_3. The training graphic is given in Figure 8. At the end of 30th epoch, accuracy value was found 97.6469, loss (train) was found 0.00384845, and loss value was found 0.0697716.

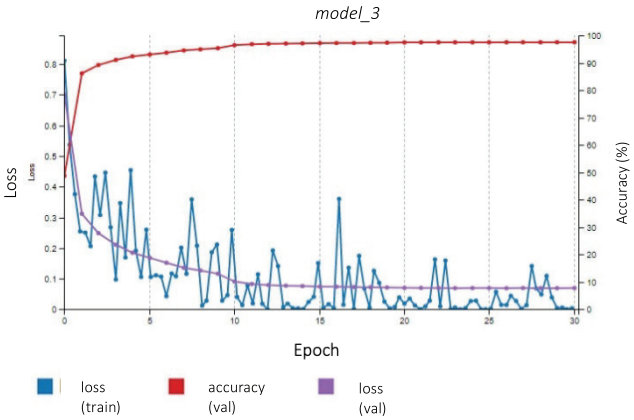


Figure 8. Model_3 training graphic

Testing of the Model

Digits interface was used in singular image tests, which is the first stage of the model. The output images of the layers, such as convolution, pooling, and normalization, can be observed separately through the digits interface. Figure 9 shows the screen shots of the results of the "alkol_var" and "alkol_yok" query on the singular image. The Images tagged (a), (b), (c), (d), (e), and (f) in Figure 9 were compared with cases including and not including alcoholic drinks in the same background. The estimation rates were found to be quite successful at the end of this comparison.

The Model's Confusion Matrix Evaluation

It is not easy to understand to the extent which the accuracy and error values obtained during model training coincide with the real values merely by reviewing the training results. Therefore, the confusion matrix rates of accurate and false classification estimation values were compared with the data that are not included within the training set. The estimation statuses of the model trained for this process were classified as true positive, false positive, true negative, and false negative as given in Figure 10.

Since there are two groups as "alcohol_yes" and "alcohol_no" in model classification, the classification values were determined in accordance with the following descriptions:

- TP (true positive): The data that include alcoholic drinks and classified as "alcohol_yes".
- FN (false negative): The data that include alcoholic drinks, however classified as "alcohol_no".
- FP (false positive): The data that do not include alcoholic drinks, however classified as "alcohol_yes".
- TN (true negative): The data that do not include alcoholic drinks and classified as "alcohol_no".

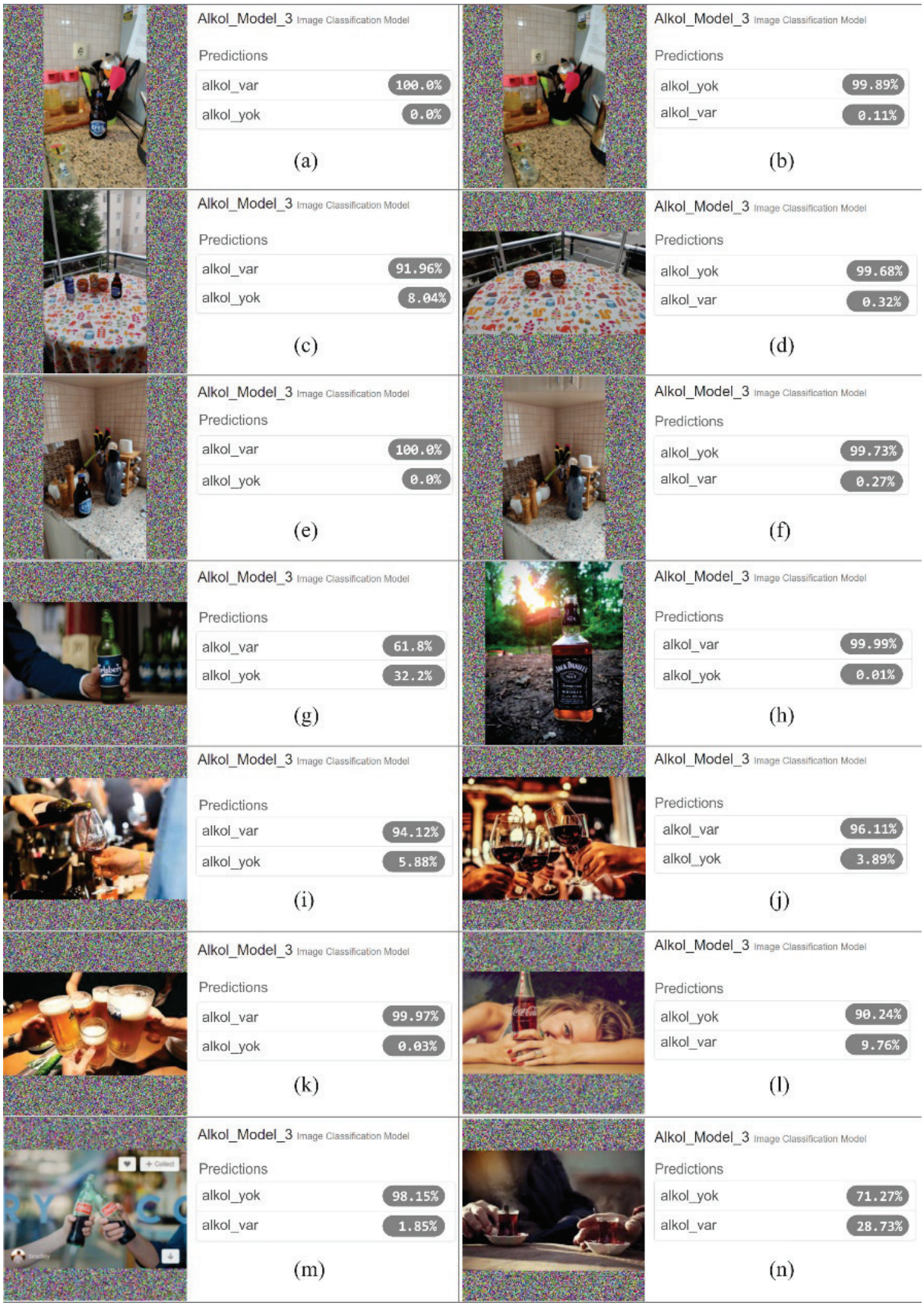


Figure 9. Examples of alcoholic drink detections through Digits interface (alkol_yok as alcohol_no & alkol_var as alcohol_yes).

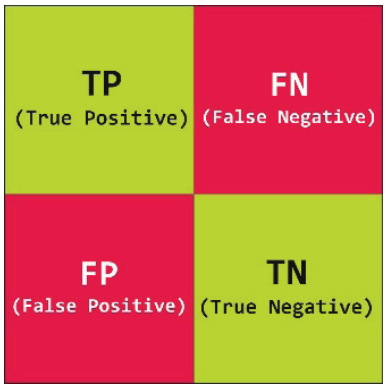


Figure 10. Confusion Matrix Evaluation metrics.

Table 4. Confusion matrix evaluation metrics formulas

Metric	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{TN+FP}$
F1 score	$2.\frac{\text{Precision}\,.\,\text{Recall}}{\text{Precision}+\text{Recall}}$

Figure 11. TP, FP, TN, and FN data score of seven different tests performed for confusion matrix calculation.

Table 5. Average values of precision, recall, accuracy, specificity and f1 score of seven different tests

Metrics	Average Value
Precision	0.975476359
Recall	0.797451923
Accuracy	0.940065113
Specificity	0.992612997
F1	0.877526188

In consideration of these data, the calculation metrics were evaluated with accuracy, precision, recall, specificity, and f1 score titles given in Table 4.

The test data to be used in confusion matrix calculation consist of data that were not included in training data set as training, verification, or test data. The classification process was completed by using the model trained with 154501 singular images obtained through the Internet. The images used here can actually be seen in websites in real life and they were blended for diversity. Table 4 shows the evaluation metrics for confusion matrix [34].

This test data set is performed with seven different tests and with approximately 22000 images per test. Figure 11 shows the TP, FP, TN, FN values obtained at the end these seven different tests. Table 5 shows the confusion matrix values obtained at the end of the test.

Firewall Application

The intended filtering structure is expected to detect and evaluate the images. The network traffic should be constantly checked in order to analyze whether the images contain harmful contents. Therefore, it has become necessary to create a proxy server. The entire traffic is transmitted through the proxy server; the proxy server directs the images that it detected in the network traffic towards CAFFE model file. It decides whether the image can be viewed or not according to the classification information coming from CAFFE model file. The application block scheme is given in Figure 12.

Mitmproxy tool was used for proxy server structure. Mitmproxy is a cluster of tools that provide an SSL/TLS featured observing, changing, and blocking proxy interactive for http/1, http/2, and websockets [35]. The entire flow is kept in a buffer with mitmproxy. It enables the users to process these data later, if required. The biggest advantage of mitmproxy tool is that it allows for interference with the network traffic through python command sequences [36].

In an ordinary http connection, the client makes a request to the proxy with "GET" method and the proxy conveys this to the server. In an open http connection, the proxy cannot read or impact the encrypted TLS data flow. On this connection, the proxy is nothing more than a facilitator. Thus, the information about the contents of the data flowing through this connection cannot be obtained. After the connection agreement with TLS is made through this line, all requests and responses become transparent

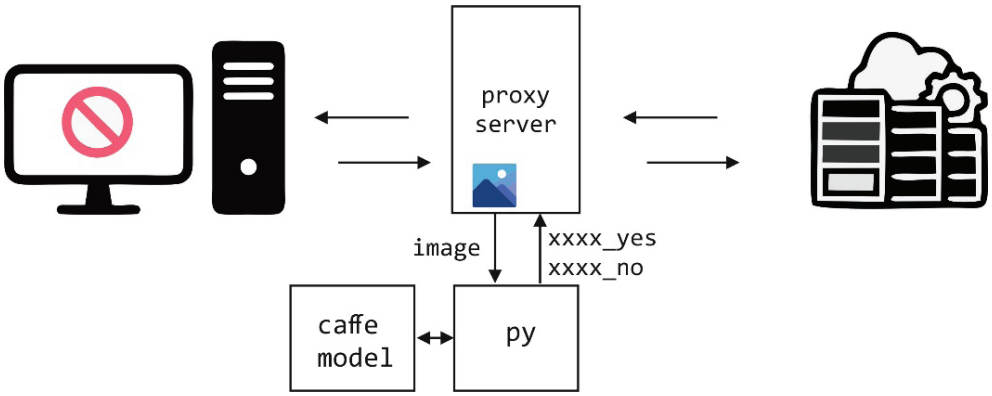


Figure 12. Application block scheme.

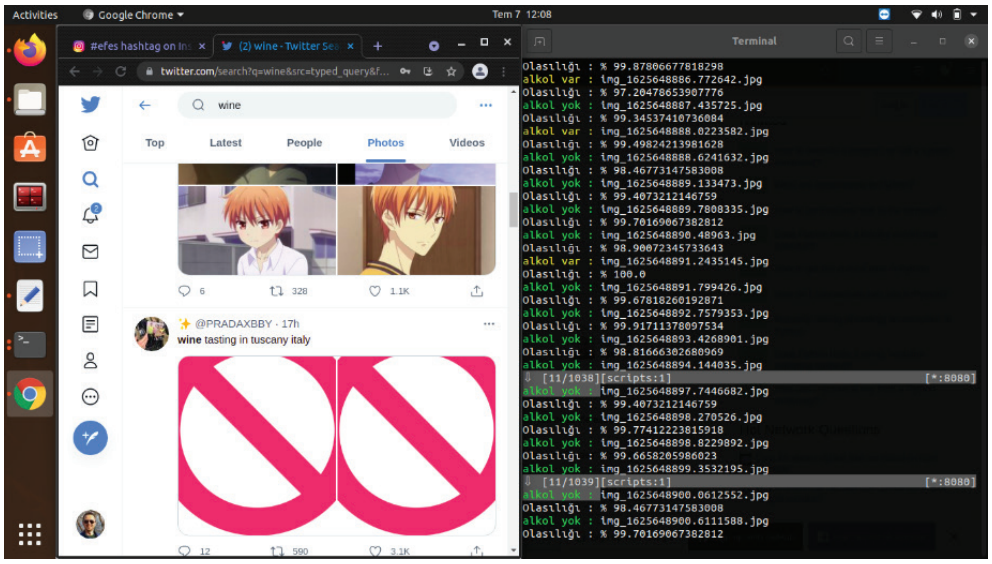


Figure 13. Screen shots of web contents with an active firewall and the terminal output.

for the proxy. Mitmproxy is located in the middle of this transparent connection flow. In decryption stage, it acts as if the server is the client, and the client is the server. For problems that may arise from certificate authority, it uses its own certificate authority and creates its own shear certificates.

First, the client establishes connection via mitmproxy and requests an "HTTP CONNECT". Secondly, the mitmproxy gives "200 Connection Established" response as if it established a connection with the client, and then it convinces the client that it is communicating with the remote server. In order to indicate the name of the main computer that in connected, it initiates a TLS connection by using SNI. Thus, the mitmproxy connects to the server and establishes a TLS connection by using the SNI main computer name specified by the client. The server gives a response to the matching certificate including CN and SAN values. After this step, mitmproxy Creates its certificate for the connection paused at the 3rd step and TLS agreement continues. The communication occurs through transmission of the requests and responses through the connection between the client and the server mediated by mitmproxy [35].

Figure 13 shows the screen shot of the integrated use of the model with mitmproxy. When the user sends a request to a website, the images sent through the network are blocked according to the estimation result of the model. In order to find out if these images are blocked, a block icon image was selected to be displayed instead of the blocked image.

Figure 14 shows the results of the screen shots of the images taken when the firewall was active and inactive according to the search results regarding alcoholic drinks performed on Twitter and Instagram.

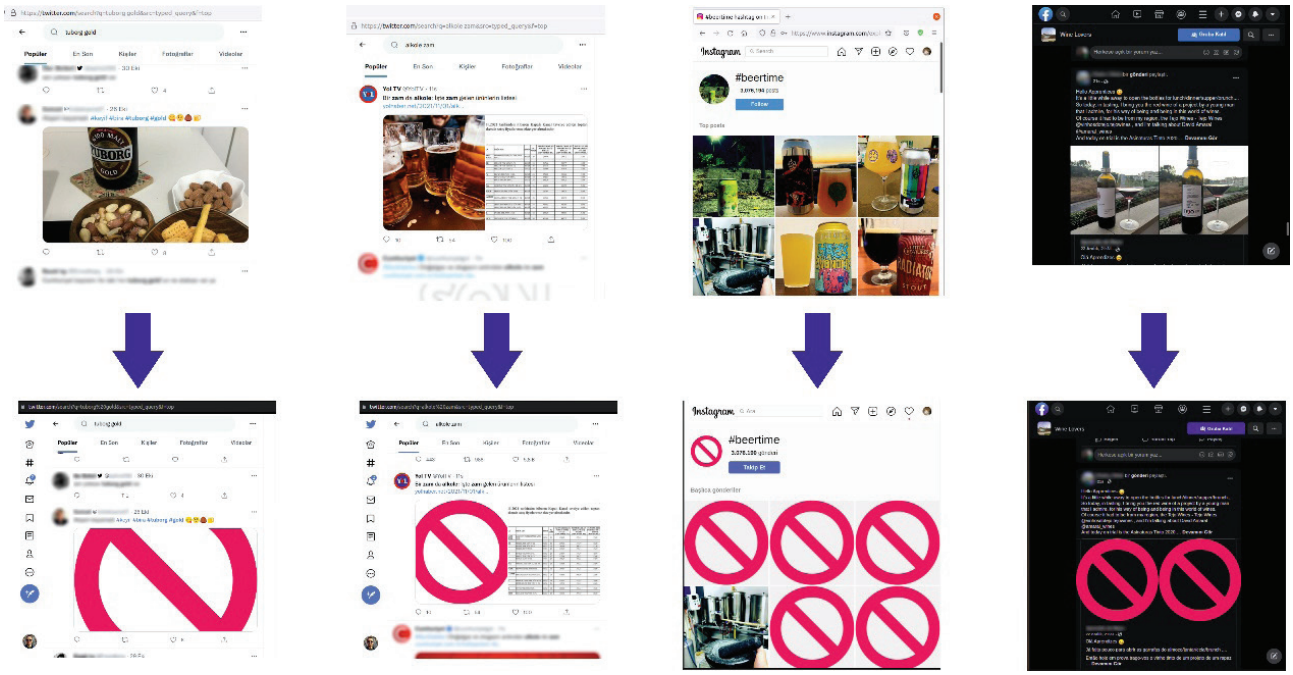


Figure 14. Screen shots of the pages viewed on the web browser when the firewall was active and inactive.

RESULTS AND DISCUSSION

A smart filtering mechanism is presented by performing image content analysis on the suggested application firewall model. The user can filter potentially harmful images by using this smart firewall. In this application model, only the alcoholic drinks were considered as harmful images, and a model was trained accordingly; the blocking process was completed successfully. However, various models can be included within the same mechanism with suitable data sets, such as weapons, violence, and 18+ adult contents. Each filtering rule can be activated or deactivated separately, when necessary.

The size of the data set, in terms of the type of the problem, is of capital importance for obtaining successful results from the training. The highest level of success was achieved only with 4.6 million image data with the means at hand. The model did not bring 100% success even in this way. However, considering the challenges of the problem, it could be regarded as a significant success. However, in order to ensure continuous success in estimations of the model, the data set may have to be updated periodically, and the training may have to be repeated with changing alcoholic or nonalcoholic drinks.

In this suggested firewall model, the image uploading speed slows down remarkably particularly because of image classification process. Graphics processor support is crucial for image classification. Setup can be performed on the proxy server through the network, however, higher process capacities for the server are recommended for efficient use. The effect of firewall on connection speed may decrease significantly with the future enhancements and developments in hardware technologies. Thus, more efficient ways for protection against harmful contents of the Internet may be developed.

Besides a firewall created through a proxy server, the trained model file can also be used in a browser add-on. In addition, a smart filtering and editing mechanism can also be used for video images to filter potentially harmful contents.

CONCLUSION

The suggested application development process was performed at six stages: collecting images for the data set, creating a data set with these images, determining and enhancing the hyper-parameters, training, success analysis through confusion matrix, and firewall application.

The present deep learning model acts as the main body for the firewall to filter the harmful contents. Alcoholic drinks were selected as the filtering subject. On the off chance of reiteration of each image selected for the data set, they were compared through means square function, and the images containing differences less than 5% were eliminated from the data set. For a successful training, the data set containing individual images have gone through duplication process, and the third data set contained a total of more than 4.6 million images.

Digits software was selected for the training. The training was performed by step-by-step enhancement with Adam optimizer algorithm by using GoogLeNet architecture in CAFFE environment. The training took 783 hours in total.

At the end of this training, 97.6469% of accuracy was achieved. It was observed that the ideal training epoch number was between 20 and 30. It was also observed that higher number of cycles had no significant impact on the accuracy result of the training.

154501 individual images were used as test data in calculation of confusion matrix. This data used for the test was performed through seven different tests with an average of 20000 images per test. These values obtained through approximately 20000 data remained constant with minor deviations. The precision was found as follows when calculated over the total values: 0.975476359, recall: 0.797451923, accuracy: 0.940065113, specificity: 0.992612997, and f1 score: 0.877526188. In coherence with the confusion matrix calculation results, the test applications also gave quite successful results.

A proxy server was required in order to control the network traffic and read the contents. Mitmproxy proxy server, which ensures flexible intervention with the network traffic with Python commands, was preferred. Thanks to mitmproxy, TLS encrypted data flow can be read, and intervened with python commands. Each image detected within the data flow is classified first through the interface created with the python commands, and CAFFE model file. At the end of this classification, an image is either blocked or not depending on whether it is harmful or not.

Tests were performed on many websites, including social media platforms, such as Facebook, Instagram, and Twitter.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 1943;5:115–133. [CrossRef]

[2] Demir I, Karaboğa HA. Modeling mathematics achievement with deep learning methods. Sigma J Eng Nat Sci 2021;39:33–40. [CrossRef]

[3] Rajkovic KM, Avramovic JM, Milic PS, Stamenkovic OS. Optimization of ultrasound-assisted base-catalyzed methanolysis of sunflower oil using response surface and artificial neural network methodologies. Chem Eng J 2013;215:82–89. [CrossRef]

[4] Zettler AH, Poisel R, Reichl I, Stadler G. Pressure Sensitive Grouting (PSG) using an artificial neural network combined with fuzzy logic. Int J Rock Mech Min Sci 1997;34:358. [CrossRef]

[5] Ma F, Sun T, Liu L, Jing H. Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network. Future Gener Comput Syst 2020;111:17–26. [CrossRef]

[6] Sabilla SI, Sarno R, Siswantoro J. Estimating gas concentration using artificial neural network for electronic nose. Procedia Comput Sci 2017;124:181–188. [CrossRef]

[7] Esen H, Esen M, Ozsolak O. Modelling and experimental performance analysis of solar-assisted ground source heat pump system. J Exp Theor Artif Intell 2017;29:1–17. [CrossRef]

[8] Esen H, Inalli M, Sengur A, Esen M. Predicting performance of a ground-source heat pump system using fuzzy weighted pre-processing-based ANFIS. Build Environ 2008;43:2178–2187. [CrossRef]

[9] Efe E, Alganci U. Determination of land cover change with multi-temporal Sentinel 2 satellite images and machine learning-based algorithms. Geomatik Derg 2023;8:27–34. [Turkish] [CrossRef]

[10] We are social. Special report - Digital 2021. Your ultimate guide to the evolving digital world. Available at: https://wearesocial.com/digital-2021. Accessed on Jul 2, 2024.

[11] Yaraş E, Yetkin Özbük RY, Çorlu P. Emmy ödüllü dizilerde alkol ve sigara ürün yerleştirme uygulamalarının içerik analizi yöntemi ile incelenmesi. Kastamonu Univ İktis İdar Bil Fak Derg 2018;20:67–84.

[12] İplikçi HG, Batu M. Digital communication and children: A content analysis of advertisements on the websites for children in Turkey. J Akdeniz İletiş 2018;29:242–256. Turkish.

[13] Uzun R. The Protection of children from media content and in media content: A study of ethical codes for children in media. J Akdeniz İletiş 2014;22:152–167.

[14] Kanbur BN. The effects of visual media and subliminal messages on child health. İstanbul Gelişim Univ Sağ Bil Derg 2020;10:94–106. [Turkish] [CrossRef]

[15] Berners-Lee CM. Cybernetics and forecasting. Nature 1968;219:202–203. [CrossRef]

[16] Lauriola I, Lavelli A, Aiolli F. An introduction to deep learning in natural language processing: Models, techniques, and tools. Neurocomput 2022;470:443–456. [CrossRef]

[17] Gupta A, Anpalagan A, Guan L, Khwaja AS. Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues. Array 2021;10:100057. [CrossRef]
[18] Fahad S, Ranjan A, Yadav J, Deepak A. A survey of speech emotion recognition in natural environment. Digit Signal Process 2021;110:102951. [CrossRef]
[19] Possemiers A, Lee I. Evaluating deep learned voice compression for use in video games. Expert Syst Appl 2021;181:115180. [CrossRef]
[20] Du X, Cai Y, Wang S, Zhang L. Owerview of deep learning. In proceedings of the 31st Youth Academic Annual Conference of Chinese Association of Automation; 2016 Nov 11–13; Wuhan, China. IEEE; 2016. pp. 159–64. [CrossRef]
[21] Pathak AR, Pandey M, Rautaray S. Application of deep learning for object detection. Procedia Comput Sci 2018;132:1706–1717. [CrossRef]
[22] Hu H, Pang L, Shi Z. Image matting in the perception granular deep learning. Knowl Based Syst 2016;102:51–63. [CrossRef]
[23] Tiken C. Deep learning applications. Master's thesis. Istanbul: Istanbul Univ; 2015.
[24] Dong S, Wang P, Abbas K. A survey on deep learning and its applications. Comput Sci Rev 2021;40:100379. [CrossRef]
[25] Yiğit ÖE, Alp S, Öz E. Prediction of bist price indices: A comparative study between traditional and deep learning methods. Sigma J Eng Nat Sci 2020;38:1693–1704.
[26] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE 1998;86:2278–2324. [CrossRef]
[27] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2017;60:84–90. [CrossRef]
[28] Wikipedia. List of alcoholic drinks. Available at: https://en.wikipedia.org/wiki/List_of_alcoholic_drinks. Accessed Jul 2, 2024.
[29] Sarkar N. Mean square error matrix comparison of some estimators in linear regressions with multicollinearity. Stat Probabil Lett 1996;30:133–138. [CrossRef]
[30] Taylor L, Nitschke G. Improving deep learning with generic data augmentation. In proceedings of the 2018 IEEE Symposium Series on Computational Intelligence (SSCI); 2017 Nov 18–21; Bangalore, India. IEEE; 2018. p. 1542–2547. [CrossRef]
[31] Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data 2019;6:60. [CrossRef]
[32] Augmentor. Available at: https://augmentor.readthedocs.io/en/master/. Accessed on Jul 2, 2024.
[33] Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, et al. Convolutional architecture for fast feature embedding. In proceedings of the 22nd ACM International Conference on Multimedia; 2015 Jun 18–19; California, USA. ACM; 2015. pp. 675–8.
[34] Amidi A, Amidi S. Machine Learning tips and tricks cheatsheet. Available at: https://stanford.edu/~shervine/teaching/cs-229/cheatsheet-machine-learning-tips-and-tricks. Accessed on Jul 2, 2024.
[35] Mitmproxy Docs. Mitmproxy. Available at: https://docs.mitmproxy.org/stable/. Accessed Jul 2, 2024.
[36] Wang Y, Xu G, Liu X, Mao W, Si C, Pedrycz W, et al. Identifying vulnerabilities of SSL/TLS certificate verification in Android apps with static and dynamic analysis. J Syst Softw 2020;167:110609. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2023.00122

Research Article

The effect of marble waste in the production of low-temperature porous material from alkali-activated fly ash

Hakan CENGİZLER[1], Muhterem KOÇ[2,*]

[1]*Department of Materials and Materials Processing Technologies, Vocational School, Manisa Celal Bayar University, Turgutlu 45410, Türkiye*
[2]*Department of Industrial Design, Kütahya Dumlupınar University, Kütahya, 43100, Türkiye*

ARTICLE INFO

Article history
Received: 09 January 2023
Revised: 16 February 2023
Accepted: 04 April 2023

Keywords:
Alkali Activation; Fly Ash; Marble Waste; Open-Pore; Sintering

ABSTRACT

The production of low-cost open-pore ceramic materials from fly ash (FA) and marble waste (MW) was investigated. The effect of MW (5-40 wt.%) on the open porosity was determined. To reduce the sintering temperature and improve the properties of porous materials, the mixtures were activated with an alkali solution. Samples pressed from FA and FA+MW mixtures were sintered at low temperature (900 °C), but sufficient strength could not be obtained. However, when these mixtures were subjected to alkali activation, pressed and sintered at 900 °C, sufficient strength and porosity values were reached. The open porosity of the MW neat specimen was 12.70%, but it increased up to 39.91% at 40 wt.% MW, which was the highest ratio used in the literature. The main phase structure was nepheline at 0-20 wt.% MW, but gehlenite became the dominant phase at 40 wt. % MW. The compressive and flexural strength values of 40 wt.% MW added specimen was determined to be 12 and 5.35 MPa, respectively. The open-pore ceramic of high MW ratio, produced by this new alternative route, has the potential for use in water purification membranes for macro filtration purposes.

Cite this article as: Cengizler H, Koç M. The effect of marble waste in the production of low-temperature porous material from alkali-activated fly ash. Sigma J Eng Nat Sci 2024;42(4):1148–1159.

INTRODUCTION

Porous ceramics have beneficial properties such as high mechanical, thermal and chemical stability, long lifetime, high permeability, ease of cleaning with many other advantages [1–3]. The microstructure of porous ceramics can be tailored and have a slight polluting effect on the environment [4] along with properties such as less susceptibility to microbial attacks and biological degradations [5]. All these superior properties, in contrast to those of polymeric porous materials, have found applications in engineering practices such as membranes in filtration processes [4,5], catalysis applications [6], thermal isolation coatings [7], and porous bricks [8].

Because of the high costs of precursor materials used to manufacture open-pore ceramics [9,10], natural clays such as kaolin and ball clay, feldspar and quartz sand as well as

*Corresponding author.
*E-mail address: muhterem.koc@dpu.edu.tr

This paper was recommended for publication in revised form by Editor in-Chief Ahmet Selim Dalkilic

industrial waste substances were used as main raw materials. In this regard, research was especially directed towards producing new generation inorganic porous products at a low price manufactured from many low-cost waste materials causing environmental pollution. FA [3,4], steel slag [10], glass waste [6], screen and monitor glass waste [11], sewage sludge [12], paper sludge [13], biological waste of water purification [14], coal extraction waste [15], sandblasting waste [16], wasted diatomaceous earths [11,16], coffee waste [17], and red mud [18] were a few of them. FA is a waste generated in large amounts during coal combustion in thermal power plants. Any attempt to recycle or re-utilize FA is crucial for the environment. FA, with the addition of pore-forming materials, were sintered and porous ceramic structures were obtained [3,10]. Additionally, it was found that geopolymerization of FA before sintering had a positive effect on the material properties [19,20]. Besides, geopolymerization made it possible to produce materials at lower temperatures [21]. Previous studies showed that the physical and mechanical properties of FA-based geopolymers exposed to elevated temperatures improved [20]. It was stated that FA-based geopolymers had large numbers of small pores, which facilitated the escape of moisture when exposed to elevated temperatures, thus causing minimal damage to the geopolymer [19]. Sintering resulted in densification and development of new phases in the structure [19] and improved the properties of geopolymers [19,20]. Furthermore, densification and formation of new phases were reported at and above 900°C [19,20,22]. In conclusion, exposure to high temperature resulted in sintering processes, structural rearrangement and new crystalline phases contributing to strength [23].

Previous work also studied the performance of porous geopolymers made with FA after exposure to high temperatures. Abdullah et al. [24] prepared geopolymer paste samples by alkali activation of FA with a mixture of sodium hydroxide and sodium silicate solution. They cured the geopolymer paste samples at 60 °C for 24 h and sintered them at the range of 600-1000°C. In a very recent study, Sawan et al. [25] investigated the in-situ formation of geopolymer foams from metakaolin. They used an alkali activator solution of sodium silicate and sodium hydroxide to prepare the initial geopolymer pastes. Afterwards, they sintered the samples at 800, 1000 and 1200°C. In the studies mentioned above [24,25], they produced porous geopolymer materials first and then subjected them to the sintering process. Thus, they investigated the effect of sintering on geopolymer materials. Geopolymer technology enables the production of high-performance ceramics at a lower temperature compared to those produced by conventional sintering [21]. However, geopolymer materials produced from FA do not have enough open porosity before and after sintering. Therefore, additives are needed to increase porosity. Carbonates decompose well below 1000 °C releasing CO_2 gas. Therefore, calcite, dolomite and soda ash containing carbonates in their structure were studied extensively as pore making agents. The recent work on porous ceramic production was also focused on exploring novel pore forming additives such as activated carbon, natural phosphates, wood sawdust, and raw materials containing carbonates, etc. [9]. In this regard, the use of MW as a carbonate source should be investigated instead of the pore making carbonate sources mentioned above. Thus, the damage to the environment will be mitigated and the production cost of porous ceramics will be reduced.

The present study investigated the utilization of MW as a pore-making agent in the production of open-pore ceramic from FA. The experimental compositions were activated with an alkali solution of Na_2SiO_3 and NaOH, which also have a fluxing effect, to reduce sintering temperature. The effect of adding MW in different ratios (5-40 wt.%) into FA on open porosity, new phase formation, mechanical and physical properties was investigated. The utilization of these industrial wastes as main raw materials to manufacture porous ceramics would be beneficial with effective solid waste management and reduced production costs.

MATERIALS AND METHODS

Materials

The Tunçbilek FA was collected from Tunçbilek thermal power plant (Kütahya, Turkey). A marble processing plant (Karaburun, Turkey) provided MW. Chemical grade NaOH and Na_2SiO_3 were purchased from Merck Chemicals GmbH.

Processing and Characterization

Five specimens of different geopolymer compositions were prepared (Table 1). The first one consisted of 100 wt. % Tunçbilek FA and was coded as MW0. The other five specimens were prepared by gradually increasing MW ratios of 5, 10, 20, and 40 wt. % coded as MW5, MW10, MW20, and MW40, respectively. The numbers following the denotation show wt. % of MW.

Table 1. Geopolymer compositions and codes

Sample code	MW0	MW5	MW10	MW20	MW40
FA (g)	100	95	90	80	60
MW (g)	-	5	10	20	40
NaOH (g)	20	20	20	20	20
Na_2SiO_3(g)	20	20	20	20	20

Firstly, the Tunçbilek FA and the MW, totally 100 grams, were homogeneously dry mixed according to the predetermined ratios (Table 1). Previously prepared 12M alkali solution of NaOH and Na_2SiO_3 was added (Table 1) and mixed for 5 minutes using a laboratory mixer. Homogeneously mixed compositions were shaped into specimens of 10x10x55 mm dimensions under 50 MPa pressure using a hydraulic press. Subsequently, they were dried at 105°C for 24 hours in a laboratory drier. Curing and drying processes were combined and carried out simultaneously in a single stage of 24 hours, which was cost-effective and time-saving. After completing the drying stage, the alkali-activated specimens were sintered at 900 °C for 30 minutes with a ramping rate of 5°C/min. Finally, the sintered specimens were characterized.

The chemical analyses of the Tunçbilek FA and the MW were carried out by X-ray fluorescence (XRF) (Panalytical Axios). Phase identification was conducted by X-ray diffraction (XRD) on raw materials and porous ceramic specimens using a Rigaku Miniflex powder diffractometer with Cu Kα radiation (λ=1.5418 Å). The measuring rate was 2°/min at 40 kW and 30 mA in the 15–70° 2θ range. Crystallographic databases used for the phase identifications were Inorganic Crystal Structure Database (ICSD) and International Centre for Diffraction Data (ICDD) PDF-2. The thermal behaviour of the Tunçbilek FA and the MW was investigated using a fully computer-controlled DTA/TG (Netzsch STA 449F3). The measurement, for the MW, was conducted between room temperature and 1100°C. It was conducted, for the Tunçbilek FA, between room temperature and 1300°C. The heating rate was 10°C/min in the air during both measurements. The particle size distribution of the Tunçbilek FA and the MW was determined using a laser size analyzer (Malvern). Scanning electron microscopy (SEM) was conducted on a Nova NanoSEM 650 scanning electron microscope to investigate the microstructure of the waste materials (FA and MW) and the sintered geopolymer specimens on the fractured surfaces. Polished cross-sections were produced by mounting sample fragments in low viscosity epoxy resin and polishing to a 1 μm finish. Samples were coated with a thin layer of Au-Pt before SEM imaging.

Bulk density, apparent porosity (open porosity), water absorption [26], and compressive strength [27] were determined according to the related standards. The compressive strength tests were carried out using a hydraulic press (SACMI 470, PIL type) with a speed of 0.5 mm/min. Five specimens were tested for each composition and the average values of the compressive strength were calculated. The flexural strength (three-point bending strength) was measured using Shimatsu 250 kN Model equipment. Again, the average values of the flexural strength were calculated from the results of five specimens tested for each composition. The linear shrinkage values of the sintered geopolymer specimens were measured by subtracting the fired length from the original green length [28].

The bulk density was measured according to Archimedes' principle. The powder density was calculated by a pycnometer. The open porosity (P) was calculated from the dry, soaked (in water) and suspended (in water) weights of a specimen using equation 1:

$$\%P: \frac{(Soaked\ Weight - Dry\ Weight)}{(Soaked\ Weight - Suspended\ Weight)} \quad (1)$$

Water absorption value represents the open porosity. It was calculated by the difference in specimen weight under over-dried and fully saturated (in water) conditions [26].

RESULTS AND DISCUSSION

Characterization of the Raw Materials

In Table 2, the chemical composition of the Tunçbilek FA used as the main material and the MW, as a pore-forming agent, was given. The sum of SiO_2, Al_2O_3 and Fe_2O_3 was 87.44 wt. % showing that the Tunçbilek FA was F class [29]. The Tunçbilek FA also contained MgO (4.48 wt. %) and CaO (1.64 wt. %) along with minor quantities of K_2O, Na_2O, and SO_3. LOI value of the Tunçbilek FA was 4.08 wt. %. The MW contained 61.65 wt. % CaO and a low amount of SiO_2 (2.20 wt. %), MgO (0.69 wt. %), Al_2O_3 (0.19 wt. %), and Fe_2O_3 (0.06 wt. %). The LOI value of the MW was 35.08 wt. %, which correlates well with the reported values [30,31].

When looked at the phase analysis, quartz, haematite and mullite were detected to be the main crystalline phases in the Tunçbilek FA (Figure 1). The origin of quartz (primary quartz) in the Tunçbilek FA was the source coal and the secondary quartz formed during combustion [32]. The mullite was the product of the solid-state reaction of decomposed clays [32] and/or occurred through the crystallization of the aluminosilicate melt [33]. The existence of haematite, which generally exists in bituminous F class FA, was because of the thermal decomposition of clay minerals during combustion [34]. Calcite was the only dominant crystalline phase in the MW with a low amount of quartz phase (Figure 1) [35]. The fact that the MW contained a high amount of CaO (61.65 wt. %) (Table 2) enabled it to be a suitable material in its use as a pore former. It was also used as a cheap pore-forming agent because it was a waste material [31,32,35].

The thermal behaviour of the Tunçbilek FA was studied by DTA/TG analyses (Figure 2a). First, an endothermic peak on the DTA curve at approximately 160°C corresponded to the evaporation of moisture, that is, dehydration of adsorbed water mechanically bonded in the form of H_2O molecules on the surface of the ash particles [36,37]. In the interval between 100 and 450°C, the removal of hydrated water was completed [38]. Subsequently, an exothermic event at 678°C was associated with the combustion of unburned coal [39]. At higher temperatures between 800 and 1200°C, the decomposition of impurities from

Table 2. Chemical analyses of Tunçbilek FA and MW

Chemical composition (%)	FA (%)	MW (%)
SiO_2	58.26	2.20
Al_2O_3	17.73	0.19
Fe_2O_3	11.45	0.06
CaO	1.64	61.65
MgO	4.48	0.69
K_2O	0.66	-
Na_2O	0.35	-
SO_3	0.26	0.05
LOI*	4.08	35.08

*: Loss of ignition

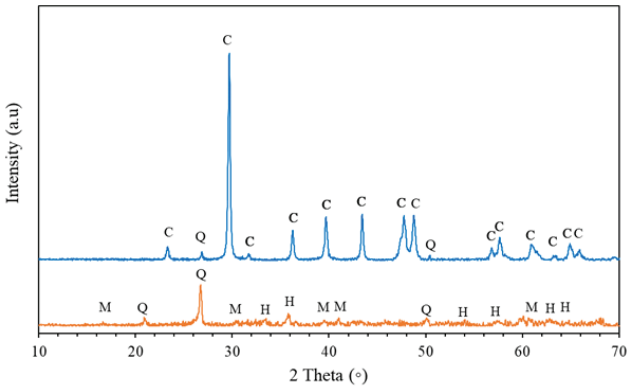


Figure 1. XRD spectrum of Tunçbilek FA (a) and MW (b). (C: calcite, Q: Quartz, M: Mullite, H: Hematite)

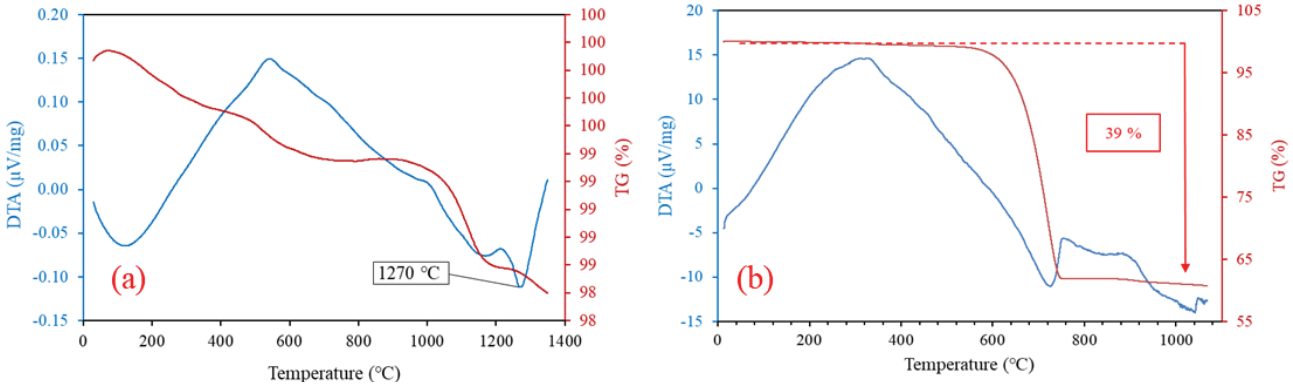


Figure 2. DTA/TG analysis of Tunçbilek FA (a) MW (b).

coal, carbonates, or sulphates took place [36,39]. The TG curve accordingly showed a continuous decrease in weight. The melting temperature of the Tunçbilek FA was 1220°C (Figure 2a).

In the DTA analysis of the MW (Figure 2b), a gradual mass loss of approximately 1.5 % was observed up to 575°C, which can be attributed to moisture loss [31,36]. A sharp endothermic reaction occurred between 620 and 760°C with a rapid downward move in the TG curve. This temperature indicated to the decomposition of calcite according to equation 2. Looking at the TG curve, the MW lost about 39% of its initial weight between 620 and 760°C [30,31]. It was considered that the high loss of ignition (LOI) (Figure 2b) (Table 2), would contribute to the formation of a porous structure in this study. Following equation 2 [30,31], it was presumed that the calcium carbonate broke down under heat and converted into CaO (solid) and CO_2 (gas), and open-pore structure formed because of the CO_2 gas released.

$$CaCO_{3(s)} \rightarrow CaO_{(s)} + CO_{2(g)} \quad (2)$$

The size distribution of the Tunçbilek FA (D_{10} 6.09 µm, D_{50} 32.5 µm and D_{90} 124 µm) ranged between 0.46 and 310 µm. However, the particle size distribution of the MW was smaller (D_{10} 2.11 µm, D_{50} 9.17 µm and D_{90} 42.14 µm) than that of the Tunçbilek FA in a close range between 0.46-66 µm (Fig. 3). Therefore, it was considered advantageous in that the pore dimensions are in a narrow range. The Tunçbilek FA and the MW were both used as received with no size-reduction.

The Tunçbilek FA mostly consisted of spherical particles so-called microspheres with some irregularly shaped angular particles (Figure 3). The MW had a microstructure consisted of irregular but dimensionally similar particles (Figure 3).

In the current experimental work, the production of open-pore ceramic from waste materials FA and MW was investigated. The alkali activation method was adopted to reduce the sintering temperature of green specimens to the lowest possible value. As seen from the DTA data in Figure 2b, the decomposition temperature of the MW was between 620 and 760°C [30,31]. The LOI value of the MW was 39 wt. % [30,31]. However, as mentioned earlier,

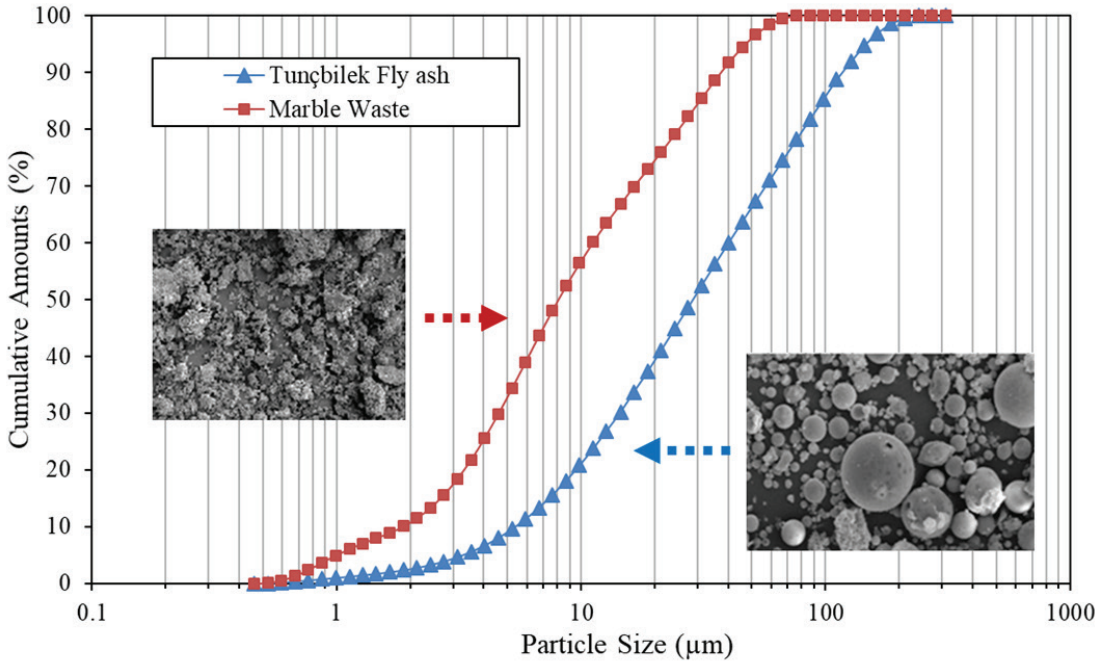


Figure 3. Particle size distribution of Tunçbilek FA and MW.

previous studies on sintering of geopolymers showed that densification and formation of new phases in the structure [23], which improved the properties of geopolymers [19,20], occurred at and above 900°C [19,20]. Therefore, in the present work, the alkali-activated specimens composed of the Tunçbilek FA and the MW were sintered at 900°C, to investigate the effect of sintering on forming an open-pore structure and contribution of new phase development to mechanical properties.

Physico-Mechanical Properties of Sintered Porous Geopolymers

FTIR analysis is principally based on the absorption of different wavelengths of infrared light and studies the composition and structure of material molecules. Besides, FTIR absorption spectroscopy is an effective tool for characterizing alkali-activated materials because of its well-known sensitivity for materials of short-range structural order [40]. Therefore, FTIR tests were conducted on the raw Tunçbilek FA and alkali-activated compositions to investigate the effect of alkali activation on the FA structural development. The results of the FTIR spectra were depicted in Figs.4a and 4b. In Figure 4a, the effect of alkali activation was determined by examining the raw and alkali-activated FA (MW0) specimen. The central band at around 1022 cm^{-1}, as the main feature of the FTIR spectra of raw FA, exhibited overlapped peaks associated with the asymmetric stretching vibrations of tetrahedral Si-O-Si or Si-O-Al bonds [41]. The intensity of this band is proportional to the reactivity of FA [42]. The peak at 1022 cm^{-1} shifted towards lower wavenumber of 983 cm^{-1} indicating the formation of geopolymer structure due to alkali activation [42,43]. The new peak observed at a lower wavenumber of 983 cm^{-1} on geopolymerization was directly linked with the transformation of Si-O-Si bonds of amorphous silica into Si-O-Al bonds of poly sialate, which meant the substitution of Si by tetrahedral Al and therefore the formation of the geopolymer structure of aluminosilicate network [44]. Another band, which appeared in the geopolymer specimen at around 1440 cm^{-1}, was absent in the Tunçbilek FA. This characteristic band appeared due to the asymmetric vibrations of CO_3^{-2} ions, which pointed to the presence of sodium carbonate because of the carbonation reaction between excess sodium and atmospheric carbon dioxide [42,44]. The band 777 cm^{-1} was connected with the symmetric stretching vibration of Si-O-Si, which was characteristic of quartz. Quartz was also detected in the XRD analysis of the Tunçbilek FA (Figure 1), which supported the FTIR results. The last band observed at around 443 cm^{-1} was assigned to Si-O- and Al-O bending vibrations characteristic for silica glass and silicates [42].

Figure 4b shows the FTIR spectra of alkali activated compositions MW0, MW5, MW10, MW20, and MW40. With the addition of the MW, new peaks at frequencies of 1409, 866, 712 cm^{-1} were detected. These peaks indicated the presence of calcite and their intensities increased with the increase in the MW addition. The FTIR peak at 983 cm^{-1} (MW0) indicating geopolymerization shifted towards lower wavenumbers 979 (MW5), 969 (MW10), 968 (MW20), and 965 cm^{-1}(MW40) with increasing MW ratios. This result herein could not be interpreted in a similar way to the above discussion for Figure 4a because of the gradual replacement of FA by increasing MW ratio in geopolymer

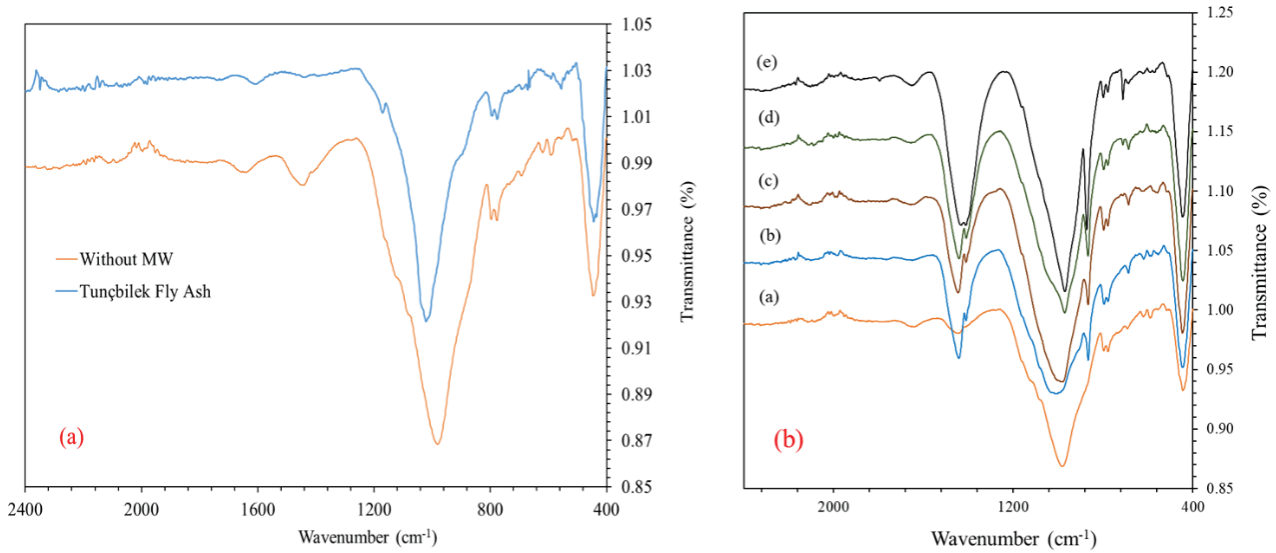


Figure 4. FTIR spectra of raw Tunçbilek FA and MW0 geopolymer (a) and geopolymers MW0 (a), MW5 (b), MW10 (c), MW20 (d), and MW40 (e) (b).

compositions. There were several articles on the formation and co-existence of the C-S-H phase within the geopolymer binder in the presence of significant amounts of calcium. A previous study reported that geopolymers with a high amount of Ca generated C-A-S-H phase along with the geopolymeric gel of N-A-S-H. In geopolymerization, Si^{4+} or Al^{3+} species react with Ca^{2+}, either in the FA or from external calcium-containing additive, to form calcium silicate hydrate gel (C-S-H), calcium alumino hydrate gel (C-A-H) or calcium aluminosilicate hydrate gel (C-A-S-H) in the presence of water [40,45-48]. Therefore, in the present study, the presumption explained in the following lines was adopted. During geopolymerization reactions, CaO from the MW ($CaCO_3$) was attacked by the alkali activation solution and turned into active species of Ca^{2+} and O^{2-}. The resultant active Ca^{2+} ions, along with Si^{4+} and Al^{3+} from the dissolution of reactive glassy FA microspheres, reacted with OH^- ions of alkali solution. Thus, C-A-S-H and N-A-S-H geopolymer gel formed [48]. It is thought that increasing MW content introduced more Ca^{2+} ions into the solution and accelerated the continuous development of geopolymer matrix, while the formation of C-S-H and C-A-S-H phases was promoted, which reflected itself as shifts in FTIR peaks towards lower wavenumbers [41,49]. These structures were expected to contribute to new phases in the sintering process and provide advantages in preserving material integrity.

Table 3 shows the change in apparent porosity and water absorption with the addition of increased MW. CO_2 gas, released on the decomposition of calcite [30,31,50,51] in the MW (equation 2), exerted pressure against the glassy phase of appropriate viscosity formed during sintering, leading to pore formation. An increase in the MW content increased apparent porosity and water absorption because of increased pore formation. While the apparent porosity and the water absorption of MW0 were respectively 12.7 and 8.92%, they increased to respective ratios of 39.92 and 28.17% in MW40. This result translates itself to more than a three-fold increase in apparent porosity and water absorption values. An increase in apparent porosity and water absorption with increasing content of the MW followed similar trends (Table 3) because water

Table 3. Physical and mechanical properties of sintered porous geopolymers

Sample code	Apparent porosity (%)	Water absorption (%)	Bulk density (g/cm³)	Linear shrinkage (%)	Flexural strength (MPa)	Compressive strength (MPa)
MW0	12.70	8.92	1.42	-0.70	19.85	34.98
MW5	16.38	10.55	1.55	1.44	15.30	31.74
MW10	22.87	14.84	1.54	1.71	12.02	27.00
MW20	31.41	20.85	1.51	1.15	6.14	17.50
MW40	39.92	28.17	1.42	0.81	5.35	12.00

absorption directly relates to the open porosity of the sintered ceramics. When compared with the apparent porosity values of porous ceramics, produced from FA and MW/or calcite wastes by various other methods for different application areas, 39.92% open porosity value of MW40 was satisfactory and even better than those of the previous studies [25]. The water absorption values were in line [41] with or even better than that of the literature [24]. Furthermore, the apparent porosity value of MW40 (39.92%) was also in good agreement with those of open-pore ceramics produced from different raw materials by various other methods [52,53]. However, it must be borne in mind that all those studies in the literature were carried out for much longer holding times, and most of them at much higher sintering temperatures.

In Table 3, the variation in bulk density and linear firing shrinkage with an increase in the MW ratio was depicted. The bulk density of the sintered alkali-activated specimens slightly decreased, as expected, with increasing apparent porosity (Table 3). MW0 composition of high melting point (1220°C) was successfully sintered at the low temperature of 900°C. High sintering temperatures between 1050-1300°C are needed to sinter F class FA [54-57]. Previous research reported that alkali activation was successful in the performance enhancement of ceramics sintered at the low temperatures [21,58]. The reason for lower temperature sintering of MW0 specimen at 900°C was the alkali activation of the Tunçbilek FA before sintering. The material produced from FA, not subjected to alkali activation, could not be sintered at 900°C, and therefore could not maintain its integrity. The bulk density increased from 1.42 to 1.55 g/cm^3 at 5 wt. % MW (MW5) and then slightly decreased with increasing ratio of the MW down to 1.42 g/cm^3 at MW40 composition. The first increase at 5 wt. % MW ratio may be because of an increasing amount of glassy phase formed by the fluxing action of calcite, which filled the pores in the structure and resulted in densification. In other words, calcite acted as a flux rather than acting as a pore-forming agent at this composition [36]. As the apparent porosity increased with the addition of the MW, the expected decrease in bulk density was not observed [31,32,35,50,51]. It is thought that this was because the MW used as a pore-making agent contained a high proportion of CaO (specific gravity 3.34 gr/cm^3). Therefore, although the apparent porosity increased, there was no significant change in the bulk density value.

Linear firing shrinkage also increased from almost 0 to 1.44% at MW5 and 1.71% at MW10 compositions, due to improved sintering then slightly decreased similarly with further increase in the MW content down to 0.81% at MW40 composition. The decrease in linear firing shrinkage can be attributed to the expansion that occurred due to increasing MW content acting as a pore former during sintering.

Microstructural Analyses of Sintered Porous Geopolymers

Figure 5 shows the SEM images of porous geopolymers produced with different amounts of the MW. The SEM micrograph of MW0 in Figure 5a showed that the structure had open and closed pores with thick pore walls. The structure was dense due to intense sintering at MW0 composition with a low apparent porosity value of 12.7% (Table 3). Due to the breakdown of $CaCO_3$ structure during sintering and the resulting CO_2 gas, the porosity increased in microstructures (Figure 5b-e) as the MW ratio increased. Although the number of large pores of MW40 appeared less in Figure 5-e than those of other compositions, it was, indeed, the specimen with the highest apparent porosity (39.92%) (Table 3). Therefore, MW0 and MW40 porous specimens were compared with each other at high magnification in Figure 6 to reveal this difference in porosity more clearly. It was observed that MW0 had large pores and its pore walls were dense (Figure 6a). On the other hand, the SEM analyses determined that MW40 contained many small pores. Consequently, the addition of the MW resulted in a more porous (39.92%) structure (Figure 6b).

Figure 7 shows the microstructure of the non-sintered MW40 composition. Compared with the microstructure after sintering (Figure 6b), the structure of the non-sintered specimen was dense, and the number of pores was low. However, after sintering, many new pores formed. Thus, a much more porous structure was observed by the removal of emitted CO_2 gas during the sintering process.

Phase Evolution of Sintered Porous Geopolymers

To study the crystalline phase evolution of fired alkali-activated specimens with increasing $CaCO_3$ (MW) addition was necessary for a better understanding of the modification of $CaCO_3$ on the pore structure. Figure 8 showed the XRD spectra and phase distribution of the specimens with different quantities of $CaCO_3$ sintered at 900 °C for 30 min. The Tunçbilek FA comprised quartz, mullite, and haematite while the MW was composed mainly of calcite and quartz (Figure 1). However, the crystalline phases of quartz, mullite and haematite in the Tunçbilek FA and calcite in the MW disappeared and transformed into the new phases of gehlenite, nepheline, quartz, orthoclase and calcium iron oxide (Figure 8) upon sintering at 900°C. The main crystalline phases in the structure of sintered geopolymers were gehlenite and nepheline. The MW contained a very high earth alkali Ca (61,65 wt. % CaO) ratio and Mg (0.69 wt. % MgO). The SiO_2 content of the MW was 2.2 wt. %. On the other hand, the Tunçbilek FA contained earth alkali Ca (1.64 wt. % CaO), Mg (4,48 wt. % MgO) and the reasonably high content of Fe (11.45 wt. % Fe_2O_3). Besides, Na^+ ions were also present in the structure due to alkali activation. All these alkaline earth oxides along with Fe_2O_3 have fluxing properties and it is thought that they contributed to the sintering of alkali-activated compositions at the low temperature of 900°C. This fluxing effect may reflect

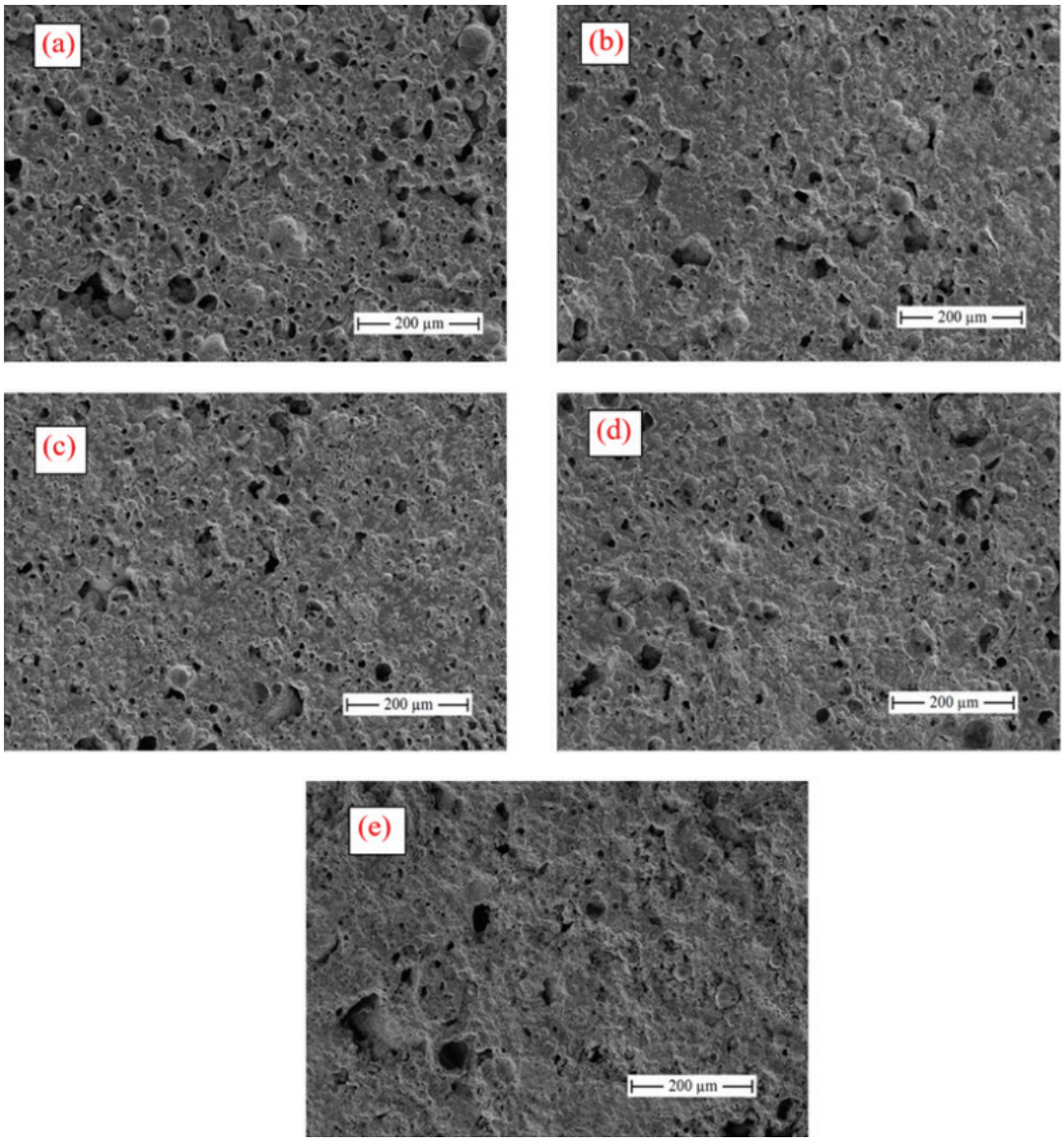


Figure 5. SEM analyses of sintered porous geopolymers MW0 (a), MW5 (b), MW10 (c), MW20 (d), MW40 (e).

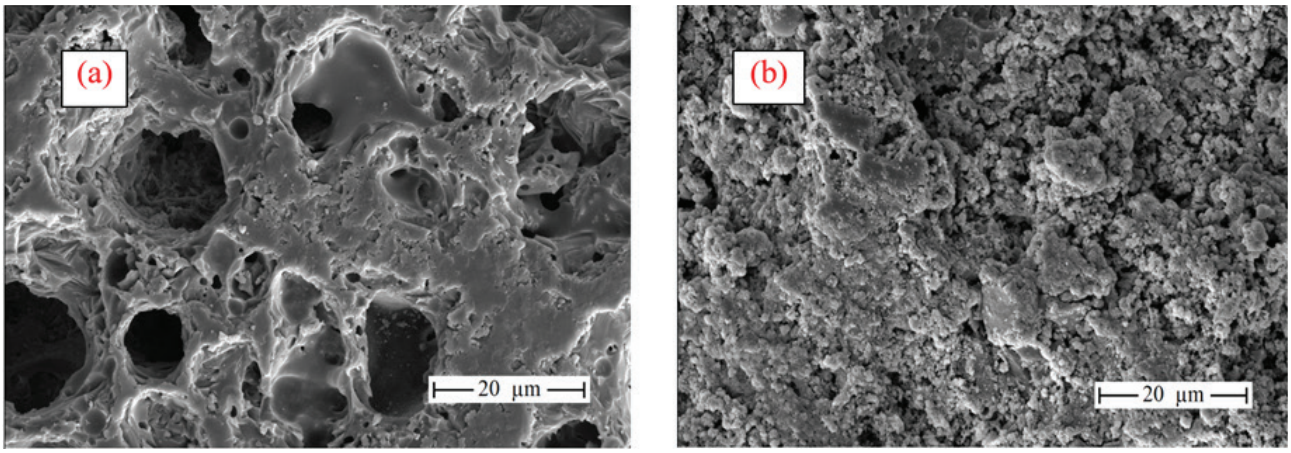


Figure 6. SEM analyses of sintered porous geopolymers; MW0 (a), MW40 (b).

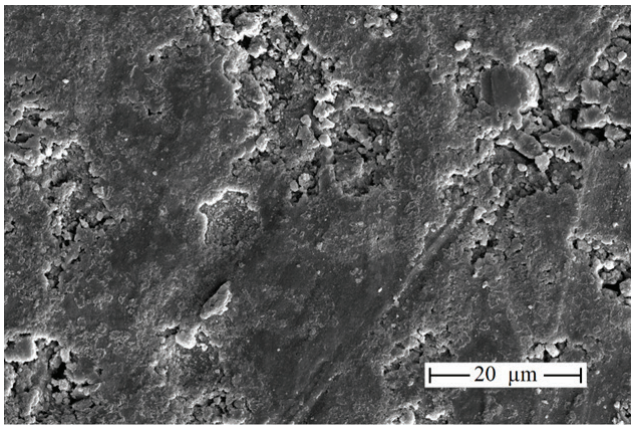


Figure 7. SEM image of non-sintered MW40 specimen.

itself as a lower glass phase viscosity, which helped to overcome the resistance against the CO_2 pressure, leading to a porous structure. The phases of mullite, hematite, calcite in the fired geopolymers were totally, and quartz was partially consumed to develop new phases of gehlenite and nepheline (Table 2, Figure 1, Figure 8).

Here, the addition of the MW had an effect on the phase structure formed. When the MW ratio was 5, 10 and 20 wt.%, the main phase was nepheline. However, when the MW addition was 40 wt.%, a low ratio of nepheline, orthoclase and calcium iron oxide phases developed in the structure, along with a large proportion of gehlenite phase. The absence of the free CaO phase was critical for the integrity and strength of the structure. The Na ions in the alkaline activation solution contributed to forming new phases such as nepheline and gehlenite at low temperature (900 °C), resulting in strong porous ceramics. As seen in Figure 8f, a very different phase structure was obtained when the MW40 composition, which was not subjected to alkali activation, was sintered at 900°C. It was determined that there was free CaO at (a) high ratio, quartz, and $Ca(OH)_2$ formed from free CaO in this phase structure. The high amount of CaO is extremely unfavourable in terms of the integrity of the structure. As can be understood from here, it was shown in this study that using a material with high $CaCO_3$ content was possible with alkali activation. In addition, the sintered sample obtained when alkali activation was not applied was easily broken with a small force applied by hand (Figure 8g). Table 3 showed the variation in the compressive and flexural strength of the porous specimens obtained with increasing MW ratio. The addition of the MW caused a decline in mechanical properties. While this decline in strength was rapid up to 20 wt. % MW addition, it diminished at 40 wt. % MW addition. The porous specimen produced without the MW (MW0) had 34.98 MPa compressive and 19.85 MPa flexural strength values, while the 40 wt. % MW added porous specimen (MW40) had 12 MPa compressive and 5.35 MPa flexural strength values. The decrease in strength values was consistent with the increased apparent porosity. However, they were acceptable for open-pore materials [53].

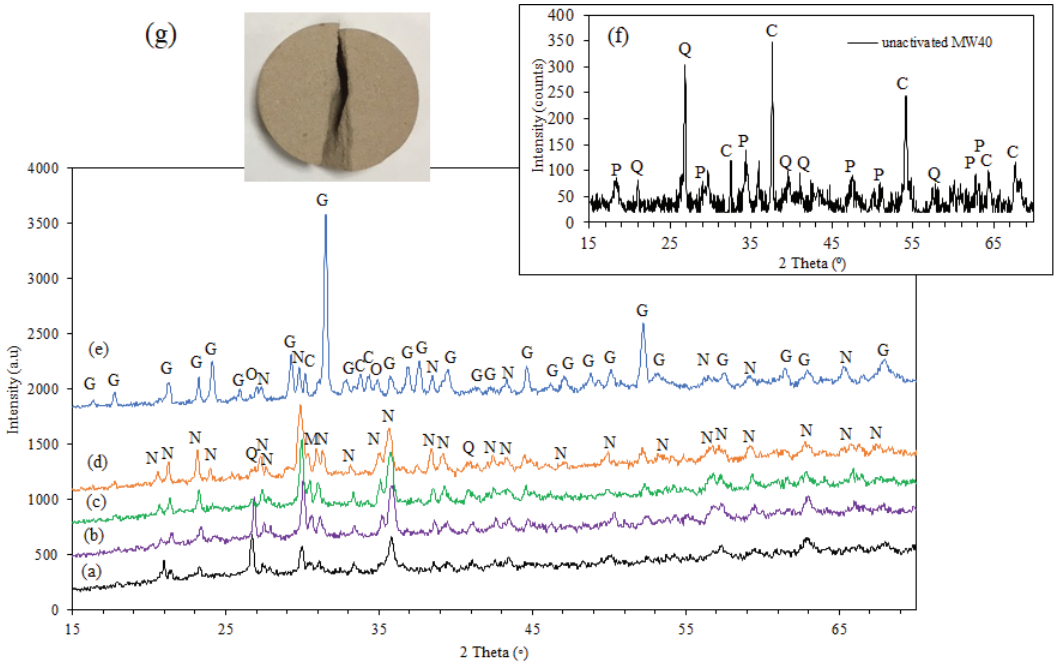


Figure 8. XRD spectra of sintered porous geopolymers; a) MW0, b) MW5, c) MW10, d) MW20, e) MW40, f) un activated MW40, g) un activated MW40 photograph (G: Gehlenite, N: Nepheline, Q: Quartz, O: Orthoclase, C: Calcium Iron Oxide, P: Portlantide (Ca(OH)2)).

CONCLUSION

The low-cost porous ceramics with sufficient apparent porosity and mechanical properties were successfully fabricated entirely from waste materials FA and MW. The experimental compositions were first alkali activated and then sintered at low temperature (900°C). The MW with high $CaCO_3$ content was used for the first time as a pore-making agent at ratios between 5-40 wt. %. It was demonstrated that an increase in the MW amount had a positive effect on the open porosity and the porosity increased proportionally with increasing MW amount. The maximum apparent porosity of the FA based open-pore ceramic containing 40 wt. % MW was 39.92%. The corresponding flexural and compressive strength values were found to be 5.35 and 12 MPa, respectively. The maximum amount of MW (40 wt. %) used in the present study was the highest in the related literature. Besides, alkali activation enabled the formation of main phases such as gehlenite (at 40 wt. % MW) and nepheline (at 0-20 wt. % MW) without free CaO in the structure as a result of the sintering process depending on the FA and MW ratios. This study showed that MW, which is an alternative waste material to $CaCO_3$ containing pore-making agents, could be used as a pore-making agent. The porous ceramic has the potential for use as a water purification membrane.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the indings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Das D, Kayal N. Thermal shock resistance of porous silicon carbide ceramics prepared using clay and alumina as additives. Trans Indian Ceram Soc 2019;78:165-171. [CrossRef]

[2] Şan O, Koç M, Cengizler H. Production of porous ceramic from clinoptilolite incorporating aluminum powder. Ceram Int 2019;45:24037-24043. [CrossRef]

[3] Mustaffar MI, Mahmud MH. Processing of highly porous glass ceramic from glass and fly ash wastes. AIP Conf Proc 2018;2031:020010. [CrossRef]

[4] Dong Y, Liu X, Ma Q, Meng G. Preparation of cordierite-based porous ceramic micro-filtration membranes using waste fly ash as the main raw materials. J Memb Sci 2006;285:173-181. [CrossRef]

[5] Wehling J, Köser J, Lindner P, et al. Silver nanoparticle-doped zirconia capillaries for enhanced bacterial filtration. Mater Sci Eng C 2015;48:179-187. [CrossRef]

[6] Sun Z, Bai C, Zheng S, Yang X, Frost RL. A comparative study of different porous amorphous silica minerals supported TiO2 catalysts. Appl Catal A Gen 2013;458:103-110. [CrossRef]

[7] Hu LF, Wang CA. Effect of sintering temperature on compressive strength of porous yttria-stabilized zirconia ceramics. Ceram Int 2010;36:1697-1701. [CrossRef]

[8] Bories C, Borredon ME, Vedrenne E, Vilarem G. Development of eco-friendly porous fired clay bricks using pore-forming agents: A review. J Environ Manage 2014;143:186-196. [CrossRef]

[9] Malik N, Bulasara VK, Basu S. Preparation of novel porous ceramic microfiltration membranes from fly ash, kaolin and dolomite mixtures. Ceram Int 2020;46:6889-6898. [CrossRef]

[10] Zong Y, Wan Q, Cang D. Preparation of anorthite-based porous ceramics using high-alumina fly ash microbeads and steel slag. Ceram Int 2019;45:22445-22451. [CrossRef]

[11] Galán-Arboledas RJ, Cotes T, Martínez C, Bueno S. Influence of waste addition on the porosity of clay-based ceramic membranes. Desalin Water Treat 2016;57:2633-2639. [CrossRef]

[12] Al-Qadhi E, Li G, Ni Y. Influence of a two-stage sintering process on characteristics of porous ceramics produced with sewage sludge and coal ash as low-cost raw materials. Adv Mater Sci Eng 2019;2019:3710692. [CrossRef]

[13] Cusidó JA, Cremades L V, Soriano C, Devant M. Applied Clay Science Incorporation of paper sludge in clay brick formulation : Ten years of industrial experience. Appl Clay Sci 2015;108:191-198. [CrossRef]

[14] Vlasova M, Rosales I, Kakazey M, Parra AP, Guardian R. Formation of porous ceramics using cullet and biological waste of water purification. Sci Sinter 2011;43:81-94. [CrossRef]

[15] Gislon ES, Simão L, Coelho K. Permeability of porous ceramic membranes obtained from waste of the coal extraction process. Mater Sci Forum 2016;881:357-361. [CrossRef]

[16] Ho C, Lo H, Lin K, Lan J. Characteristics of porous ceramics from prepared from sandblasting waste and waste diatomite by co-sintering process. 2019;38:321-328. [CrossRef]

[17] Manni A, El Haddar A, Hassani IZEA, El Bouari A, Sadik C. Valorization of coffee waste with Moroccan clay to produce a porous red ceramics. 2019;58:211-220. [CrossRef]
[18] Wang Q, Yu H, Ben T, Li Q, Li F, Xu H, et al. Preparation of lightweight high-strength thermal insulation and decoration integration porous ceramics using red mud. 2019;1:91-98. [CrossRef]
[19] Aziz IH, Al Bakri Abdullah MM, Yong HC, Ming LY, Hussin K, Surleva A, et al. Manufacturing parameters influencing fire resistance of geopolymers: A review. 2016;233:721-733. [CrossRef]
[20] Rickard WDA, Temuujin J, Van Riessen A. Thermal analysis of geopolymer pastes synthesised from five fly ashes of variable composition. J Non Cryst Solids 2012;358:1830-1839. [CrossRef]
[21] Jaya NA, Mustafa M, Ghazali CMR, Hussain M, Hussin K, Ahmad R. Kaolin geopolymer as precursor to ceramic formation. 2016;78:01061. [CrossRef]
[22] Rickard W, Kealley C, Van Riessen A. Thermally ınduced microstructural changes in fly ash geopolymers: Experimental results and proposed model. 2015;939:929-940. [CrossRef]
[23] Villaquirán-Caicedo MA, Gutiérrez RM De. Synthesis of ceramic materials from ecofriendly geopolymer precursors. Mater Lett 2018;230:300-304. [CrossRef]
[24] Al-Bakri Abdullah MM, Jamaludin L, Hussin K, Bnhussain M, Ghazali CMR, et al. Fly ash porous material using geopolymerization process for high temperature exposure. Int J Mol Sci 2012;13:4388-4395. [CrossRef]
[25] Sawan SEA, Zawrah MF, Khattab RM, Abdel-shafi AA. In-situ formation of geopolymer foams through addition of silica fume: Preparation and sinterability. Mater Chem Phys 2020;239:121998. [CrossRef]
[26] ASTM. Standard test methods for apparent porosity, water absorption, apparent specific gravity, and bulk density of burned refractory brick and shapes by boiling water. Available at: https://edisciplinas.usp.br/pluginfile.php/3773002/mod_resource/content/0/Norma%20determina%C3%A7%C3%A3o%20densidade%20a%20verde.pdf. Accessed on Jul 4, 2024.
[27] ASTM. Standard test methods for sampling and testing brick and structural clay tile. Available at: https://cdn.standards.iteh.ai/samples/105428/ba8f623c3f4e4545b5c9d55e2f9fd59b/ASTM-C67-C67M-20.pdf. Accessed on Jul 4, 2024.
[28] ASTM. Standard test method for drying and firing shrinkages of ceramic whiteware clays. Available at: https://cdn.standards.iteh.ai/samples/66723/35a6cf0afa2140cbaae8d19afaac3f95/ASTM-C326-09.pdf. Accessed on Jul 4, 2024.
[29] ASTM. Standard specification for coal fly ash and raw or calcined natural pozzolan for use in concrete. Available at: https://webstore.ansi.org/standards/astm/astmc61812. Accessed on Jul 4, 2024.
[30] Sutcu M, Alptekin H, Erdogmus E, Er Y, Gencel O. Characteristics of fired clay bricks with waste marble powder addition as building materials. Constr Build Mater 2015;82:1-8. [CrossRef]
[31] Munir MJ, Abbas S, Nehdi ML, Kazmi SMS, Khitab A. Development of eco-friendly fired clay bricks incorporating recycled marble powder. J Mater Civ Eng 2018;30:100006. [CrossRef]
[32] Williams RP, Van Riessen A. Determination of the reactive component of fly ashes for geopolymer production using XRF and XRD. Fuel 2010;89:3683-3692. [CrossRef]
[33] Hemmings RT, Berry EE. On the glass ın coal fly ashes: Recent advances. MRS Proc 1987;113:3. [CrossRef]
[34] Hemmings RT, Berry EE, Cornelius BJ, Scheetz BE. Speciation in size and density fractionated fly ash ıı. characterization of a low-calcium, high-ıron fly ash. MRS Proc 1986;86:81. [CrossRef]
[35] Bilgin N, Yeprem HA, Arslan S, Bilgin A, Günay E, Maroglu M. Use of waste marble powder in brick industry. Constr Build Mater 2012;29:449-457. [CrossRef]
[36] Sokolář R, Vodová L, Grygarová S, Štubňa I, Šín P. Mechanical properties of ceramic bodies based on calcite waste. Ceram Int 2012;38:6607-6612. [CrossRef]
[37] Wons W, Rzepa K, Reben M, Murzyn P, Sitarz M, Olejniczak Z. Effect of thermal processing on the structural characteristics of fly ashes. J Mol Struct 2018;1165:299-304. [CrossRef]
[38] Terzi A, Pavlovi V. Novel Utilization of Fly Ash for High-Temperature Mortars : Phase Composition, Microstructure and Performances Correlation. Int J Appl Ceram Technol 2015;12:133-146. [CrossRef]
[39] Goga F, Dudric R, Cormos C, Imre F, Bizo L, Misca R. Fly ash from thermal power plant, raw material for glass-ceramic. Environ Eng Manag J 2013;12:337-342. [CrossRef]
[40] Xiaolu GUO, Huisheng SHI, Maosong LIN, Wenjing D. Effects of calcium contents in class C fly ash geopolymer. Adv Mat Res 2013;687:508-513. [CrossRef]
[41] Mustafa AM, Al Bakri Abdullah MM, Kamarudin H, Bnhussain M, Nizar IK, Razak RA, et al. The processing, characterization, and properties of fly ash based geopolymer concrete. Rev Adv Mater 2012;30:90-97.
[42] Alehyen S, Achouri MEL, Taibi M. Characterization, microstructure and properties of fly ash-based geopolymer. J Mater Environ Sci 2017;8:1783-1796.
[43] Khan I, Azizli K, Sufian S, Siyal A, Man Z. Sodium silicate free geopolymer as coating material: adhesion to steel. In proceedings of 1st International Electronic Conference on Materials. 2014 May 26-June 10; Basel, Switzerland. 2014. [CrossRef]
[44] Kioupis D, Kavakakis C, Tsivilis S, Kakali G. Synthesis and characterization of porous fly ash-based geopolymers using Si as foaming agent. Adv Mater Sci Eng 2018;2018:1942898. [CrossRef]

[45] Yip CK, Deventer JSJV. Microanalysis of calcium silicate hydrate gel formed within a geopolymeric binder. J Mater Sci 2003;8:3851-3860.
[46] Dong M, Feng W, May F, Elchalakani M, Li GK, Karrech A. Development of a high strength geopolymer by novel solar curing. Ceram Int 2017;43:11233-11243. [CrossRef]
[47] Yu X, Chen L, Komarneni S, Hui C. Fly ash-based geopolymer: Clean production, properties and applications. J Clean Prod 2016;125:253-267. [CrossRef]
[48] Azimi EA, Abdullah MMAB, Vizureanu P, Salleh MAAM, Sandu AV, Chaiprapa J, et al. Strength development and elemental distribution of dolomite/fly ash geopolymer composite under elevated temperature. Materials (Basel) 2020;13:1015. [CrossRef]
[49] Zhao X, Liu C, Zuo L, Wang L, Zhu Q, Wang M. Investigation into the effect of calcium on the existence form of geopolymerized gel product of fly ash based geopolymers. Cem Concr Compos 2019;103:279-292. [CrossRef]
[50] Harabi A, Boudaira B, Bouzerara F, Foughali L, Zenikheri F, Guechi A, et al. Porous ceramic supports for membranes prepared from kaolin (DD3) and calcite mixtures. Acta Phys Pol A 2015;127:1164-1166. [CrossRef]
[51] Klosek-Wawrzyn E, Malolepszy J, Murzyn P. Sintering behavior of kaolin with calcite. Procedia Eng 2013;57:572-582. [CrossRef]
[52] Studart R, Gonzenbach UT, Tervoort E, Gauckler LJ. Processing routes to macroporous ceramics: A review. J Am Ceram Soc 2006;89:1771-1789. [CrossRef]
[53] Vakifahmetoglu C, Zeydanli D, Colombo P. Porous polymer derived ceramics. Mater Sci Eng Rep 2016;106:1-30. [CrossRef]
[54] Biernacki JJ, Vazrala AK, Leimer HW. Sintering of a class F fly ash. Fuel 2008;87:782-792. [CrossRef]
[55] Lingling X, Wei G, Tao W, Nanru Y. Study on fired bricks with replacing clay by fly ash in high volume ratio. Constr Build Mater 2005;19:243-247. [CrossRef]
[56] Biernacki JJ, Mogula NR, Dunne JK, Nagolu RR. Kinetics of sintering for a class-F fly ash: A sintering model. In: Brandth, Glinicki MA, Olek J, Leung CKY, editors. Brittle Matrix Composites 10. 1st ed. Cambridge: Woodhead Publishing; 2012. p. 71-89. [CrossRef]
[57] Erol M, Ku S. Characterization of sintered coal fly ashes. Fuel 2008;87:1334-1340. [CrossRef]
[58] Luo Y, Ma S, Liu C, Zhao Z, Zheng S, Wang X. Effect of particle size and alkali activation on coal fly ash and their role in sintered ceramic tiles. J Eur Ceram Soc 2017;37:1847-1856. [CrossRef]

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1160–1168, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00094

Research Article

Cognitive activity detection and tracing system

Onur YILDIRIM[1,*], Çağla KANDEMIR[2], Emre KARDAŞLAR[3], Emre SÜMER[4]

[1]*Department of Computer Engineering, Bilkent University, 06800, Türkiye*
[2]*Vodafone, Istanbul, 34398, Türkiye*
[3]*Akakce, Ankara, 06530, Türkiye*
[4]*Department of Computer Engineering, Baskent University, Ankara, 06790, Türkiye*

ARTICLE INFO

Article history
Received: 19 January 2023
Revised: 29 April 2023
Accepted: 30 August 2023

Keywords:
Alzheimer; Cognitive Activity; Data Analysis; Dementia; Feedback; Game; Tracking

ABSTRACT

Cognitive problems like Dementia and Alzheimer's are usually challenging to diagnose but can be noticed by some signs of their symptoms. The most common symptoms are confusion, trouble finding the right word, memory loss, and difficulty concentrating. This study aims to design a cognitive activity detection and tracing system that contains games and analyzes users' performances then displays detailed statistics to the users. The proposed Cognitive Activity Detection and Tracing System (CADTS) is software that contains different kinds of games from different categories inside its body that aims to measure cognitive activity by utilizing formulations in the context of the games and give feedback to users concerning the performance analyses done. The purpose of these analyses is to catch the signs of symptoms. An insight into a possible scoring system is provided, and as our results, several descriptive statistics are shared based on the tests conducted.

Cite this article as: Yıldırım O, Kandemir Ç, Kardaşlar E, Sümer E. Cognitive activity detection and tracing system. Sigma J Eng Nat Sci 2024;42(4):1160–1168.

INTRODUCTION

Dementia, Alzheimer and other cognitive problems are encountered too often, and early diagnosis may be necessary for the treatment to be more effective. Diagnosis of cognitive problems is usually difficult, but there are symptoms that can be a sign of one of those cognitive problems.

The purpose of this work is to create a desktop application that is suitable for all people of all ages to play and evaluate them with their scores in the background compared to other users. Each game is built to evaluate a specific attribute, and the evaluation of a user performance of that game is done related to specific indicators. Low scores do not represent a diagnosis of any cognitive problem since these games aim not to diagnose any disease [1]. The detection and screening processes of cognitive diseases are complex, and the screening tool's accuracy is disputable [2]. However, these low scores might be a sign of a problem that may be diagnosed as a cognitive illness with the further investigation since the labeling process of the scores is based on comparisons with other players.

Games selected that fit into categories are developed by the Unity Game Engine [3]. Firebase [4] database is used

*Corresponding author.
*E-mail address: o.yildirim@bilkent.edu.tr

This paper was recommended for publication in revised form by Editor in-Chief Ahmet Selim Dalkilic

for keeping the statistics then those statistics are analyzed. For this paper, a literature review and research of similar studies, software, and hardware specification, determining the four-game categories, developing a game for one of the categories, and interpreting the saved scores are done. After that, the development of the games for the other categories, some improvements to these games, and the user interface are implemented. After the study's goal is achieved, an application that provides people, especially people at risk of neurological conditions, with detailed reports of their performance is actualized. Users are evaluated in four different age ranges defined by the United Nations [5]. Each user is compared to other users in their age range for the specific game they played.

There are a lot of other applications that try to do the evaluation task in different ways; what sets our work apart from others is the feedback mechanism. This feedback mechanism allows users to track their past performances through a line chart, which can be e-mailed to them or their relative.

REVIEW OF LITERATURE

There are many studies conducted that show the relation between games and cognitive activities [6,7,8,9,10]. One study [6] is conducted on middle-school students with stealth assessment embedded into Plants vs. Zombies 2. In this study, the researchers created a model to test problem-solving skills and implemented this model into the game with Bayesian networks. To validate the stealth assessment, they ran a pilot study on middle-school students to collect data, analyze it and update the model. This loop continued throughout the process to measure students' problem-solving skills more accurately each iteration. They stated that their study shows a correlation between the game and problem-solving skills. Another study [10] also aims to measure problem-solving skills by utilizing stealth assessment in a game called Use Your Brainz. The researchers focus on two external problem-solving tests, Raven's Progressive Matrices [11] and MicroDYN [12]. The observation was done on 55 7th-grade students, and each played the game for about three hours across three consecutive days. The researchers state that the problem-solving estimates derived from the game correlate with the external measures. Therefore, their stealth assessment is valid.

There are a lot of existing methods and tests for measuring attention. Posner cueing task [13], Attention Network Test (ANT) [14], and Conners Continuous Performance Test [15] can be given as examples. This study focuses on visual search tasks under the attention category. In studies around visual search tasks, there are many mentions of the term "selective attention" [16]. Two visual search task examples are given in Figure 1. The example on the left side asks the user to find the number of green or red lines whereas the example on the right asks the user to find the number "2" that is blent into several numbers "5"s.

The brain is exposed to vast amounts of sensory information throughout the day. Nevertheless the brain, especially in focus, cannot fully process all of this information (i.e., traffic noise, music in the background). "Selective attention" can be defined as the brain ignoring certain stimuli (noise) to focus on an object or task. This function of the brain has been studied along with visual search tasks to be integrated into our study in the form of a mini-game under the attention category [18].

Memory is acquiring, storing, retaining, and retrieving information. Encoding, Storage, and Retrieval are the main processes of memory. On the other hand, working memory is a function that allows humans to simultaneously store and process information in immediate consciousness for a limited duration [19].

According to the studies available, there are two kinds of tests on memory. These are direct and indirect memory tests. Direct memory tests use the subject's personal experiences and aim to see whether or not the subject can give behavioral evidence on that subject. With indirect memory tests, the subject is asked to complete some cognitive or motor memory-related activities.

Indirect tests fall under four categories [20]:

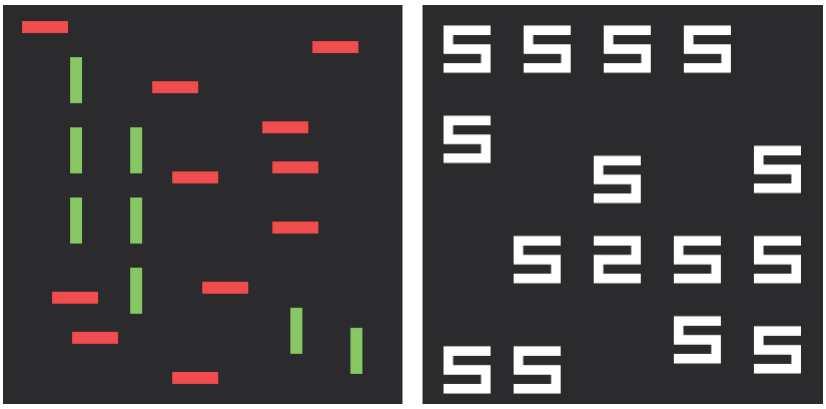


Figure 1. Visual search task (adapted from [17] with permission Visual Attention Lab).

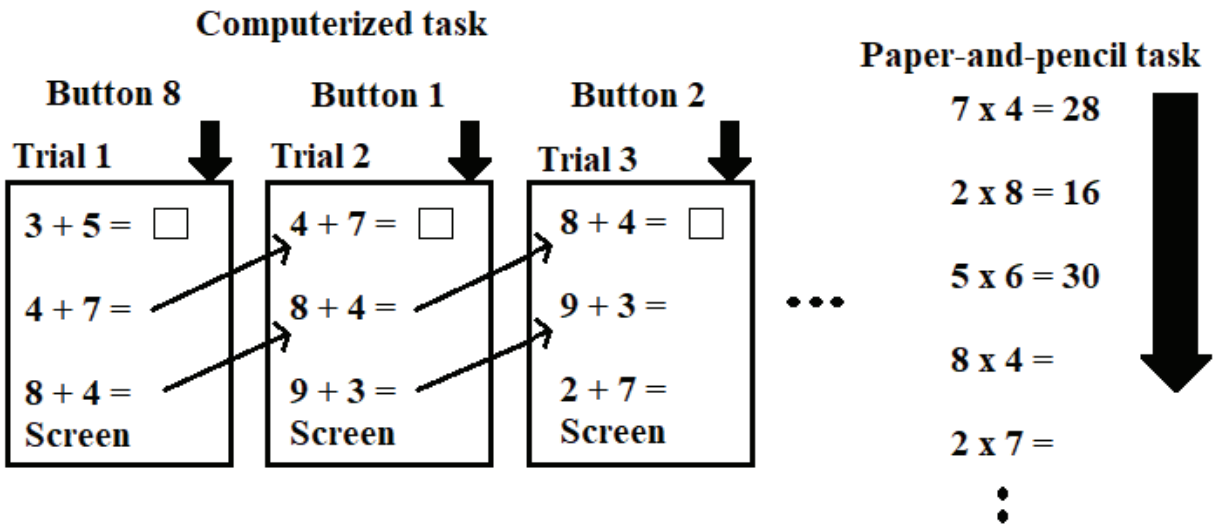


Figure 2. Schema of examples of training tasks used in this study. Training tasks consisted of computerized (addition, subtraction, and multiplication) and paper-and-pencil (addition, subtraction, and multiplication) tasks (adapted from [21], under the Creative Commons Attribution License).

- Tests of conceptual, factual, lexical, and perceptual knowledge
- Tests of procedural knowledge
- Measures of evaluative response
- Other measures of behavioral change

Suppose the tests of conceptual, factual, lexical and perceptual knowledge (especially factual) are examined in detail. In that case, it is seen that tasks that fall under this category focus more on the retrieval of information. In factual domains, the subject is asked to answer questions of general knowledge, react to certain stimuli, and categorize them [20].

In these kinds of tasks, the measurement parameters are accuracy or the latency of a correct response. The more the prior exposure to stimuli increases, the more the accuracy or latency decreases. This is called Direct/Repetition Priming [20]

Another study has been made to observe the effects of fast simple numerical calculations (FSNC) on neural systems. In this study, simple mathematical problems are asked of participants, like single-digit addition, subtraction, and multiplication, as shown in Figure 2. The ability to complete an FSNC task correlates with processing speed, quantitative ability or knowledge, and general intelligence. Previous studies of psychological interventions showed that cognitive interventions involving arithmetic or FSNC tasks lead to improvements in performance on untrained cognitive tasks (transfer effects) among the elderly as well as dementia patients. Besides the effects on the performance improvement, FSNC is used to trace the brain activity and measure participants' performance.

Studies have shown that fast simple numerical calculations (FSNC) can be used to measure a person's cognitive skills. According to a study that researches the effects of FSNC training in neural systems, by the changes on levels of regional gray matter volume (rGMV) and regional cerebral blood flow (rCBF), the effects of FSNC training can be observed [21]. Changes in rGMV and rCBF provide direct evidence that FSCN can measure cognitive skills.

Other studies are not mentioned in this section but can be accessed in [22]. Besides the studies done in the academic field, there are also a lot of commercial applications like Lumosity, Elevate, etc. These commercial applications have advertised or are still advertising themselves as "cognitive training" applications. There have been studies on the effectiveness of these applications, and there have been conflicting results. Brain training exercises can be effective in the long term [23]; however, a study that focuses on Lumosity in its experiments concludes by stating that Lumosity lacks strong evidence for improving the cognitive skills of the players [24]. In a study conducted [25], it is shown that a group of players who played Portal 2, which is a puzzle game with no intention of improving cognitive skills, outperforms the other group of players that played Lumosity on some tests.

CADTS STRUCTURE

CADTS structure is divided into four games and a feedback mechanism. The categories those games belong to are memory, problem solving, maths and attention. The use-case diagram of the system is explained in Figure 3.

Memory

The game that has been developed for the Memory category contains four buttons in four different colors. These

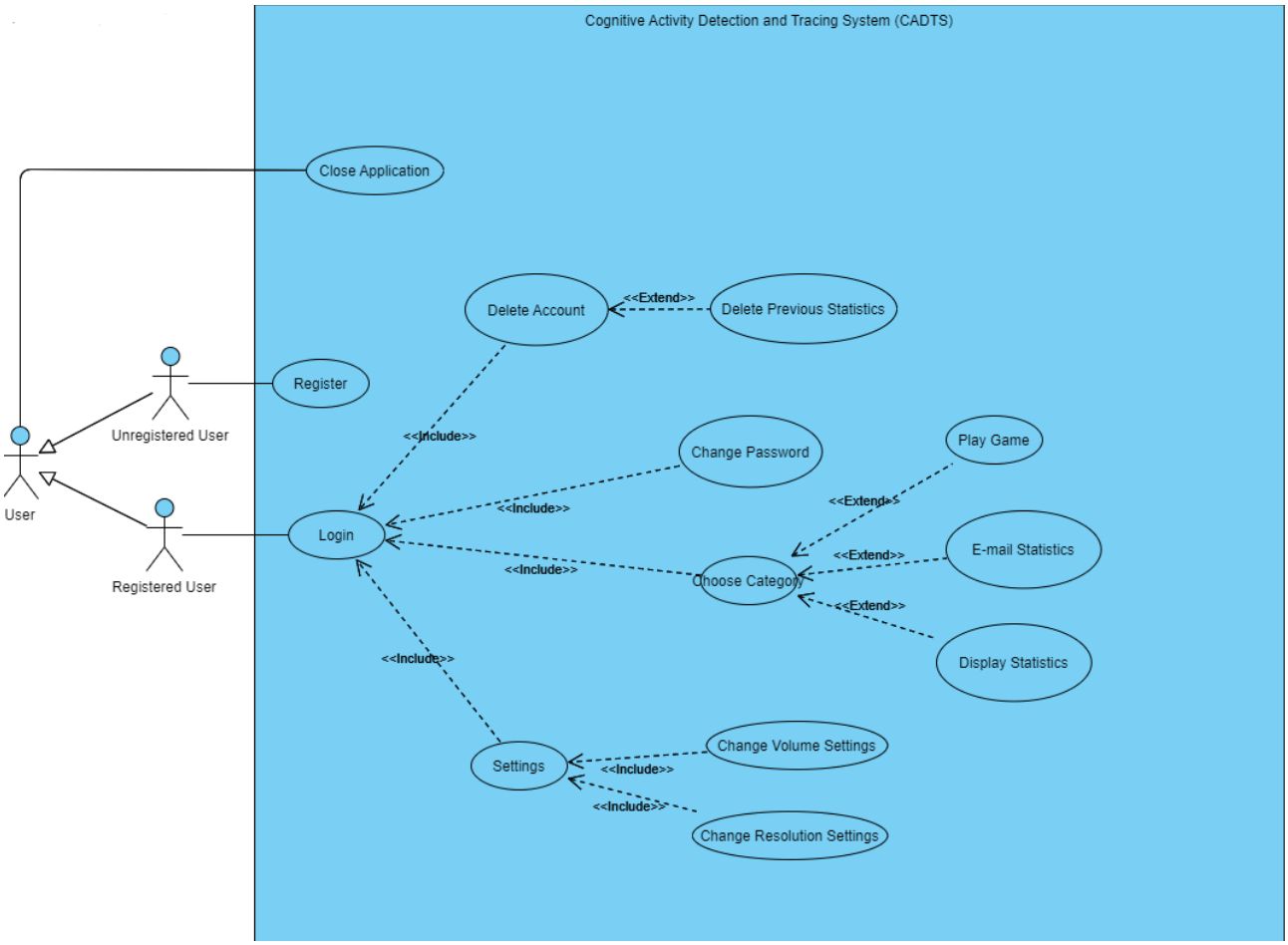


Figure 3. Use case diagram of CADTS application.

buttons start to light up in random orders three times in the beginning. After the sequence is given, the user is expected to push the buttons in the same order. After every level the user manages to provide the correct answer, another random button is added to the initial sequence. If a wrong button is pushed, the game ends. A visual from the memory game is given in Figure 4.

Figure 4. Buttons of the memory game.

This game falls under the category of indirect tests and, more specifically, tests of conceptual, factual, lexical, and perceptual knowledge. In these kinds of tests, the measure parameters are the accuracy and latency of a correct response. According to this information, the scoring of this game is calculated using the number of buttons correctly pushed (b_j) and the time (t_j) in which the level is completed. Ten points are added for every correct button push (b_m) , and at the end of the level, the total score is divided by 1% of the time passed and then multiplied by the level coefficient (l_j).

For the playthrough i with n levels, the memory score (ms_i) is computed as follows:

$$ms_i = \sum_{j=0}^{n} (b_j \; b_m \; / (t_j \; t_m)) l_j \tag{1}$$

where b_j is the number of correct button pushes, t_j is the total time passed in level j, and l_j is the level index (i.e 1 for the first level, 2 for the second). The multipliers b_m and t_m in Equation 1 is set to 10 and 100, respectively, for our experiments.

Problem Solving

Problem-solving game design is based on the definition of problem-solving skill. Problem-solving is usually explained in four steps [26]:

1. Defining the problem
2. Generating alternative solutions
3. Evaluating solutions and selecting the best solution among the alternatives
4. Implementing the solution

The developed game aims to build a path from the pipe connected to the valve to the exit. Users can interact with two types of pipes, and each click rotates them 90 degrees in clockwise direction. The visuals of the pipe types are given in Figure 5. Water starts flowing from top to bottom, and rotating the pipes changes the direction the water is flowing.

Figure 5. I type and L type pipes.

The game consists of three levels with different pipe counts for each level. Pipe types and their rotations are randomly generated at each level's start. That design choice aims to evaluate every player almost reasonably since every level is randomly generated, and some people can encounter manageable levels while others can encounter challenging levels. The first level is a 4x7 map, the second level is a 5x8 map, and the final level is a 5x12 map; with that design, there is usually an average difficulty for each gameplay cycle.

A few variables are tracked during the gameplay to evaluate players' problem-solving skills when all the levels are completed. A model is created for the problem evaluation mentioned:

The model mentioned in Table 1 consists of two main attributes: planning the solution pathway, and efficient usage of given resources. If a player turns on the valve and the water flow cannot reach the exit, the player could not generate a solution and loses points. The time and length of the path are evaluated as resources, and less those values, the higher the score players earn. A final indicator of efficient usage of given resources is the rotation count for each pipe. Each pipe has four states, and the starting state of every pipe is randomly generated as mentioned before. Rotating a pipe three times yields every state a pipe can be in an order. Rotating a pipe for the fourth time results in a cycle (c_i), and the pipe goes back to its starting state, which is evaluated as non efficient, and players lose points for performing that.

For the playthrough i, the problem-solving score (ps_i) is computed as follows:

$$ps_{i=}t_i^{-1}\, t_{m\_}w_i w_{m\_}c_i\, c_{m\_}\, p_i\, p_m \tag{2}$$

where t_i is the total time passed, w_i is the number of wrong attempts, c_i is the total number of cycles, and p_i is the total number of pipes in the solutions. The multipliers in Equation 2 t_m, w_m, c_m and p_m are set accordingly for the scores to be balanced between players. For instance, in our experiments, players are punished more for the wrong attempts (w_m=15) than the pipe cycles they perform (c_m=5). The utilized values of the other multipliers t_m and p_m are 2500 and 1, respectively.

Maths

While designing the game, it is considered that the games should be playable for all ages. This brought a challenge since older gamers generally have difficulty playing games with certain boundaries like time, game speed, etc. According to the research, the games quickly become overwhelming for elders because of the decline in processing speed and reaction time [27]. In order to prevent this, the game's design was kept simple. Instead of using complex mathematical problems, simple numerical calculations are used. The numbers used in those calculations also kept small to decrease the player›s cognitive load.

Balloons are used to visualize the calculations. All balloons in the game have simple calculations written on their bodies, as given in Figure 6. The players need to enter the result of those calculations into the input field to blow up

Table 1. Indicators used to evaluate score in problem solving game

Measured attribute	Example(s) of indicators
Planning the solution pathway	- Water reaches the exit when the valve is turned on
Efficient usage of given resources	- Number of pipes that water passes through from start to exit (p_i)
	- Total time passed from start to the end of all levels (t_i)
	- Rotate count for each pipe should be <4 in order to avoid cycles

Figure 6. Balloon with calculations.

the balloons and earn points. There is an obstacle at the top of the game screen; if the balloons touch the obstacle, the balloons blow up. The game ends if the balloons blow up by touching the obstacle five times. The users need to blow up the balloons by entering the correct result for the balloons. The overall structure of the game is given in Figure 7.

The difficulty scaling is achieved by increasing the speed of the new spawning balloon. Considering the playthrough for the elders and younger players the balloon's speed increases by 0.25 units/seconds according to the speed of the last balloon. According to our user tests and feedback, 0.25 units/second is optimal for all the players. After a certain level which can differ for every player, the player cannot calculate the results since the balloons become too fast for the player. This is the ending condition for the math game since there is not any time limit used in this game.

Players earn 10 points for every correct answer and lose 3 points for every wrong answer. Entering the wrong answer does not blow up the balloon; the player still can enter the correct answer for that balloon before the balloon touches the obstacle. The scores of the players are compared among themselves according to a range of ages. The players get motivated to play the game when they see that their peers get better scores; they also get motivated when they see their scores increase over time. The feeling of progress is the key to the motivation of the players.

For the playthrough i, the math score (mas_i) is computed as given in Equation 3,

$$mas_{i\,=}co_i\,co_{m-}w_i w_m \tag{3}$$

where w_i is the number of wrong answers, and co_i is the total number of correct answers. The multipliers co_m and $_{wm}$ are used with the values of 10 and 3, respectively.

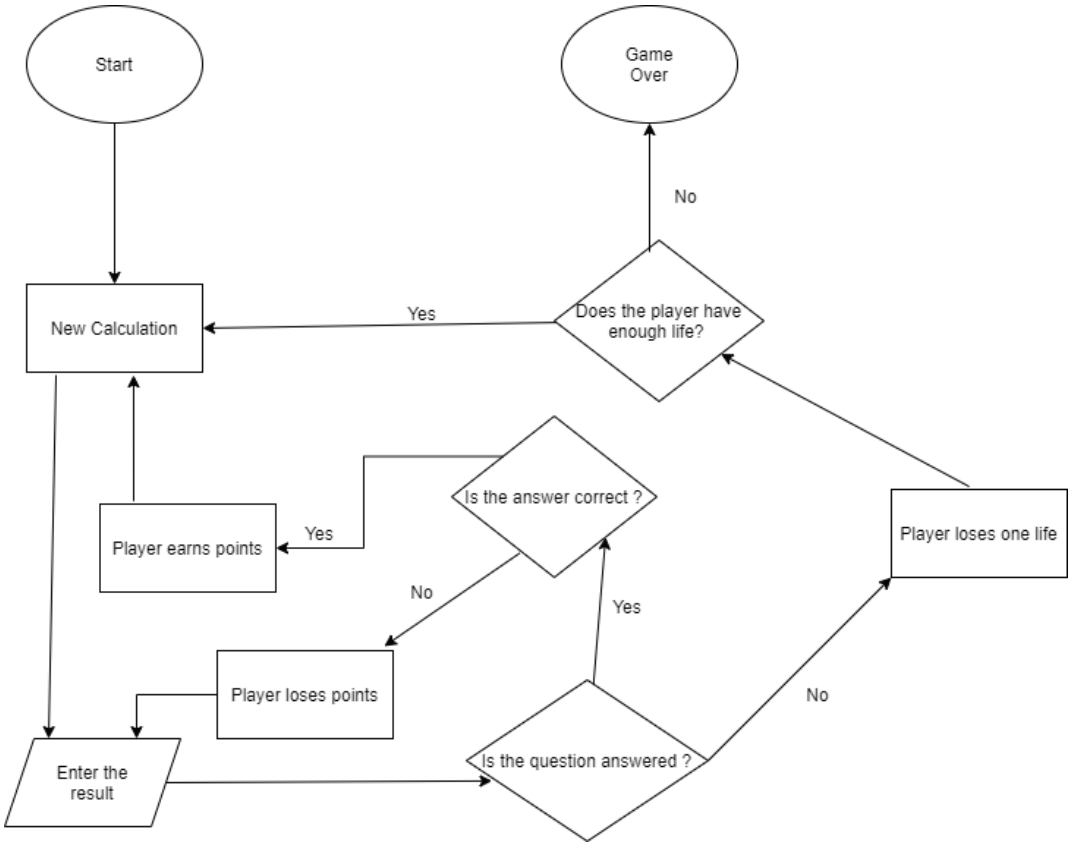


Figure 7. Flowchart representation of the math game.

Attention

The attention skills of a person can be analyzed in different branches. The design process of the game in the attention category focuses on the "selective attention" branch. In the game view, two pasteboards and two different random generated strings are written on both. Players are expected to select the pasteboard with left and right arrow keys which contain the uppercase letter string written on it. After the player chooses a pasteboard, new random strings are generated and written on the pasteboards again, this cycle continues until the time is up, which is displayed on the top right. If a player makes five mistakes back to back, the game will be ended directly to limit the player and increase the quality of the data collected; on the other hand, three seconds will be added to the time limit for every five correct answers given back to back. The main goal of the game is to spot the string that includes the uppercase letter and select the postcard that the string is written on with the correct arrow key as fast as possible for each iteration. An example in-game visual from the attention game is given in Figure 8.

Figure 8. In-game screenshot of two postcards with strings written on them.

The parameters used in evaluating players' performance in visual search tests are the number of correct/incorrect answers and the reaction time given to these answers separately. Reaction time (r_i), is the time passed between the appearance of the strings and the input taken by the player. In the evaluating the player score, the arithmetic mean of those reaction times (ar_i) is used.

For the playthrough i, the attention score (as_i) is computed as given in Equation 4,

$$as_i = ar_i^{-1} r_m + co_i\, co_{m-} w_i w_m \quad (4)$$

where $_{ari}$ is the average reaction time (arithmetic mean of $_{ri}$ values), w_i is the number of wrong answers and co_i is the total number of correct answers. The multipliers r_m, w_m, and co_m are used with the values of 2.5, 2.5, 5, respectively.

Feedback Mechanism

The feedback mechanism is a simple yet informative system. It is divided into two sub-mechanisms where one is visible to the user in the statistics menu. At the same time, the other is based on an e-mail system to inform the relatives responsible for the user.

The first sub-mechanism is percentage-based comparisons displayed individually in the statistics menu for every game. Three statistics are displayed to the user in this branch. The first statistic compares a user's last score to the average of their previous scores. An example message is "Your last score is 21.4% better than your average score". The second statistic is similar, but instead of comparing to the user's past scores, the last score is compared to other users' average scores in the same age range. The final displayed statistic is a comparison between the user's average score and the average of the other users' scores within the same age range. These messages help users see if they are performing better or worse compared to themselves and others in their own age interval.

The second sub-mechanism, which discriminates our study from the others, is a function that creates a line chart using all past scores of the user in the selected game category and sends an e-mail to their registered relative. It provides an easy-to-interpret visual to see whether the user's performance decreases or increases.

PRELIMINARY RESULTS

Preliminary tests were conducted on eight individuals between ages 25-64 and have no confirmed diagnosis of any cognitive problems. A developed application was sent to those individuals, and there were no restrictions for accessibility to any game; players could play each game in their

Table 2. The data collected during the preliminary tests

	Attention	Math	Memory	Problem solving
Highest score	350.38	407	55157.5	270.93
Lowest score	34.42	118	1182.14	79.11
Average score	112.96	244.23	21383.84	139.38
Total games played	35	47	25	71

environment for an unlimited number of times to produce more data. Table 2 is divided by category names since there is only one game for each category.

Although each game has its tutorial section, it is not possible to know if users followed the tutorial on their first playthrough to get an understanding of the game or not. To take problem-solving game as an example, it is not correct to say the person who scored 139.38 might be showing a sign of a cognitive problem from statistics at first glance. This low score might be the individual's first game, and the individual might have spent time understanding of the game or made a mistake while trying.

It is incorrect to bind a score to a sign of a problem in short-term statistics observation. This study requires regular data collection with minimum technical issues by the users' side in the long term to interpret the data.

CONCLUSION

This study proposes the CADTS system, which includes a game under the categories of memory, problem solving, mathematics and attention. For each game, the game's structure and the scoring function is explicitly described in detail. CADTS has been tested by people of various ages, and a small collection of data is recorded in the target database, which gives a clue about skill distribution in categories and ages. In commercial applications similar to CADTS, in terms of the structure of the games, the scoring systems are not explicitly stated. Thus, with this study, an insight into a possible scoring system for different types of games is contributed. Our results show that the scores of people with no known cognitive illnesses are relatively close. However, for a generalized conclusion, more data is required. The controversy over the efficacy of brain training applications [24,25] is known and is expected to remain a research area.

With the consultation of neuroscientists, current games can be further improved for more accurate results, and new games can be implemented. The current comparing system relies on the users' performances and makes comparisons between them; this system can be further improved with a detailed classification mechanism that uses more data.

Based on these data, this study could be progressed into a product that can catch signs of dementia and some of the other issues related to those skills and also keep track of the diagnosed people. With all the improvements mentioned, this study could be used in professional applications in the future.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Lumosity - HCP. Available at: https://www.lumosity.com/hcp/research/completed, Accessed on May 06, 2023.

[2] Taylor-Rowan M, Nafisi S, Owen R, Duffy R, Patel A, Burton J, et al. Informant-based screening tools for dementia: An overview of systematic reviews. Psychol Med 2023;53:580–589.

[3] Haas, J. K. (2014). A history of the unity game engine (Graduation thesis). Worcester, USA: Worcester Polytechnic Institute.

[4] Firebase. Available at: https://firebase.google.com/ Accessed on Jan 08, 2020.

[5] Global population age structures and sustainable development. Available at: Available at: https://www.un.org/En/Development/Desa/Population/Events/Pdf/Expert/25/2016-EGM_Nicole%20Mun%20Sam%20Lai.pdf Accessed on Jan 08, 2020.

[6] Shute VJ, Moore GR, Wang L. Measuring problem solving skills in plants vs. zombies 2. International Educational Data Mining Society, 8th, Madrid, Spain, Jun 26-29, 2015. [CrossRef]

[7] Shute VJ. Stealth assessment in computer-based games to support learning. Comput Games Instruct 2011;55:503–524.

[8] Huizenga J, Admiraal W, Akkerman S, ten Dam G. Cognitive and affective effects of learning history by playing a mobile game. Leading Issues in Games-Based Learning Research, 2011:1–11.

[9] Shute VJ, Rieber L, Van Eck R. Games... and... learning. Trends Issues Instruct Design Technol 2011;3:1–31.

[10] Shute VJ, Wang L, Greiff S, Zhao W, Moore G. Measuring problem solving skills via stealth assessment in an engaging video game. Comput Hum Behav 2016;63:106–117. [CrossRef]

[11] Raven JC. Standardization of progressive matrices, 1938. Br J Med Psychol 1941;19:137–150. [CrossRef]

[12] Wüstenberg S, Greiff S, Funke J. Complex problem solving-More than reasoning? Intelligence 2012;40:1–14. [CrossRef]

[13] Posner MI. Orienting of attention. Q J Exp Psychol 1980;32:3–25. [CrossRef]
[14] MacLeod JW, Lawrence MA, McConnell MM, Eskes GA, Klein RM, Shore DI. Appraising the ANT: Psychometric and theoretical considerations of the Attention Network Test. Neuropsychology 2010;24:637. [CrossRef]
[15] Folsom R, Levin P. Conners' Continuous Performance Test. In: Volkmar, F.R. (eds) Encyclopedia of Autism Spectrum Disorders. New York, USA: Springer; 2013. [CrossRef]
[16] Treisman AM. Strategies and models of selective attention. Psychol Rev 1969;76:282-299. [CrossRef]
[17] Visual Attention Lab, search.bwh.harvard.edu/new research.html Accessed on Jan 08, 2020.
[18] Müller HJ, Krummenacher J. Visual search and selective attention. Vis Cogn 2006;14:389–410. [CrossRef]
[19] Wen ZE, Juffs A, Winke P. Measuring working memory. In The Routledge handbook of second language acquisition and language testing (pp. 167–176). London: Routledge; 2020. [CrossRef]
[20] Richardson-Klavehn A, Bjork RA. Measures of memory. Ann Rev Psychol 1988;39:475–543. [CrossRef]
[21] Takeuchi H, Nagase T, Taki Y, Sassa Y, Hashizume H, Nouchi R, et al. Effects of fast simple numerical calculation training on neural systems. Neural Plast 2016;2016:5940634. [CrossRef]
[22] Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM. A systematic literature review of empirical evidence on computer games and serious games. Comput Educ 2012;59:661–686. [CrossRef]
[23] Hampshire A, Sandrone S, Hellyer PJ. A large-scale, cross-sectional investigation into the efficacy of brain training. Front Hum Neurosci 2019;13:221. [CrossRef]
[24] Bainbridge K, Mayer RE. Shining the light of research on Lumosity. J Cogn Enhanc 2018;2:43–62. [CrossRef]
[25] Shute VJ, Ventura M, Ke F. The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills. Comput Educ 2015;80:58–67. [CrossRef]
[26] Gick ML Problem-solving strategies. Educ Psychol 1986;21:99–120. [CrossRef]
[27] Fua K, Gupta S, Pautler D, Farber I. Designing serious games for elders.» In Conference: Proceedings of the 8th International Conference on the Foundations of Digital Games (FDG 2013) At: Chania, Crete, Greece, pp. 291–297, 2013.

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1169–1182, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00095

Research Article

Performance estimation of honeynet system for network security enhancement via copula linguistic

Muhammad SALIHU ISA[1,2], Jinbiao WU[1,*], Ibrahim YUSUF[3]

[1]*School of Mathematics and Statistics, Central South University, Changsha 410083, China*
[2]*Department of Mathematics, Yusuf Maitama Sule University, Kano, 700282, Nigeria*
[3]*Department of Mathematical Sciences, Bayero University, Kano, 700006, Nigeria*

ARTICLE INFO

Article history
Received: 31 January 2023
Revised: 18 April 2023
Accepted: 04 April 2023

Keywords:
Cyber; Honeynet; Intrusion detection system (IDS); Network; Security

ABSTRACT

Honeypots are computer systems that deceive cyber attackers into believing they are ordinary computer systems designed for invasion, when in fact they are primarily designed to collect data about attack methods, resulting in better protection and defense against malicious actors. As a result, developing reliability metrics for measuring the performance, strength, and effectiveness of honeypot deception is advantageous. Despite extensive and mature research on honeynet system, reliability modeling, analysis and performance prediction and evaluation, based on copula techniques for accurately testing, estimating and optimizing the overall performance of honeynet systems remain lacking. To start, a copula approach for analyzing and optimizing the performance of honeynet systems was proposed. Any honeynet system's performance can be classified based on its availability, dependability and profit generated. As a result, the current paper sought to investigate the performance of a multi-state honeynet system in terms of availability, dependability and expected profit. This paper examines two types of repairs. Type I repairs are known as general repairs and they are used to recover from a partial or non-lethal failure to a perfect state, whereas Type II repairs are known as copula repairs they are used to recover from a complete or lethal failure to a perfect state. For the sake of generality, the supplementary variable technique and Laplace transforms were used to develop the performance models that are essential to this research, such as availability, reliability, mean time to failure (MTTF), sensitivity and profit function. The models' numerical validation was fully carried out. The results are shown in tables and figures, enabling us to draw the conclusion that Type II repair is a superior repair policy. Type II repair, according to the findings, can more accurately portray system structure and states while still allowing for efficient assessment.

Cite this article as: Salihu Isa M, Wu J, Yusuf I. Performance estimation of honeynet system for network security enhancement via copula linguistic. Sigma J Eng Nat Sci 2024;42(4):1169–1182.

***Corresponding author.**
*E-mail address: wujinbiao@csu.edu.cn

This paper was recommended for publication in revised form by Editor in Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey

INTRODUCTION

In today's technological and computerized world, the internet stands as an important system for both service providers and consumers. The accuracy and consistent availability of service are critical to a successful of any venture. As a result, it is critical for service providers to safeguard their servers against numerous cyber-attacks. Due to the use of network explosives in recent years, computer systems and the internet have raised numerous security concerns. Computer crime is constantly on the rise. Based on known facts, mitigation strategies are devised for safeguard and prevention of attacks. Security preventive and corrective measures range from keeping intruders away from a network or system, to protecting and preventing internet communication, to limiting the spread and serious harm caused by computer viruses.

A honeynet is a network that is designed to capture hackers by hosting deliberate vulnerabilities on a decoy server. The primary goal is to put network security to the test by inviting attacks. This method enables security experts to investigate a real attacker's activities and strategies for enhancing network security. The penetration testing of intrusion detection system (IDS) has become a critical component of enterprises that prevents cybercriminal activity by protecting the network, resources and sensitive data. So far, several ways to thwarting harmful activity have been presented and implemented.

Albahar et al. (2020), Althubiti et al. (2018) and AlHamouz et al. (2017) an intrusion detection system (IDS) detects intrusions in two ways: signature-based IDS and anomaly-based IDS. Signature-based detection looks for a "signature" pattern or known attacks. This type of IDS requires regular updates to currently common signatures or identities to ensure that the intruders' database is up to date. However, attackers can change minor details in signatures so that databases do not recognize them. As a result, a new attack type may not be detected because the signature does not exist in the database. Furthermore, the larger the databases, the more processing is required to analyse and verify each connection. In contrast to signature-based IDS, anomaly-based detection is used to detect known and unknown attacks based on learning their behaviour in a computer network by specifying observations that deviate from a basic model and informing the network's administrator to take necessary actions. The ability to detect unknown attacks is the primary advantage of anomaly-based detection.

Many researchers have proposed various intrusion detection systems as a result of their importance. Among these systems, machine learning models, specifically neural networks, can effectively detect malicious network activity by being trained with enough intrusion detection recorded data. Non-neural network machine learning models, such as SVM, have limitations such as low repetition attack detection rates, detection instability, and training process complexity.

Auto encoders and variational auto encoders (VAEs) are two neural network models that have been used for anomaly detection. Auto encoders are made up of sequentially linked encoder and decoder networks. An encoder can compress the input data, and a decoder can reconstruct the input data. Auto encoders try to reduce reconstruction error (the difference between decoder output and original input). To detect anomalies, this error is used as an anomaly score. Small reconstruction errors are associated with normal data, whereas larger reconstruction errors are associated with anomalous data.

Related Work

Researchers have put in significant effort to developed methods of defending, protecting, and improving the honeypot's security system. To cite few, Agrawal and Tapaswi (2017) proposed intrusion detection mechanism called honeypot intrusion detection system meant for detecting and preventing external and internal malicious users gaining access to wireless network. Kondra et al. (2016) developed an intrusion detection technique which will extract the details of the attacker. Naik et al. (2021) proposed method that allow honeypot to explore and estimate the malicious users' fingerprint using fuzzy inference and principal components analysis. Paryathia et al. (2021) analyzes the technique of anti-identification thinking, signature, and the theoretical basis of game. Isa et al. (2023) explore on reliability analysis of computer network which comprises of three subsystems: router, workstation and hub. Yusuf et al. (2021) consider a distributed system with five standby subsystems A (the clients), B (two load balancers), C (two distributed database servers), D (two mirrored distributed database serves) and E (centralized database server) is considered arranged as series-parallel system. Kasongo and Sun (2020) created five supervised models using a filter-based information reduction method and compared their performance on the UNSW-NB 15 dataset, the UNSW-NB15 is a network intrusion dataset that contains nine different attacks, includes DoS, worms, Backdoors, and Fuzzers. The dataset contains raw network packets. Disha and Waheed (2022) proposed machine learning techniques for intrusion detection systems and analyzed model performance by training and testing the Long-Short Term Memory, Multilayer Perceptron, Decision Tree, Gradient Boosting Tree, AdaBoost and Gated Recurrent Unit for the binary classification task. Isa et al. (2021) investigate the performance measures of network with transparent bridge as follows 1-out-of-2: G, 2-out-of-3: F, a bridge unit and 3-out-of-5: G schemes.

Arqub and Hammour (2014) explore on continuous genetic algorithm as an efficient solver for systems of second-order boundary value problems where smooth solution curves were used throughout the evolution of the algorithm to obtain the required nodal values of the unknown variables, Aydin et al. (2022) estimate the coliform values of the Tekkekoy deep sea discharge system, which is chosen as

an application area, by using a radial-based artificial neural network structure, Sekerci and Aydin (2022) writes on production-distribution network system for a company, which is active in producing bottled natural spring water was established. In Kenan et al. (*2021)* classification algorithms were used to classify electromyography and depth sensor data. Tolga and Ali (2022) use artificial neural networks to predict the risk size of the BLEVE event. Bakar and Murat (2022) examine the net single premiums of multiple life annuities using stochastic rates of return and dynamic life table under the assumption of dependency of spouses' future lifetimes. In pology, Adem (2023) introduces the concept of intuitionistic fuzzy hyper soft. Certain properties of intuitionistic fuzzy hyper soft (IFH) topology are investigated, including the IFH basis, IFH subspace, IFH interior, and IFH cloure. Arqul et al. (2021) use extended reproducing kernel Hilbert space technique to analyse and numerically solve fuzzy fractional differential equations with Atangana-Baleanu-Caputo differential operators. Maryam et al. (2023) investigate codes over the direct product of two finite commutative chain rings. The parity-check matrix's standard form is determined. Alazzam et al. (2020) examined the performance of IDS using a binary classifier called Decision Tree (DT). Belgrana et al. (2021) suggested a condensed nearest neighbors neural network to reduce feature dimensionality and computational time, as well as a radial basis function neural network to achieve performance learning on the network security laboratory-knowledge discovery in databases (NSL-KDD) dataset. Gu and Lu (2021) suggested an effective solution for intrusion detection that combines SVM with the Nave Bayes algorithm to differentiate intrusion and normal cases. Lee et al. (2020) offered a hybrid technique in which the authors recommended a deep sparse auto encoder for feature selection in the data pre-processing step. Isa et al. (2022) Explore on reliability analysis of computer network which comprises of three subsystems: router, workstation and hub. Mauro et al. (2020) gave an experimental study for Network Intrusion Mitigation application of Neural Network methods. Arqub et al. (2014) publish an article on numerical approximation of solutions with Troesch's and Bratu's problems.

Kelly et al. (2020) emphasize the necessity of using publicly accessible vulnerability intelligence information and indicators of compromise obtained via honeypots to inform an organization's Situational Awareness operations, using a similar methodology as in this paper. Sethia and Jeyasekar (2019) developed a honeypot as a security measure to protect an establishment from the detrimental and malicious acts of malwares by examining the honeypot's network logs. Arqub et al. (2021) consider a numerical approach to solve groups of fuzzy fractional integrodifferentials (FFIDEs) with Atangana–Baleanu–Caputo (ABC) fractional distributed order derivatives. The solution-based approach lies in generating infinite orthogonal basis from kernel functions, where an uncertain condition is fulfilled.

Numerous studies in the field of reliability engineering have shown that effective performance analysis can help to avoid disasters, protection, safety and save time, money, or both. To cite few, Xie et al. (2021) investigated and examined the performance of a safety system that is vulnerable to cascading failures that cause the appearance of further failures. In the paper, a unique technique for mitigating and preventing cascading failure is provided. Xie et al. (2019) suggested performance and an approximation approach for medium-frequency hazardous failures in safety instrumental systems prone to cascade failures. Yusuf et al. (2020) analyzed the performance of computer system using copula linguistic. Colledani et al. (2019) offer a method for evaluating the performance of unstable manufacturing systems that takes into account unknown machine reliability predictions.

Reliability modeling or analysis and performance prediction and evaluation, based on copula techniques for accurately testing, estimating and optimizing the overall performance of honeynet systems remain lacking, copula technique is a powerful tool for describing variable dependence and has received much attention in a variety of fields of study. Numerous researchers have used the copula method to explore the performance of complex repairable systems and have reported improved operational performance. However, the issue of whether copula-based reliability, performance, strength, and effectiveness of the given honeynet system has not been thoroughly investigated. This motivates us to evaluate and investigate the honeypot's availability and performance analysis using the gumbel hougaard family copula.

DESCRIPTION AND NOTATION OF THE SYSTEM

Notations

q: Variable representing time.

s: representing variable of Laplace transform

m_1: stand for rate of failure of unit in production subsystem

m_2: stand for rate of failure of honeypot in honeypot subsystem

m_{s1}: stand for rate of failure of switch I

m_{s2}: stand for rate of failure of switch II

m_r: stand for rate of failure of router

m_h: stand for rate of failure of honey sensor

$l_1(x)$: stand for rate of repair by general repair of unit in production subsystem

$l_2(y)$: stand for rate of repair by general repair of honeypot in honeypot subsystem

$\eta_o(x)$: stand for rate of repair by copula of unit in production subsystem

$\eta_o(y)$: stand for rate of repair by copula of honeypot in honeypot subsystem

$\eta_o(n)$: stand for rate of repair by copula of switch I

$\eta_o(z)$: stand for rate of repair by copula of switch II

$\eta_o(r)$: stand for rate of repair by copula of router

$\eta_o(z)$: stand for rate of repair by copula of honey sensor

$H_i(t)$: stand for chance of the system sojourning in S_i state at instants for i =0 to 14.

$\bar{H}(s)$: stand for Laplace transformation of state transition probability $H(t)$.

H_i (x, q): stand for chance of the system sojourning in S_i with x variable of repair and variable time q.

P_i (y_2, t): stand for chance of the system sojourning in S_i with y_2 variable of repair variable y_1 and variable time t.

P_i (y_3, t): stand for chance of the system sojourning in S_i with y_3 variable of repair variable y_1 and variable time t.

P_i (y_4, t): stand for chance of the system sojourning in S_i with y_4 variable of repair variable y_1 and variable time t.

P_i (y_5, t): stand for chance of the system sojourning in S_i with y_5 variable of repair variable y_1 and variable time t.

$E_p(t)$: Expected profit during the time interval [0, t)

Z_1, Z_2: Revenue and service cost per unit time, respectively.

$m_0(x)$: The expression of joint probability according to Gumbel-Hougaard family Copula definition is given as: $c_\theta(u_1(x), u_2(x)) = exp\left(x^\theta + \{log\,\phi\,(x)^\theta\}^{\frac{1}{\theta}}\right)$, $1 \leq \theta \leq \infty$. Where $\mu_1 = \phi(x)$ and $u_2 = e_x$.

System Description

The diagram, depicted in Figure 1, portrays a Honeynet system that implements Gen III honeynet solution architecture. The system consists of system users that include an attacker on one side, accessing production systems network via the Internet, a router, a honeynet sensor called a honeywall gateway, a real service network of production systems (system-1, system-2 and system-3) and the network of honeypots with data capture capability (OS-1, OS-2 and OS-3).

The router typically implements a hidden firewall , which serves as first access control mechanism. The production system network applies a honeynet security technology. The honeynet implements a honeynet sensor which is the most important tool in the entire honeynet solution. The honeyewall is a computer server that serves as a layer 2 gateway device to supervise outbound data and separate the honeynet from other production systems. The honeynet sensor supports interception of SSL connections and make decision about the incoming traffic into the system. It determines if the traffic is malicious and thus redirect it to a honeypots or it is valid and thus redirect it to the real production system. Ultimately, the honeynet sensor performs three essential functions, viz: data control, which involves controlling the flow of data so that the attacker does not realize being in the honeynet and ensuring that the honeynet system is not used to attack other systems in the event of system compromise; data capture, which involves capturing all the data regarding movements and actions within the honeynet; and data collection, which involves the ability to securely tranfer all the captured data to a central database/log service, also implemented within the honeynet sensor. Furthermore, the honeypots are computer systems that duplicate and disguise themselves as real production systems in order to lure an attacker. The honeypots are controlled by the honeywall. They typically implements Sebek/Qebek monitoring tool. When the honeypots receive a malicious request from attacker, the systems invisibly monitor and capture

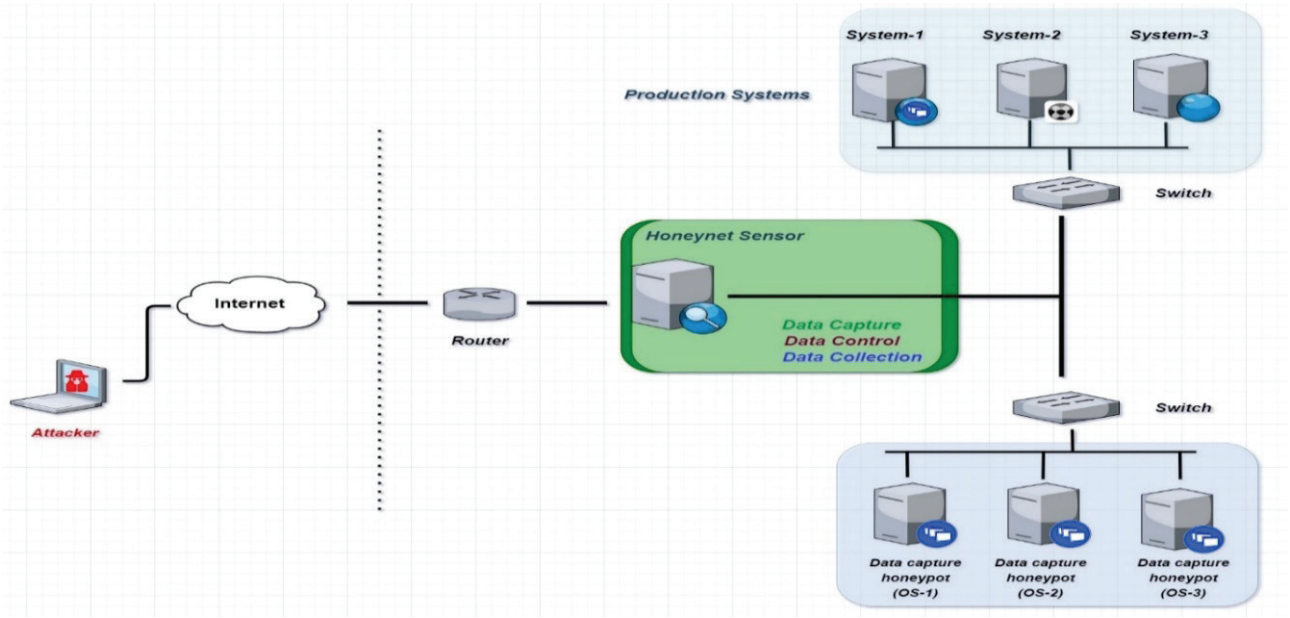


Figure 1. Reliability block diagram of the honeynet system.

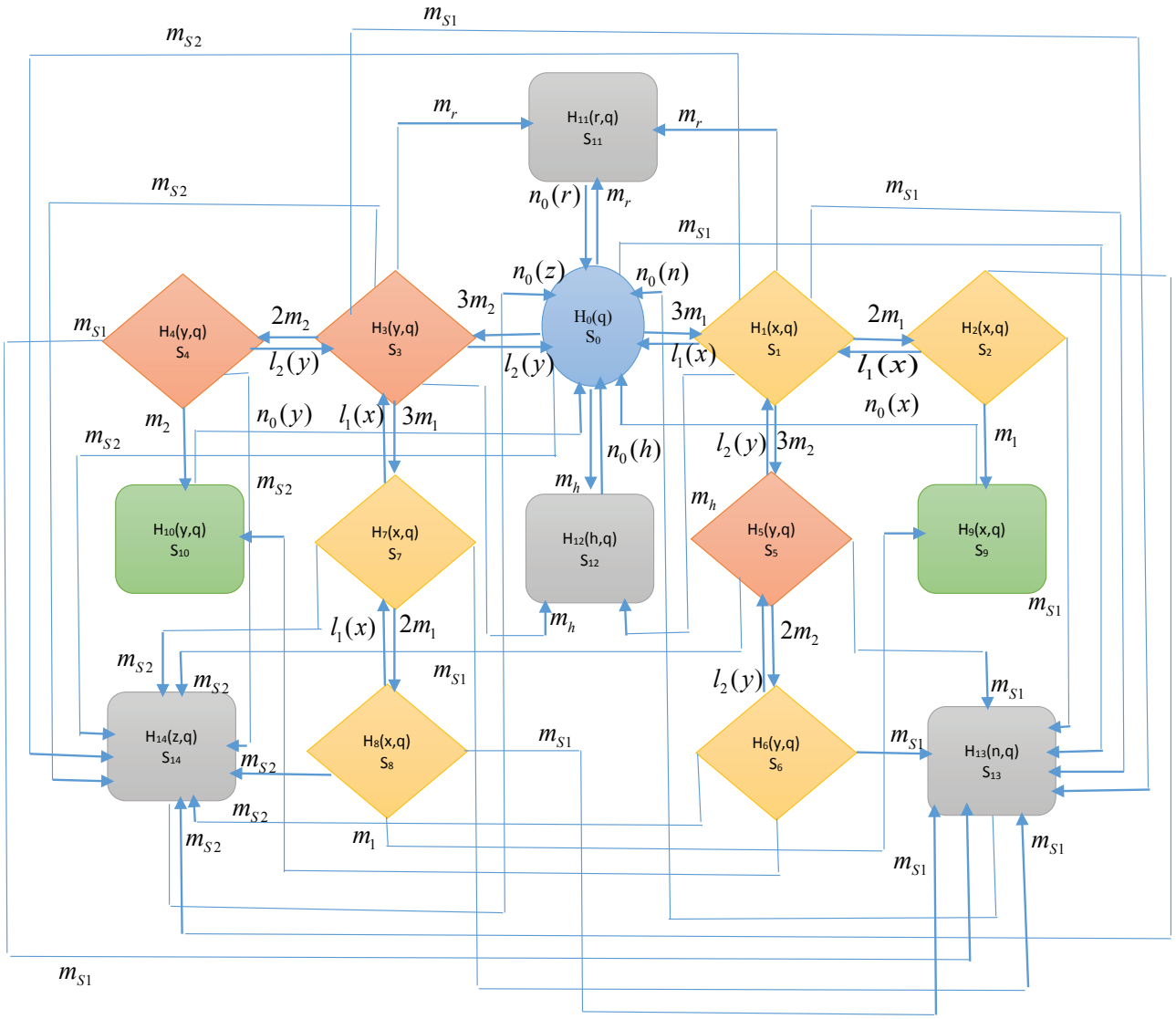


Figure 2. Transition diagram of the honeynet system.

activities of the attacker in the honeypots and send the captured data to the central log in the honeynet sensor for analysis.

HONEYNET MODEL FORMULATION

The supplementary variable technique and Laplace transforms were used to create reliability models for honeynet system analysis. A probabilistic approach was used to generate the differential equations from the transition diagram above. These equations were then solved using initial and boundary conditions to obtain steady state probabilities, which serve as the basis for the development of reliability models.

The following partial differential equations are obtained via Figure 2:

$$\left(\frac{\partial}{\partial q}+3m_1+3m_2+m_r+m_h+m_{s_1}+m_{s_2}\right)H_0(q)=\int_0^\infty l_1(x)H_1(x,q)dq+\int_0^\infty l_2(y)H_3(y,q)dq+\int_0^\infty n_0(z)H_{14}(z,q)dq+\int_0^\infty n_0(x)H_9(x,q)dq+\int_0^\infty n_0(y)H_{10}(y,q)dq+\int_0^\infty n_0(r)H_{11}(r,q)dq+\int_0^\infty n_0(h)H_{12}(h,q)dq+\int_0^\infty n_0(n)H_{13}(n,q)dq \quad (1)$$

$$\left(\frac{\partial}{\partial q}+\frac{\partial}{\partial x}+2m_1+3m_2+m_r+m_h+m_{s_1}+m_{s_2}+l_1(x)\right)H_1(x,q)=0 \quad (2)$$

$$\left(\frac{\partial}{\partial q}+\frac{\partial}{\partial x}+m_1+m_{s_1}+m_{s_2}+l_1(x)\right)H_2(x,q)=0 \quad (3)$$

$$\left(\frac{\partial}{\partial q}+\frac{\partial}{\partial y}+3m_1+2m_2+m_r+m_h+m_{s_1}+m_{s_2}+l_2(y)\right)H_3(y,q)=0 \quad (4)$$

$$\left(\frac{\partial}{\partial q}+\frac{\partial}{\partial y}+m_2+m_{s_1}+m_{s_2}+l_2(y)\right)H_4(y,q)=0 \quad (5)$$

$$\left(\frac{\partial}{\partial q}+\frac{\partial}{\partial y}+2m_2+m_{s_1}+m_{s_2}+l_2(y)\right)H_5(y,q)=0 \quad (6)$$

$$\left(\frac{\partial}{\partial q}+\frac{\partial}{\partial y}+m_2+m_{s_1}+m_{s_2}+l_2(y)\right)H_6(y,q)=0 \quad (7)$$

$$\left(\frac{\partial}{\partial q}+\frac{\partial}{\partial x}+2m_1+m_{s_1}+m_{s_2}+l_1(x)\right)H_7(x,q)=0 \quad (8)$$

$$\left(\frac{\partial}{\partial q}+\frac{\partial}{\partial x}+m_1+m_{s_1}+m_{s_2}+l_1(x)\right)H_8(x,q)=0 \quad (9)$$

$$\left(\frac{\partial}{\partial q}+\frac{\partial}{\partial x}+n_0(x)\right)H_9(x,q)=0 \quad (10)$$

$$\left(\frac{\partial}{\partial q}+\frac{\partial}{\partial y}+n_0(y)\right)H_{10}(y,q)=0 \quad (11)$$

$$\left(\frac{\partial}{\partial q}+\frac{\partial}{\partial r}+n_0(r)\right)H_{11}(r,q)=0 \quad (12)$$

$$\left(\frac{\partial}{\partial q}+\frac{\partial}{\partial h}+n_0(h)\right)H_{12}(h,q)=0 \quad (13)$$

$$\left(\frac{\partial}{\partial q}+\frac{\partial}{\partial n}+n_0(n)\right)H_{13}(n,q)=0 \quad (14)$$

$$\left(\frac{\partial}{\partial q}+\frac{\partial}{\partial z}+n_0(z)\right)H_{14}(z,q)=0 \quad (15)$$

Boundary conditions

$$H_1(0,q)=3m_1H_0(q) \quad (16)$$

$$H_2(0,q)=2m_1H_1(0,q) \quad (17)$$

$$H_3(0,q)=3m_2H_1(q) \quad (18)$$

$$H_4(0,q)=2m_2H_3(0,q) \quad (19)$$

$$H_5(0,q)=3m_2H_1(0,q) \quad (20)$$

$$H_6(0,q)=2m_2H_5(0,q) \quad (21)$$

$$H_7(0,q)=3m_1H_3(0,q) \quad (22)$$

$$H_8(0,q)=2m_1H_7(0,q) \quad (23)$$

$$H_9(0,q)=m_1\left(H_2(0,q)+H_8(0,q)\right) \quad (24)$$

$$H_{10}(0,q)=m_2\left(H_4(0,q)+H_6(0,q)\right) \quad (25)$$

$$H_{11}(0,q)=m_r\left(H_0(q)+H_1(0,q)+H_3(0,q)\right) \quad (26)$$

$$H_{12}(0,q)=m_h\left(H_0(q)+H_1(0,q)+H_3(0,q)\right) \quad (27)$$

$$H_{13}(0,q)=m_{s_1}\begin{pmatrix}H_0(q)+H_1(0,q)+H_2(0,q)+H_3(0,q)+H_4(0,q)\\+H_5(0,q)+H_6(0,q)+H_7(0,q)+H_8(0,q)\end{pmatrix} \quad (28)$$

$$H_{14}(0,q)=m_{s_2}\begin{pmatrix}H_0(q)+H_1(0,q)+H_2(0,q)+H_3(0,q)+H_4(0,q)\\+H_5(0,q)+H_6(0,q)+H_7(0,q)+H_8(0,q)\end{pmatrix} \quad (29)$$

Initial condition

$$H_t(0)=\begin{cases}1, & t=0\\ 0, & t\neq 0\end{cases}$$

MODEL'S SOLUTION

The Laplace transformation of equations (1) to (29) with the help of initial condition to obtain:

$$\begin{aligned}&\left(s+3m_1+3m_2+m_r+m_h+m_{s_1}+m_{s_2}\right)\overline{H}_0(\mathrm{s})=1+\int_0^\infty l_1(x)\overline{H}_1(x,\mathrm{s})dx\\&+\int_0^\infty l_2(y)\overline{H}_3(y,\mathrm{s})dy+\int_0^\infty n_0(x)\overline{H}_9(x,s)dr+\int_0^\infty n_0(y)\overline{H}_{10}(y,s)dy\\&+\int_0^\infty n_0(r)\overline{H}_{11}(r,s)dr+\int_0^\infty n_0(h)\overline{H}_{12}(h,s)dr+\int_0^\infty n_0(n)\overline{H}_{13}(n,s)dn\\&+\int_0^\infty n_0(z)\overline{H}_{14}(z,s)dz\end{aligned} \quad (30)$$

$$\left(s+\frac{\partial}{\partial x}+2m_1+3m_2+m_r+m_h+m_{s_1}+m_{s_2}+l_1(x)\right)\overline{H}_1(x,\mathrm{s})=0 \quad (31)$$

$$\left(s+\frac{\partial}{\partial x}+m_1+m_{s_1}+m_{s_2}+l_1(x)\right)\overline{H}_2(x,\mathrm{s})=0 \quad (32)$$

$$\left(s+\frac{\partial}{\partial y}+3m_1+2m_2+m_r+m_h+m_{s_1}+m_{s_2}+l_2(y)\right)\overline{H}_3(y,s)=0 \quad (33)$$

$$\left(s+\frac{\partial}{\partial y}+m_2+m_{s_1}+m_{s_2}+l_2(y)\right)\overline{H}_4(y,s)=0 \quad (34)$$

$$\left(s+\frac{\partial}{\partial y}+2m_2+m_{s_1}+m_{s_2}+l_2(y)\right)\overline{H}_5(y,s)=0 \quad (35)$$

$$\left(s+\frac{\partial}{\partial y}+m_2+m_{s_1}+m_{s_2}+l_2(y)\right)\overline{H}_6(y,s)=0 \quad (36)$$

$$\left(s+\frac{\partial}{\partial x}+2m_1+m_{s_1}+m_{s_2}+l_1(x)\right)\overline{H}_7(x,s)=0 \quad (37)$$

$$\left(s+\frac{\partial}{\partial x}+m_1+m_{s_1}+m_{s_2}+l_1(x)\right)\overline{H}_8(x,s)=0 \quad (38)$$

$$\left(s+\frac{\partial}{\partial x}+n_0(x)\right)\overline{H}_9(x,s)=0 \quad (39)$$

$$\left(s+\frac{\partial}{\partial y}+n_0(y)\right)\overline{H}_{10}(y,s)=0 \quad (40)$$

$$\left(s+\frac{\partial}{\partial r}+n_0(r)\right)\overline{H}_{11}(r,s)=0 \quad (41)$$

$$\left(s+\frac{\partial}{\partial h}+n_0(h)\right)\overline{H}_{12}(h,s)=0 \quad (42)$$

$$\left(s+\frac{\partial}{\partial n}+n_0(n)\right)\overline{H}_{13}(n,s)=0 \quad (43)$$

$$\left(s+\frac{\partial}{\partial z}+n_0(z)\right)\overline{H}_{14}(z,s)=0 \quad (44)$$

Boundary conditions

$$\overline{H}_1(0,s)=3m_1\overline{H}_0(s) \quad (45)$$

$$\overline{H}_2(0,s)=2m_1\overline{H}_1(0,s) \quad (46)$$

$$\overline{H}_3(0,s)=3m_2\overline{H}_0(s) \quad (47)$$

$$\overline{H}_4(0,s)=2m_2\overline{H}_3(0,s) \quad (48)$$

$$\overline{H}_5(0,s)=3m_2\overline{H}_1(0,s) \quad (49)$$

$$\overline{H}_6(0,s)=2m_2\overline{H}_5(0,s) \quad (50)$$

$$\overline{H}_7(0,s)=3m_1\overline{H}_3(0,s) \quad (51)$$

$$\overline{H}_8(0,s)=2m_1\overline{H}_7(0,s) \quad (52)$$

$$\overline{H}_9(0,s)=m_1\left(\overline{H}_2(0,s)+\overline{H}_8(0,s)\right) \quad (53)$$

$$\overline{H}_{10}(0,s)=m_2\left(\overline{H}_4(0,s)+\overline{H}_6(0,s)\right) \quad (54)$$

$$\overline{H}_{11}(0,s)=m_r\left(\overline{H}_0(s)+\overline{H}_1(0,s)+\overline{H}_3(0,s)\right) \quad (55)$$

$$\overline{H}_{12}(0,s)=m_h\left(\overline{H}_0(s)+\overline{H}_1(0,s)+\overline{H}_3(0,s)\right) \quad (56)$$

$$\overline{H}_{13}(0,s)=m_{s_1}\left(\begin{matrix}\overline{H}_0(s)+\overline{H}_1(0,s)+\overline{H}_2(0,s)+\overline{H}_3(0,s)+\overline{H}_4(0,s)\\+\overline{H}_5(0,s)+\overline{H}_6(0,s)+\overline{H}_7(0,s)+\overline{H}_8(0,s)\end{matrix}\right) \quad (57)$$

$$\overline{H}_{14}(0,s)=m_{s_2}\left(\begin{matrix}\overline{H}_0(s)+\overline{H}_1(0,s)+\overline{H}_2(0,s)+\overline{H}_3(0,s)+\overline{H}_4(0,s)\\+\overline{H}_5(0,s)+\overline{H}_6(0,s)+\overline{H}_7(0,s)+\overline{H}_8(0,s)\end{matrix}\right) \quad (58)$$

Condition of Initials

$H_0(0)=1$, but other state transition probability is 0 at this time. Therefore, we have the following solution. (59)

$$\overline{H}_0(s)=\frac{1}{K(s)} \quad (60)$$

$$\overline{H}_1(s)=\frac{3m_1}{K(s)}\left\{\frac{1-\bar{s}_l\left(s+2m_1+3m_2+m_r+m_h+m_{s_1}+m_{s_2}\right)}{s+2m_1+3m_2+m_r+m_h+m_{s_1}+m_{s_2}}\right\} \quad (61)$$

$$\overline{H}_2(s)=\frac{6m_1^2}{K(s)}\left\{\frac{1-\bar{s}_l\left(s+m_1+m_{s_1}+m_{s_2}\right)}{s+m_1+m_{s_1}+m_{s_2}}\right\} \quad (62)$$

$$\overline{H}_3(s)=\frac{3m_2}{K(s)}\left\{\frac{1-\bar{s}_l\left(s+3m_1+2m_2+m_r+m_h+m_{s_1}+m_{s_2}\right)}{s+3m_1+2m_2+m_r+m_h+m_{s_1}+m_{s_2}}\right\} \quad (63)$$

$$\overline{H}_4(s)=\frac{6m_2^2}{K(s)}\left\{\frac{1-\bar{s}_l\left(s+m_2+m_{s_1}+m_{s_2}\right)}{s+m_2+m_{s_1}+m_{s_2}}\right\} \quad (64)$$

$$\overline{H}_5(s)=\frac{9m_1m_2}{K(s)}\left\{\frac{1-\bar{s}_l\left(s+2m_2+m_{s_1}+m_{s_2}\right)}{s+2m_2+m_{s_1}+m_{s_2}}\right\} \quad (65)$$

$$\overline{H}_6(s)=\frac{18m_1m_2^2}{K(s)}\left\{\frac{1-\bar{s}_l\left(s+m_2+m_{s_1}+m_{s_2}\right)}{s+m_2+m_{s_1}+m_{s_2}}\right\} \quad (66)$$

$$\overline{H}_7(s)=\frac{9m_1m_2}{K(s)}\left\{\frac{1-\bar{s}_l\left(s+2m_1+m_{s_1}+m_{s_2}\right)}{s+2m_1+m_{s_1}+m_{s_2}}\right\} \quad (67)$$

$$\overline{H}_8(s)=\frac{18m_1^2m_2}{K(s)}\left\{\frac{1-\bar{s}_l\left(s+m_1+m_{s_1}+m_{s_2}\right)}{s+m_1+m_{s_1}+m_{s_2}}\right\} \quad (68)$$

$$\overline{H}_9(s)=\left(\frac{6\xi_1^3+18m_1^3m_2}{K(s)}\right)\left\{\frac{1-\bar{s}_{n_0}(s)}{s}\right\} \quad (69)$$

$$\overline{H}_{10}(s)=\left(\frac{6\xi_2^3+18m_2^3m_1}{K(s)}\right)\left\{\frac{1-\bar{s}_{n_0}(s)}{s}\right\} \quad (70)$$

$$\overline{H}_{11}(s)=\left(\frac{m_r+3m_1m_r+3m_2m_r}{K(s)}\right)\left\{\frac{1-\bar{s}_{n_0}(s)}{s}\right\} \quad (71)$$

$$\overline{H}_{12}(s)=\left(\frac{m_h+3m_1m_h+3m_2m_h}{K(s)}\right)\left\{\frac{1-\bar{s}_{n_0}(s)}{s}\right\} \quad (72)$$

$$\overline{H}_{13}(s)=\frac{1}{K(s)}\begin{pmatrix}m_{s_1}+3m_1m_{s_1}+3m_1^2m_{s_1}+3m_2m_{s_1}+6m_2^2m_{s_1}+9m_1m_2m_{s_1}\\+18m_1m_2^2m_{s_1}+9m_1m_2m_{s_1}+18m_1^2m_2m_{s_1}\end{pmatrix}\left\{\frac{1-\bar{s}_{n_0}(s)}{s}\right\} \quad (73)$$

$$\overline{H}_{14}(s)=\frac{1}{K(s)}\begin{Bmatrix}m_{s_2}+3m_1m_{s_2}+3m_1^2m_{s_2}+3m_2m_{s_2}+9m_1m_2m_{s_2}\\6m_2^2m_{s_2}+9m_1m_2m_{s_2}+18m_1m_2^2m_{s_2}+18m_1^2m_2m_{s_2}\end{Bmatrix}\left\{\frac{1-\bar{s}_{n_0}(s)}{s}\right\} \quad (74)$$

However, K(s) is;

$$K(s)=\left\{\begin{pmatrix}s+3m_1+3m_2+\\m_r+m_h+m_{s_1}+m_{s_2}\end{pmatrix}-\left(\Delta_0+\left[\Delta_1+\Delta_2+\Delta_3\right]\bar{s}_{n_0}(s)\right)\right\} \quad (75)$$

Where

$$\Delta_0=\left\{3m_1\bar{s}_l\left(s+2m_1+3m_2+m_r+m_h+m_{s_1}+m_{s_2}\right)+3m_2\bar{s}_l\left(s+3m_1+m_2+m_r+m_h+m_{s_1}+m_{s_2}\right)\right\},$$
$$\Delta_1=\left\{\left(6m_1^3+18m_1^3m_2\right)+\left(6m_2^3+18m_2^3m_1\right)+\left(m_r+3m_1m_r+3m_2m_r\right)+\left(m_h+3m_1m_h+3m_2m_h\right)\right\},$$
$$\Delta_2=\left\{m_{s_1}+3m_1m_{s_1}+3m_1^2m_{s_1}+3m_2m_{s_1}+6m_2^2m_{s_1}+9m_1m_2m_{s_1}+18m_1m_2^2m_{s_1}+9m_1m_2m_{s_1}+18m_1^2m_2m_{s_1}\right\}$$
and
$$\Delta_3=\left\{m_{s_2}+3m_1m_{s_2}+3m_1^2m_{s_2}+3m_2m_{s_2}+6m_2^2m_{s_2}+9m_1m_2m_{s_2}+18m_1m_2^2m_{s_2}+9m_1m_2m_{s_2}+18m_1^2m_2m_{s_2}\right\}$$

The sum of Laplace transformed state transition probabilities that the system is working are as follows:

$$\overline{H}_{up}(s)=\left[\overline{H}_0(s)+\overline{H}_1(s)+\overline{H}_2(s)+\overline{H}_3(s)+\overline{H}_4(s)+\overline{H}_5(s)+\overline{H}_6(s)+\overline{H}_7(s)+\overline{H}_8(s)\right] \quad (76)$$

$$\overline{H}_{up}(s)=\frac{1}{K(s)}\begin{Bmatrix}1+3m_1\left(\frac{1-\bar{s}_l\left(s+2m_1+3m_2+m_r+m_h+m_{s_1}+m_{s_2}\right)}{s+2m_1+3m_2+m_r+m_h+m_{s_1}+m_{s_2}}\right)+6m_1^2\left(\frac{1-\bar{s}_l\left(s+m_1+m_{s_1}+m_{s_2}\right)}{s+m_1+m_{s_1}+m_{s_2}}\right)+\\3m_2\left(\frac{1-\bar{s}_l\left(s+3m_1+2m_2+m_r+m_h+m_{s_1}+m_{s_2}\right)}{s+3m_1+2m_2+m_r+m_h+m_{s_1}+m_{s_2}}\right)+6m_2^2\left(\frac{1-\bar{s}_l\left(s+m_1+m_{s_1}+m_{s_2}\right)}{s+m_2+m_{s_1}+m_{s_2}}\right)+\\9m_1m_2\left(\frac{1-\bar{s}_l\left(s+2m_2+m_{s_1}+m_{s_2}\right)}{s+2m_2+m_{s_1}+m_{s_2}}\right)+18m_1m_2^2\left(\frac{1-\bar{s}_l\left(s+m_2+m_{s_1}+m_{s_2}\right)}{s+m_2+m_{s_1}+m_{s_2}}\right)+\\9m_1m_2\left(\frac{1-\bar{s}_l\left(s+2m_1+m_{s_1}+m_{s_2}\right)}{s+2m_1+m_{s_1}+m_{s_2}}\right)+18m_1^2m_2\left(\frac{1-\bar{s}_l\left(s+m_1+m_{s_1}+m_{s_2}\right)}{s+m_1+m_{s_1}+m_{s_2}}\right)\end{Bmatrix} \quad (77)$$

$$\overline{H}(s)_{down}(s)=1-\overline{H}_{up}(s)) \quad (78)$$

ANALYSIS OF THE MODEL FOR DIFFERENT CIRCUMSTANCES

Availability Analysis

Suppose that $S_{n_0}(s)=\bar{S}_{\exp[r^\theta+\{\log l(r)\}^\theta]^{1/\theta}}(s)=\frac{\exp[r^\theta+\{\log l(r)\}^\theta]^{1/\theta}}{s+\exp[r^\theta+\{\log l(r)\}^\theta]^{1/\theta}}$, $\bar{S}_l(s)=\frac{l}{s+l}$, assuming failure rates as $m_1 = 0.011$, $m_2 = 0.012$, $m_r = 0.013$, $m_h = 0.014$, $m_{s_1} = 0.015$, $m_{s_1} = 0.016$, $l_1(r) = l_2(r) = 1$. Therefore, substituting those values in equation (77), the subsequent equation follows :

$$\overline{H}_{up}(q)=\begin{bmatrix}-0.000384e^{-1.04200q}-0.000453e^{-1.04300q}\\-0.000557e^{-1.05300q}-0.000547e^{-1.05500q}\\+0.023049e^{-2.78291q}-0.009251e^{-1.17509q}\\+5.615146e^{-1.11551q}+0.988140e^{-0.00925q}\end{bmatrix} \quad (79)$$

When time (t) is use as q = 0, 1,...,10 in equation (79), Table 1 obtained

Table 1. Availability analysis of the system

q	0	1	2	3	4	5	6	7	8	9	10
Availability	1.0000	0.9832	0.9816	0.9795	0.9771	0.9744	0.9717	0.9691	0.9664	0.9637	0.9610

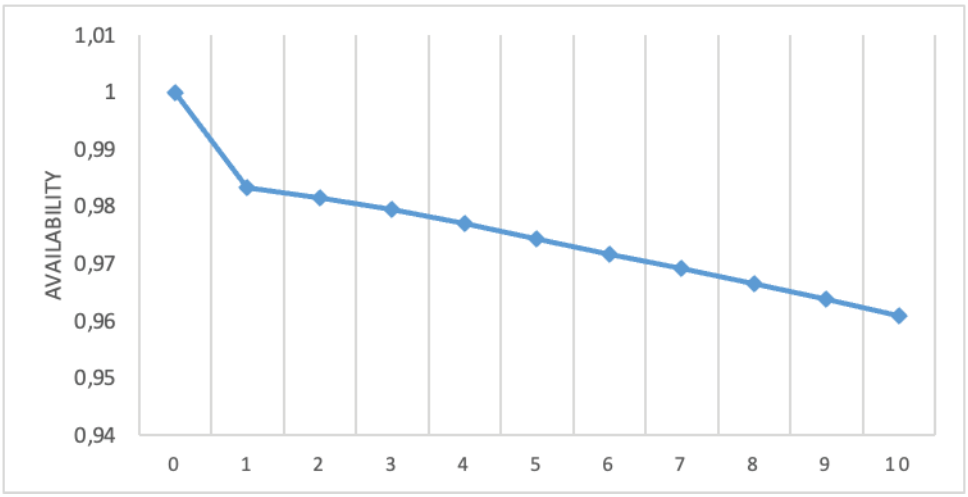


Figure 3. Honeynet availability analysis.

Reliability Analysis

If *l, n* are declared to be zero and values of failure rate as follows: $m_1 = 0.011$, $m_2 = 0.012$, $m_r = 0.013$, $m_h = 0.014$, $m_{s_1} = 0.015$, $m_{s_1} = 0.016$. Then we have,

$$\mathrm{Re}\,l(q) = \begin{bmatrix} 3e^{-0.11600q} + 0.008848e^{-0.02100j} + 0.016054e^{-0.05300q} + 0.010625e^{-0.04300q} \\ -5.052027e^{-0.12700q} + 0.016500e^{-0.05500q} + 3e^{-0.11500q} \end{bmatrix} \quad (80)$$

For q = 0, 1...10 in equation (80),

Mean Time to Failure (MTTF)

Assuming all repairs to zero while s tends zero in equation (77), MTTF expression is obtained as:

$$MTTF = \lim_{s\to 0} \overline{H}_{up}(s) = \frac{1}{s + 3m_1 + 3m_2 + m_r + m_h + m_{s_1} + m_{s_2}} \left\{ 1 + \frac{3m_1}{2m_1 + 3m_2 + m_r + m_h + m_{s_1} + m_{s_2}} + \frac{6m_1^2}{m_1 + m_{s_1} + m_{s_2}} + \frac{18m_1^2 m_2}{m_1 + m_{s_1} + m_{s_2}} + \frac{3m_2}{3m_1 + 2m_2 + m_r + m_h + m_{s_1} + m_{s_2}} + \frac{3m_2^2}{m_2 + m_{s_1} + m_{s_2}} + \frac{9m_1 m_2}{2m_2 + m_{s_1} + m_{s_2}} + \frac{18m_1 m_2^2}{m_2 + m_{s_1} + m_{s_2}} + \frac{9m_1 m_2}{2m_1 + m_{s_1} + m_{s_2}} \right\} \quad (81)$$

Table 2. Reliability analysis of the system

q	0	1	2	3	4	5	6	7	8	9	10
Reliability	1.0000	0.9455	0.8907	0.8363	0.7830	0.7311	0.6811	0.6332	0.5875	0.5442	0.5032

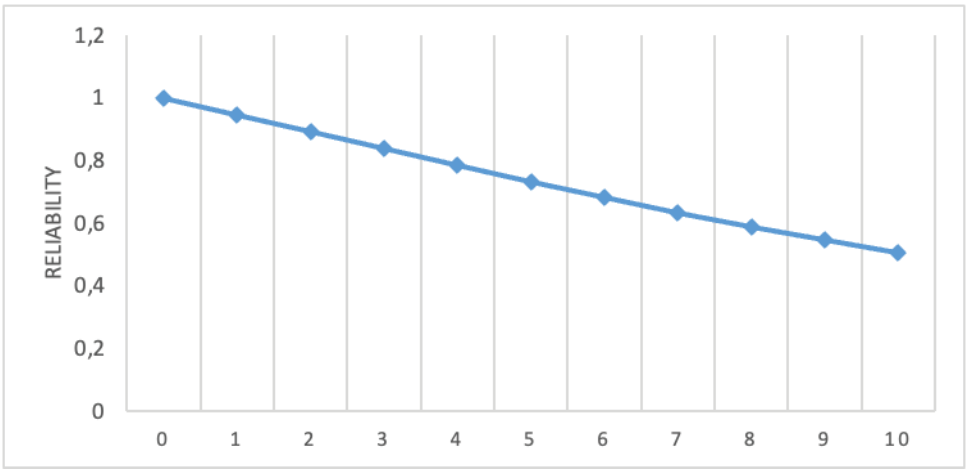


Figure 4. Honeynet Reliability Analysis.

Table 3. MTTF of the system

Failure Rate	$MTTF\ m_1$ (a)	$MTTF\ m_2$ (b)	$MTTF\ m_r$ (c)	$MTTF\ m_h$ (d)	$MTTF\ m_{s_1}$ (e)	$MTTF\ m_{s_2}$ (f)
0.001	4.4962	15.4494	15.2130	15.4035	15.9030	16.1377
0.002	4.5353	15.2001	15.0268	15.2130	15.6754	15.9030
0.003	4.5692	14.9628	14.8449	15.0268	15.4546	15.6754
0.004	4.5987	14.7365	14.6670	14.8449	15.2403	15.4546
0.005	4.6246	14.5202	14.4931	14.6670	15.0321	15.2403
0.006	4.6474	14.3132	14.3230	14.4931	14.8296	15.0321
0.007	4.6676	14.1147	14.1566	14.3230	14.6327	14.8296
0.008	4.6856	13.9241	13.9938	14.1566	14.4411	14.6327
0.009	4.7016	13.7408	13.8345	13.9938	14.2546	14.4411

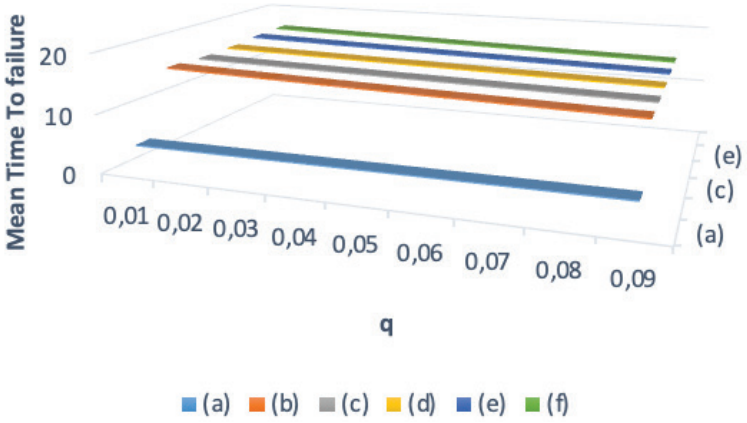


Figure 5. Honeynet MTTF Analysis.

Assuming $m_1 = 0.011$, $m_2 = 0.012$, $m_r = 0.013$, $m_h = 0.014$, $m_{s_1} = 0.015$, $m_{s_1} = 0.016$ and varying the required failure rate as 0.001, 0.002...0.009 in equation (81) while others kept constant Table 3 below is obtained

Sensitivity Analysis

The computation of sensitivity MTTF is studied through the partial differentiation of MTTF with respect to the failure rates $m_1 = 0.011$, $m_2 = 0.012$, $m_r = 0.013$, $m_h = 0.014$, m_{s_1}

Table 4. Sensitivity analysis of the system.

Failure Rate	$\frac{\partial(MTTF)}{m_1}$ (I)	$\frac{\partial(MTTF)}{m_2}$ (II)	$\frac{\partial(MTTF)}{m_r}$ (III)	$\frac{\partial(MTTF)}{m_h}$ (IV)	$\frac{\partial(MTTF)}{m_{s_1}}$ (V)	$\frac{\partial(MTTF)}{m_{s_2}}$ (VI)
0.001	42.0655	-255.6961	-188.3255	-192.7575	-231.0825	-238.4650
0.002	36.3253	-243.1247	-184.0386	-188.3255	-224.1070	-231.0825
0.003	31.5751	-231.6505	-179.8908	-184.0386	-217.5019	-224.1070
0.004	27.6088	-221.1508	-175.8762	-179.8908	-211.2352	-217.5019
0.005	24.2705	-211.5148	-171.9893	-175.8762	-205.2790	-211.2352
0.006	21.4402	-202.6442	-168.2250	-171.9893	-199.6086	-205.2790
0.007	19.0249	-194.4527	-164.5782	-168.2250	-194.2023	-199.6086
0.008	16.9513	-186.8651	-161.0443	-164.5782	-189.0406	-194.2023
0.009	15.1615	-179.8160	-157.6187	-161.0443	-184.1061	-189.0406

Table 5. Profit of the system

q	$E_p(q)$ $D_2 = 0.1$	$E_p(q)$ $D_2 = 0.2$	$E_p(q)$ $D_2 = 0.3$	$E_p(q)$ $D_2 = 0.4$	$E_p(q)$ $D_2 = 0.5$	$E_p(q)$ $D_2 = 0.6$
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.8878	0.7878	0.6878	0.5878	0.4878	0.3878
2	1.7702	1.5702	1.3702	1.1702	0.9702	0.7702
3	2.6509	2.3509	2.0509	1.7509	1.4509	1.1509
4	3.5293	3.1293	2.7293	2.3293	1.9293	1.5293
5	4.4051	3.9051	3.4051	2.9051	2.4051	1.9051
6	5.2782	4.6782	4.0782	3.4782	2.8782	2.2782
7	6.1487	5.4487	4.7487	4.0487	3.3487	2.6487
8	7.0164	6.2164	5.4164	4.6164	3.8164	3.0164
9	7.8815	6.9815	6.0815	5.1815	4.2815	3.3815
10	8.7439	7.7439	6.7439	5.7439	4.7439	3.7439

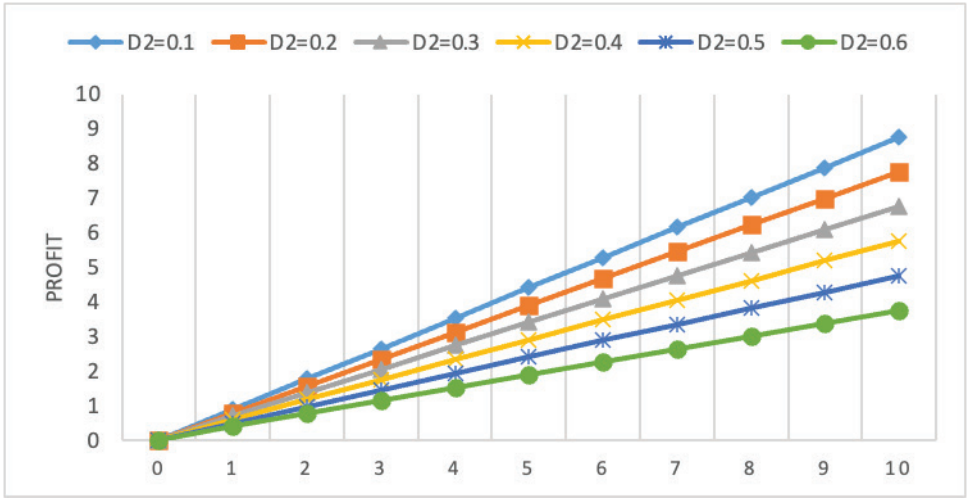


Figure 6. Honeynet Profit Analysis.

Table 6. Profit of the system

q	$E_p(q)$ $D_1 = 2$	$E_p(q)$ $D_1 = 4$	$E_p(q)$ $D_1 = 6$	$E_p(q)$ $D_1 = 8$	$E_p(q)$ $D_1 = 10$	$E_p(q)$ $D_1 = 12$
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.9757	2.9514	4.9272	6.9029	8.8787	10.8544
2	1.9405	5.8811	9.8217	13.7622	17.7028	21.6434
3	2.9019	8.8038	14.7057	20.6076	26.5095	32.4114
4	3.8586	11.7172	19.5759	27.4345	35.2932	43.1518
5	4.8102	14.6204	24.4307	34.2409	44.0512	53.8614
6	5.7565	17.5130	29.2695	41.0260	52.7825	64.5390
7	6.6974	20.3948	34.0922	47.7896	61.4870	75.1844
8	7.6329	23.2658	38.8987	54.5331	70.1646	85.7975
9	8.5630	26.1261	43.6892	61.2522	78.8153	96.3784
10	9.4878	28.9757	48.4636	67.9514	87.4393	106.9272

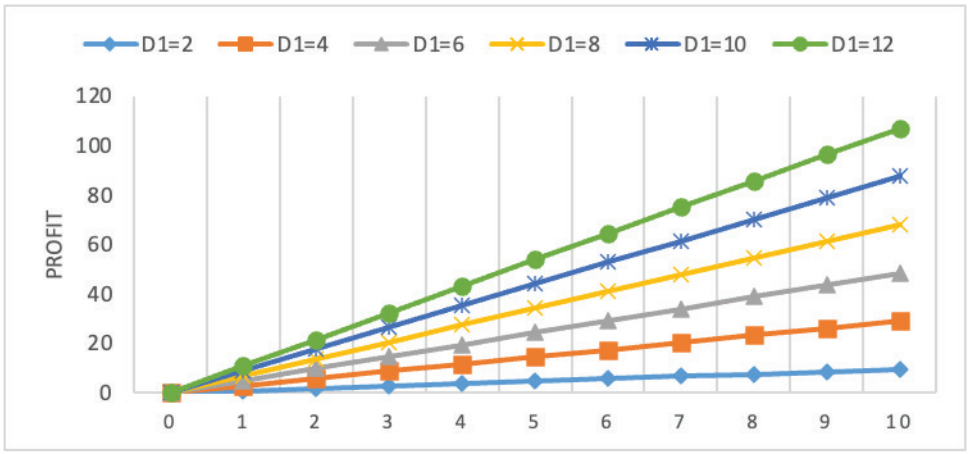


Figure 7. Honeynet Profit Analysis.

= 0.015, m_{s_1} = 0.016 of the system, by introducing the set of parametric variation of the failure rates from the resulting expression, we calculated the sensitivity as shown in Table 5 and the corresponding value in Figure 5

COST (Revenue fixed)

The service cost (D_2) is varied and revenue (D_1) fixed as 0.1, 0.2..., 0.6 for time interval, as q = 0, 1...10 in equation (83).

$$E_p(q) = D_1 \int_0^q H_{up}(q)dq - D_2 q \tag{82}$$

$$E_p(q) = D_1 \begin{Bmatrix} 0.000369e^{-1.04200q} + 0.000434e^{-1.04300q} + 0.000529e^{-1.05300q} + \\ 0.000518e^{-1.05500q} - 0.008282e^{-2.78291q} + 0.007873e^{-1.17509q} + \\ 5.033668e^{-1.11551q} - 355.528670e^{-0.00277q} + 355.5272 \end{Bmatrix} - D_2(q) \tag{83}$$

The revenue (D_1) is fixed and service cost (D_2) is varied as 0.1, 0.2…, 0.6 for time interval, as q = 0, 1…10 in equation (83).

COST (Service cost fixed)

The service cost (D_2) is fixed and revenue (D_1) varied as 0.1, 0.2..., 0.6 for time interval, as q = 0, 1...10 in equation (83).

RESULTS AND DISCUSSION

The objective of this section is to express numerical experiment so as to see effect of the parameters on the performance of each and every honeynet subsystem. The findings in terms of honeynet availability were briefed in the figure (3) through figure (7) and table (1) through table (5) above. The following figures have shown the simulations of availability with respect time (t) of honeynet system and it was observed that system availability decreases with increase in time (t), on the other hand the honeynet system reliability analysis shows from their respective figures above that as the time increases reliability decreases. However, in another dimension the average time to honeynet system failure (MTTF) were analyzed based on different failure rate by varying it from 0.001,.., 0.009 fixing m_1 = 0.011, m_2 = 0.012, m_r = 0.013, m_h = 0.014, m_{s_1} = 0.015, m_{s_1} = 0.016. MTTF honeynet system decreases with increase in failure rate in all the cases, MTTF Sensitivity was checked in this article to determine how the honeynet system was sensitive to the change in parameter and was identified that as the failure rate increases seems to be decreasing. Cost analysis on the other hand have been investigated on the service cost from (0.0) through (10) for the honeynet system throughout the findings it was observed that cost in terms of fixed revenue it happens that cost increases with time, also if the service is fixed the cost increases. To this fact, the honeynet system require optimal maintenance action in order to avoid huge downfall and adequate the life span of the network.

CONCLUSION

In this research, the honeynet sensor supports interception of SSL connections and make decision about the incoming traffic into the system. It determines if the traffic is malicious and thus redirect it to a honeypots or it is valid and thus redirect it to the real production system. Ultimately, the honeynet sensor performs three essential functions, viz: data control, which involves controlling the flow of data so that the attacker does not realize being in the honeynet and ensuring that the honeynet system is not used to attack other systems in the event of system compromise; data capture, which involves capturing all the data regarding movements and actions within the honeynet; and data collection, which involves the ability to securely tranfer all the captured data to a central database/log service, also implemented within the honeynet sensor. Furthermore, the honeypots are computer systems that duplicate and disguise themselves as real production systems in order to lure

an attacker. The honeypots are controlled by the honeywall. They typically implements Sebek/Qebek monitoring tool. When the honeypots receive a malicious request from attacker, the systems invisibly monitor and capture activities of the attacker in the honeypots and send the captured data to the central log in the honeynet sensor for analysis, this implies that the availability of all the honeypot need to be checked and protected at all cost. Despite extensive and mature research on honeynet system, reliability modeling, analysis, and performance prediction and evaluation, copula-based techniques for accurately testing, estimating and optimizing the overall performance of honeynet systems remain lacking.

The research work presented will help plant management to shun away an erroneous performance assessment caused by poor system design. Failure occurrence, monitoring of condition can be extended and incorporated to allow management in approving the optimal replacement/maintenance time.

ACKNOWLEDGEMENT

This research is supported by the National Social Science Foundation of China (Grant No. 20BTJ044) and the Provincial Natural Science Foundation of Hunan Grant (Grant No. 2024JJ5453).

AUTHORSHIP CONTRIBUTIONS

Muhammad Salihu Isa initiate the model and do all the writing and mathematical analysis while Jinbiao Wu and Ibrahim Yusuf helps in editing and supervision.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Paryathia P, Chintab A, Patnala CM. A Honey Pot Implementation for Security Enhancement in IOT System using AES and Key management. Turk J Comput Math Educ 2021;12:5206–5214. [CrossRef]

[2] Naik N, Jenkins P, Savage N. A computational intelligence enabled honeypot for chasing ghosts in the wires. Complex Intell Syst 2021;7:477–494. [CrossRef]

[3] Kondra JR, Bharti SK, Mishra SK, Babu KS. Honeypot-Based Intrusion Detection System: A Performance Analysis. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom); 2016. p. 2347–2351

[4] Agrawal N, Tapaswi S. The performance analysis of honeypot based intrusion detection system for wireless network. Int J Wirel Inf Netw 2017;24:14–21. [CrossRef]

[5] Kasongo SM, Sun Y. Performance analysis of intrusion detection systems using a feature selection method on the UNSW-NB15 dataset. J Big Data 2020;7:1–20. [CrossRef]

[6] Disha RA, Waheed S. Performance analysis of machine learning models for intrusion detection system using gini impurity-based weighted random forest (GIWRF) feature selection technique. Cybersecurity 2022;5:1. [CrossRef]

[7] Alazzam H, Sharieh A, Sabri KE. A feature selection algorithm for intrusion detection system based on pigeon inspired optimizer. Expert Syst Appl 2020;148:113–249. [CrossRef]

[8] Belgrana FZ, Benamrane N, Hamaida MA. Network Intrusion Detection System using Neural Network and Condensed Nearest Neighbors with Selection of NSL-KDD Influencing features. In: 2020 IEEE International Conference on Internet of Things and Intelligence System; 2020. p. 23–29. [CrossRef]

[9] Mauro DM, Galatro G, Liotta A. Experimental review of neural-based approaches for network intrusion management. IEEE Trans Netw Serv Manag 2020;17:2480–2495. [CrossRef]

[10] Kelly C, Pitropakis N, Mylonas A, McKeown S, Buchanan WJ. A comparative analysis of honeypots on different cloud platforms. Sensors. 2021;21:2433. [CrossRef]

[11] Sethia V, Jeyasekar A. Malware Capturing and Analysis using Dionaea Honeypot. In: 2019 International Carnahan Conference on Security Technology; 2019 Oct 1-3; Chennai, India. p. 1–4. [CrossRef]

[12] Lee J, Pak J, Lee M. Network Intrusion Detection System using Feature Extraction Based on Deep Sparse Autoencoder. In: 2020 International Conference on Information and Communication Technology Convergence; 2020. p. 1282–1287. [CrossRef]

[13] Gu J, Lu S. An effective intrusion detection approach using SVM with naive bayes feature embedding. Comput Secur 2021;103:102–158. [CrossRef]

[14] Isa MS, Yusuf I, Ali UA, Suleiman K, Yusuf B, Ismail AL. Reliability analysis of multi-workstation computer network configured as series-parallel system via gumbel - hougaard family copula. Int J Oper Res 2022;19:13–26.

[15] Isa MS, Abubakar MI, Ibrahim KH, Yusuf I, Tukur I. Performance analysis of complex series parallel computer network with transparent bridge using copula distribution. Int J Reliab Risk Saf Theory Appl 2021;4:47–59. [CrossRef]

[16] Xie L, Lundteigen MA, Liu YL. Common Cause Failures and Cascading Failures in Technical Systems, Similarities, Differences and Barriers. In Haugen S, Barros A, Gulijk C, Kongsvik T, Vinnem JE, (editors). Safety and Reliability - Safe Societies in a Changing World. Proceedings of ESREL 2018, June 17-21, 2018, Trondheim, Norway. [CrossRef]

[17] Xie L, Lundteigen MA, Liu YL. Performance analysis of safety instrumented systems against cascading failures during prolonged demands. Reliab Eng Syst Saf 2021;216. [CrossRef]

[18] Yusuf I, Ismail AL, Singh VV, Ali UA, Sufi NA. Performance analysis of multi computer system consisting of three subsystems in series configuration using copula repair policy. SN Comput Sci 2020;1:241. [CrossRef]

[19] Colledani M, Tolio T, Yemane A. Production Quality Improvement During manufacturing systems ramp-up. J Manuf Sci Technol 2019;23. [CrossRef]

[20] Althubiti SA, Jones EM, Roy K. LSTM for Anomaly-Based Network Intrusion Detection. In: 28th International Telecommunication Networks and Applications Conference; 2018. [CrossRef]

[21] AlHamouz S, Abu-Shareha A. Hybrid Classification Approach Using Self-Organizing Map and Back Propagation Artificial Neural Networks for Intrusion Detection. In: 10th International Conference on Developments in eSystems Engineering (DeSE); 2017. [CrossRef]

[22] Albahar M, Alharbi A, Alsuwat M, Aljuaid H. A hybrid model based on radial basis function neural network for intrusion detection. Int J Adv Comput Sci Appl 2020;11:781–791. [CrossRef]

[23] Arqub OA, Singh J, Alhodaly M. Adaptation of kernel functions-based approach with Atangana-Baleanu-Caputo distributed order derivative for solutions of fuzzy fractional Volterra and Fredholm integrodifferential equations. Math Meth Appl Sci 2021;46:7228. [CrossRef]

[24] Hammour ZA, Arqub OA, Momani S, Nabil S. Optimization Solution of Troesch's and Bratu's Problems of Ordinary Type Using Novel Continuous Genetic Algorithm. Discret Dyn Nat Soc 2014;2014:401696. [CrossRef]

[25] Arqub OA, Hammour ZA. Numerical solution of systems of second-order boundary value problems using continuous genetic algorithm. Inf Sci 2014;279:396–415. [CrossRef]

[26] Arqub OA, Singh J, Banan M, Alhodaly M. Reproducing kernel approach for numerical solutions of fuzzy fractional initial value problems under the Mittag-Leffler kernel differential operator. Math Meth Appl Sci 2021;46:7965–7986. [CrossRef]

[27] Kenan E, Mustafa CK, Boru B. Comparison of gesture classification methods with contact and non-contact sensors for human-computer interaction. Sigma J Eng Nat Sci 2021;40:219–226.

[28] Şekerci AZ, Aydın N. A stochastic model for facility locations using the priority of fuzzy AHP. Sigma J Eng Nat Sci 2022;40:649–662. [CrossRef]

[29] Aydın Er B, Şişman A, Ardalı Y. Applicability of radial-based artificial neural networks (RBNN) on coliform calculation: A case of study. Sigma J Eng Nat Sci 2022;40:724–731. [CrossRef]

[30] Tolga B, Ali FG. BLEVE risk effect estimation using the Levenberg-Marquardt algorithm in an artificial neural network model. Sigma J Eng Nat Sci 2022;40:877–893.

[31] Bakar O, Murat B. Applicability of radial-based artificial neural networks (RBNN) on coliform calculation: A case of study. Sigma J Eng Nat Sci 2021;40:235–242.

[32] Adem Y. Intuitionistic fuzzy hypersoft topology and its applications to multi-criteria decision-making. Sigma J Eng Nat Sci 2023;41:106–118.

[33] Maryam B, Rashid R, Karim S. On codes over product of finite chain rings. Sigma J Eng Nat Sci 2023;41:145–155.

[34] Isa MS, Yusuf I, Ali UA, Jinbiao W. Series-parallel computer system performance evaluation with human operator using gumbel hougaard family copula. In: Computational Intelligence in Sustainable Reliability Engineering. 2023. p. 109–127. [CrossRef]

[35] Yusuf I, Ismail AL, Sufi NA, Ambursa FU, Sanusi A, Isa MS. Reliability Analysis of Distributed System for Enhancing Data Replication using Gumbel Hougaard Family Copula Approach Joint Probability Distribution. J Ind Eng Int 2021;17:59–78.

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1183–1193, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2023.00133

Research Article

Modeling asymmetrically dependent automobile bodily injury claim data using Khoudraji Copulas

Emel KIZILOK KARA[1], Sibel AÇIK KEMALOĞLU[2,*]

[1]*Department of Actuarial Science, Kırıkkale University, Kırıkkale, 71451, Türkiye*
[2]*Department of Statistics, Ankara University, Ankara, 06800, Türkiye*

ARTICLE INFO

Article history
Received: 03 February 2023
Revised: 12 March 2023
Accepted: 13 May 2023

Keywords:
Asymmetric Copula; Archimedean Copula; Automobile Bodily Injury Claims Data; Directional Dependence

ABSTRACT

Linear-dependent variables are typically modeled through the Spearman correlation, a classical statistical technique. In reality, the dependence between the data cannot always be linear. The copula approach has often been a popular tool for modeling dependent data in these cases. Archimedean copulas, which can model mostly symmetrical data, are also among the copula families used for this purpose. Recently, asymmetric copula models have been developed to model unsymmetrical-dependent variables. The dependency measure is calculated using directional dependency coefficients instead of the Spearman correlation when the data is asymmetrical. Appropriate asymmetric model selection is made with the help of these measurements.

In the study, first, dependency parameters corresponding to different Spearman coefficients were obtained for Archimedean copula families, and asymmetric copulas were derived from them. Then, simulation data were obtained for these parameter values to determine the effect of asymmetry on data modeling, and directional dependency measures were found. In addition, the study methodology was applied to automobile bodily injury claims data, which is a real dataset with an asymmetric structure. Here, we used two different asymmetric models: the Khoudraji copula KC models, which are created by multiplying independent and Archimedean copulas, and the LCC models, which are linear-convex combinations of Archimedean copulas. Finally, the appropriate model was selected according to the directional dependency coefficients, and the results were interpreted.

Cite this article as: Kızılok Kara E, Açık Kemaloğlu S. Modeling asymmetrically dependent automobile bodily injury claim data using Khoudraji Copulas. Sigma J Eng Nat Sci 2024;42(4):1183–1193.

INTRODUCTION

Determining the dependency structure of variables is important in many areas of research, including statistics, finance, engineering, and actuaries. In deciding this dependence structure, the Spearman correlation, which is one of the standard statistical methods, is used. This method can model variables with a normal distribution, i.e., linear

***Corresponding author.**
*E-mail address: sibelacik@gmail.com

This paper was recommended for publication in revised form by Editor in Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey

dependent variables. However, the dependency coefficient may often differ in the real data's lower and upper tail regions. It may also be affected by whether the data are symmetrical or asymmetrical. For situations where the correlation becomes complex, known statistical models are insufficient. Analytical methods have recently been developed for this purpose. In this sense, copulas, first proposed by Sklar [1], are popular methods used to model dependent variables.

Copulas are preferred because they allow for more realistic data modeling and are capable of generating joint distributions for statistical modeling of dependent variables without restrictions on the marginal distributions of each variable. Nelsen [2], Salvadori and De Michele [3], Genest and Favre [4], Joe [5], Durante and Sempi [6], and Hong et al. [7] showed statistical behavior of copulas as dependent variables.

Moreover, there are studies on the use of copulas in fields such as hydrology [8-10], earthquakes [11], finance [12-15], wind [16], ocean [17, 18], climate science [19], and bioinformatics [20, 21].

On the other hand, most parametric copula models, such as Archimedean copulas, can only be applied to data with symmetric dependence. Some authors have highlighted this shortcoming of existing copula approaches in their studies, such as Genest and Favre [4], Kim et al. [20], and Sungur [22,23]. Indeed, most data have an asymmetric dependence structure. In such data, ignoring asymmetry affects the identification of the dependence structure and subsequent calculations. Some authors such as [24-28] have recently contributed to the development of asymmetric copula construction to eliminate this deficiency. These include various techniques used in multivariate data modeling to capture asymmetric dependence.

Our study focuses on the asymmetric Khoudraji copula (KC) and linear convex combination (LCC) copula families, which can be easily generated using Archimedean copulas. Khoudraji copulas were first developed by Khoudraji [29] and consist of Archimedean copula families and independent (product) copula families. Subsequently, authors such as Nelsen [2], Rodríguez-Lallena and Ubeda-Flores [24], Klement and Mesiar [25], Liebscher [30], Durante [31], Quessy and Kortbi [32], Siburg et al. [33], and Bezak et al. [34] developed asymmetric copulas. In addition, Zhang et al. [35] examined several asymmetric copula functions capable of modeling both linear and nonlinear asymmetric dependence structures using Khoudraji copulas between ocean variables. Moreover, Zhang et al. [36] demonstrated the advantages of asymmetric copulas with Khoudraji copulas and compared them with traditional copula approaches for modeling site soil data. Besides, Lin et al. [37], Bai et al. [38], and Huang and Dong [39] compared the performance of symmetric copula, Khourdaji copula, and traditional conditional modeling methods on bivariate wave data. LCC copulas are also constructed from Archimedean copulas and their linear convex combinations. Authors such as Ma and Zhang [18], Siburg [33], and Wu [40] have investigated the asymmetric properties of these asymmetric families. Recently, some authors ([12], [15], [41]) have also shown that asymmetric copulas provide more realistic and accurate results when modeling asymmetric multivariate data.

While model selection with classical models considers well-known selection criteria such as AIC, KS Cramer-von Mises, and MSE, model selection with asymmetric models uses directional dependence measures calculated based on conditional copula functions. Model selection with directional dependence of copulas is a statistical approach that involves choosing the most appropriate copula function to describe the dependency structure between two or more random variables. The importance of this study is the selection of the most appropriate model with the directional dependency method, which considers the direction of the relationship, among the asymmetric models created with Khoudraji copulas. The advantage of this method is that it allows more accurate and flexible modeling of the dependency structures of asymmetric data.

This study analyzed data on "automobile bodily injury claims" from CASdataset [42]. Frees and Wang [43] modeled these data using the classical copula approach. However, since the data are both dependent and asymmetric, it is necessary to work with asymmetric-dependent models. To this end, unlike the previous study, we apply the asymmetric copula approach to this dataset and determine appropriate models according to directional dependence measures. Thus, this modeling improves the risk assessment and decision-making process by providing valuable information and tools for insurers, researchers, and other stakeholders in the auto insurance industry.

The remainder of this article is organized as follows. The first section introduces copulas and asymmetric copula models. The Archimedean copulas used in this study and the Khoudraji asymmetric copula models derived from them are given. The parameter values corresponding to Spearman correlations were found. The symmetric and asymmetric measurements were analyzed, and the results obtained by simulation for these parameters are presented in tables and graphs. In the next section, one-parameter and two-parameter KCC and LCC models are applied to "Automobile Bodily Injury Claims" data, and model selection is performed according to directional dependence measures. The results are discussed in the last section.

COPULA THEORY AND MEASURES OF ASYMMETRY

Definition of the Copula

A copula is a mathematical function that describes the dependence structure between multiple variables, regardless of their marginal distributions. The idea behind copulas is to separate the modeling of the dependencies between variables from the modeling of the marginal distributions

of individual variables. This allows for more flexibility in modeling complex dependencies, especially when the underlying data do not follow a normal distribution. Copulas have been widely used in various fields, such as finance, economics, and actuarial science, to model multivariate data and estimate the risk of various financial instruments.

The Sklar theorem, first introduced by economist Sklar [1], plays an important role in copula theory, which is a powerful tool for modeling the dependencies between variables in multivariate data. It is defined as

Sklar's Theorem: Let H be an n-dimensional distribution function with marginal distributions $F_1,\ldots,F_n$. An n-dimensional copula C for all $x \in R_n$ is given by

$$H(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)) \quad (1)$$

If $F_1,\ldots,F_n$ are continuous, then C is unique. Conversely, if $F_1,\ldots,F_n$ are marginal distribution functions and C is a copula, a function $H(x_1,\ldots,x_n)$ with marginal distributions $F_1,\ldots,F_n$ is defined by Eq. (1).

The theorem states that any multivariate distribution function can be represented as the copula of its marginal distribution functions and a uniformly distributed random variable on the unit hypercube. This result allows for flexible modeling of the dependence structure between different variables and has important applications in fields such as finance, insurance, and actuarial science.

A unique bivariate copula C: $[0,1]^2 \rightarrow [0,1]$ is defined as $C(u,v) = F_{XY}(F_X^{-1}(u), F_Y^{-1}(v))$, where $F_X^{-1}(u)$ and $F_Y^{-1}(v)$ are the inverse distribution functions of X and Y, respectively.

Asymmetric Copula Models

Different types of copulas have been proposed in the literature, such as Archimedean copulas, elliptical copulas, extreme-value copulas, vine copulas, and empirical copulas. Many commonly used copula families, such as the Gaussian, Clayton, Gumbel, and Frank copulas, have the property of exchangeability, which means that the copula function is symmetric with respect to its arguments. This means that $C(u,v) = C(v,u)$ for all $u,v \in [0,1]$. This property is desirable because it allows a simpler and more intuitive interpretation of the dependence structure. However, this can be a limitation if the data have asymmetric dependence.

To eliminate this deficiency, recently, asymmetric copulas have been constructed in various ways. One simple way to construct an asymmetric copula is by using the rotation method. In this method, using a rotation matrix R, the standard copula is transformed into a new copula with a different dependency structure. The rotation is then defined as $C_R(u,v) = C(R^*(u,v))$ where C_R is the rotated copula and $R^*(u,v)$ is the rotation of the standard coordinates.

Ma and Zhang [18], and Zhang et al. [35] have described and implemented other methods of creating an asymmetric copula. This study focuses on Khoudraji copulas (KC), developed by multiplying copulas, and LCC copulas constructed using their linear convex combinations. KC and LCC copulas have been widely used in different fields and applications, particularly in finance, in modeling dependency structures in various types of data. They have also been used in bioinformatics, environmental sciences, and engineering to model the dependence between variables. These studies fit the empirical data better than some traditional copulas when the data exhibit asymmetric dependence.

Our study uses these models to analyze automobile bodily injury claim data and determine the best-fit copula models. We choose appropriate models to analyze automobile bodily injury claim data and make inferences.

In the following section, we introduce the mathematical form of the KC and LCC models.

Khoudraji copula (KC) model:

Khoudraji copulas are a class of asymmetrical copulas first introduced by Khoudraji [29]. Later, Liebscher [30] defined its general form. They may model both positive and negative dependency, and a wide variety of dependency constructs can be captured by varying their parameters.

The mathematical model of a Khoudraji copula function $C_{\alpha,\beta}: I^2 \rightarrow I$ is defined as

$$C_{\alpha,\beta}(u,v) = C_1\left(u^{\bar{\alpha}}, v^{\bar{\beta}}\right) C_2\left(u^{\alpha}, v^{\beta}\right) \quad (2)$$

where $\phi = (\alpha, \beta, \theta)$, $\alpha, \beta \in (0,1)$, $\alpha \neq 1/2$, $\beta \neq 1/2$, $\alpha + \bar{\alpha} = 1$, $\beta + \bar{\beta} = 1$.

In Eq. (1), if C_1 is independent copula $C(u,v) = uv$, and C_2 is a symmetrical Archimedean copula family with a dependency parameter θ, $C_{\alpha,\beta}$ is called a Khoudraji copula. For $\beta = 1 - \alpha$ one parameter Khoudraji copula families (KC1-model),

$$C_{KC1}(u,v) = u^{1-\alpha} v^{\alpha} C(u^{\alpha}, v^{1-\alpha}) \quad (3)$$

and for $\alpha \neq \beta$, two-parameter Khoudraji copula families (KC2-model)

$$C_{KC2}(u,v) = u^{1-\alpha} v^{1-\beta} C\left(u^{\alpha}, v^{\beta}\right) \quad (4)$$

are expressed as in Eqs. (3) and (4), respectively.

Linear convex combination (LCC) model:

It is possible to create asymmetrical copulas using linear and convex combinations of the copulas. However, the resulting pattern remains a symmetrical copula when direct linear-convex combinations are produced with symmetric copula functions. Using the method in Wu [40], basic copulas can be modified to include asymmetrical features. This method involves deriving a new asymmetric copula by modifying the basic copulas using a weighting function

that allows for asymmetry. A new asymmetric copula can be derived using this methodology as follows:

$$\breve{C}_k(u_1,\dots,u_n) = C(u_1,\dots,u_{k-1},1,u_{k+1},\dots,u_n) - C(u_1,\dots,u_{k-1},1-u_k,u_{k+1},\dots,u_n) \quad (5)$$

where $C(\cdot)$ is the n-dimensional base copula.

An LCC copula can be constructed to capture the asymmetric characteristics of a multivariate variable based on Eq. (6) as follows:

$$C_{LCC}(u_1,\dots,u_n) = \sum_{k=0}^{n} p_k \breve{C}_k(u_1,\dots,u_n) \quad (6)$$

where $0 \le p_k \le 1$ and $\sum_{k=0}^{n} p_k = 1$. Thus, an asymmetric copula can be constructed by linear convex combinations of $\breve{C}_k(\cdot)$. There are many copula families that can be specified for base copula C. For instance, a bivariate copula $C(u,v)$ according to Eq. (5) can be written as follows:

$$C(u,v) = u + v - 1 + C_0(1-u,1-v)$$
$$\breve{C}_1(u,v) = v - C(1-u,v) = u - C_0(u,1-v)$$
$$\breve{C}_2(u,v) = u - C(u,1-v) = v - C_0(1-u,v)$$

Thus, the constructed one-parameter (LCC1-model) and two-parameter (LCC2-model) asymmetric copulas by linear combination can be given in Eqs. (7) and (8), respectively.

$$C_{LCC1}(u,v) = p_0 C(u,v;\theta_1) + p_1 \breve{C}_1(u,v;\theta_2) \quad (7)$$

$$C_{LCC2}(u,v) = p_0 C(u,v) + p_1 \breve{C}_1(u,v) + p_2 \breve{C}_2(u,v) \quad (8)$$

Directional Dependence Measures

Directional dependence refers to the ability to measure the degree and direction of dependence between two or more variables. When the coefficient of association between variables is linear, it can be measured by the Spearman correlation, whereas when it is nonlinear, it can be measured by the Spearman coefficient based on the copula, expressed as follows:

$$\rho_C = 12 \iint_0^1 C(u,v)dudv - 3 \quad (9)$$

Therefore, ρ_C^2 can be used to calculate the ratio of variables explaining each other. However, when there is asymmetricity in the data structure, the directional dependence coefficients determined according to regression-based copula functions will not be the same relative to each other. Therefore, this situation will cause the ratio of the explained variance to be different.

Accordingly, when dependence is symmetric, the regression functions for U and V have the same linear form, and the same model can be used to predict both U and V. However, when the dependence is asymmetric, the regression functions for U and V will not be the same, and different models will be needed to estimate the regression functions for U and V separately. Detailed information on directional dependence can be found in Sungur [22, 23], Jung et al. [44], and Kim and Kim [45].

The directional dependence coefficients using the copula regression functions (in the directions of U to V ($U \to V$) and V to U ($V \to U$) can be obtained by an approximate calculation method as follows:

$$\tilde{\rho}^{(2)}_{U\to V} = \frac{12}{S}\sum_{s=1}^{S}\left(\tilde{r}_{V|U}(u_s)\right)^2 - 3 \quad (10)$$

and

$$\tilde{\rho}^{(2)}_{V\to U} = \frac{12}{S}\sum_{s=1}^{S}\left(\tilde{r}_{U|V}(v_s)\right)^2 - 3 \quad (11)$$

where,

$$\tilde{r}_{V|U}(u) = 1 - \frac{1}{S}\sum_{s=1}^{S} C_u(v_s) \text{ and}$$
$$C_u(v) \equiv P(V \le v | U = u) = \frac{\partial C(u,v,\Phi)}{\partial u},$$

$$\tilde{r}_{U|V}(v) = 1 - \frac{1}{S}\sum_{s=1}^{S} C_v(u_s) \text{ and}$$
$$C_v(u) \equiv P(U \le u | V = v) = \frac{\partial C(u,v,\Phi)}{\partial v},$$

Φ is the parameter set. $\tilde{r}_{V|U}(u)$ and $\tilde{r}_{U|V}(v)$ are approximately calculated copula regression functions over the pseudo- observations, $(u_s, v_s) \in (0,1)^2$ for the size of the pseudo-observation, S.

Simulation Study of Archimedean and Khoudraji Copula Models

A simulation study was conducted to examine the symmetrical and asymmetrical dependency structures. Archimedean copulas for symmetrical models and Khoudraji copulas for asymmetrical models are considered. First, dependency parameter values corresponding to various correlation values were obtained for the mentioned models (Table 1), and parameter estimates of 1000 data pairs produced with these parameters were made (Table 2). Additionally, the asymmetry test was applied to these simulated data using Cramer-von Misses statistics. As shown in Table 3, simulated data from symmetric models are symmetric (p-values > 0.05), while simulated data from asymmetric models are asymmetric (p-values < 0.05).

Contour plots showing symmetrical and asymmetrical dependency structures are visually presented in Figure 1 for $\rho = 0.4$. These graphs show that the ones drawn for the

Table 1. Parameter values of symmetric and asymmetric copulas for various correlations

ρ	Clayton	KC1-Clayton		Frank	KC1-Frank		Gumbel	KC1-Gumbel	
	θ	α	θ	θ	α	θ	θ	α	θ
0.1	0.143	0.122	2.8	0.603	0.333	1.9	1.072	0.274	1.2
0.2	0.311	0.287	2.7	1.224	0.459	3.7	1.156	0.160	3.0
0.3	0.511	0.352	5.6	1.883	0.511	6.9	1.257	0.246	6.099
0.4	0.759	0.492	14.5	2.610	0.507	19.7	1.382	0.611	9.599

ρ	Clayton	KC2-Clayton			Frank	KC2-Frank			Gumbel	KC2-Gumbel		
	θ	α	β	θ	θ	α	β	θ	θ	α	β	θ
0.2	0.311	0.4	0.6	1.9	1.224	0.4	0.7	3.4	1.156	0.2	0.4	3.8
0.4	0.759	0.6	0.9	2	2.610	0.9	0.9	3.1	1.382	0.6	0.9	1.7
0.6	1.505	0.8	0.9	2.9	4.466	0.7	0.9	8.5	1.755	0.7	0.8	3.1
0.8	3.185	0.9	0.9	7.1	7.902	0.9	0.9	13.8	2.581	0.9	0.9	4.1

Table 2. Parameter estimation results for data pairs generated by simulation (N =1000)

ρ	Clayton	KC1-Clayton			Frank	KC1-Frank			Gumbel	KC1-Gumbel		
	θ	α	θ	$\hat{\rho}$	θ	α	θ	$\hat{\rho}$	θ	α	θ	$\hat{\rho}$
0.1	0.1397	0.1053	4.7777	0.0907	0.5827	0.8922	6.9387	0.0913	0.7191	0.1407	1.2546	0.0935
0.2	0.3356	0.2147	4.3232	0.2007	1.2279	0.3412	4.5981	0.2001	1.1743	0.1728	3.0878	0.2033
0.3	0.4927	0.3915	6.9374	0.2993	1.8870	0.5441	6.9271	0.2970	1.2605	0.2368	5.7554	0.3024
0.4	0.7326	0.4964	13.0285	0.4052	2.7652	0.5165	20.6832	0.3935	1.4015	0.6148	9.5974	0.4033

ρ	Clayton	KC2-Clayton				Frank	KC2-Frank				Gumbel	KC2-Gumbel			
	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\rho}$	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\rho}$	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\rho}$
0.2	0.3387	0.3354	0.9999	1.6262	0.2044	1.2498	0.4036	0.9965	2.7880	0.2050	1.1670	0.2385	0.4178	3.3837	0.2074
0.4	0.7554	0.5809	0.9924	1.8154	0.3957	2.4613	0.7612	0.8330	3.9251	0.4183	1.3966	0.5867	0.8345	1.8024	0.3970
0.6	1.5020	0.7857	0.8652	2.9955	0.5936	4.5125	0.7183	0.9114	8.2398	0.6039	1.7922	0.7433	0.8660	2.6038	0.6050
0.8	3.1726	0.8895	0.8963	7.0759	0.7897	7.9359	0.9098	0.8739	14.5948	0.7998	2.5988	0.9206	0.8845	4.0058	0.7934

Archimedean copulas are symmetrical, and those drawn for the KC1 and KC2 models are irregular and asymmetrical.

The ρ_C^2, $\rho_{u\to v}^{(2)}$ and $\rho_{v\to u}^{(2)}$ values for the asymmetric KC1 and KC2 models of the data pairs produced in different correlations were calculated using equations (9), (10), and (11) and are presented in Table 4. In this table, ρ_S and ρ_S^2 show the Spearman correlation values and coefficients of determination of the produced data, respectively. Here, it can be seen that the directional dependence values from V to U and from U to V are different from the values of ρ_C^2. This indicates that $\rho_{u\to v}^{(2)}$ and $\rho_{v\to u}^{(2)}$values should be used instead of ρ_C^2 because the dependency structures of the KC models are asymmetrical. Moreover, Jung et al. [44] showed that the directional dependence coefficients are theoretically different from ρ_C^2 asymmetric models for some parameters of the generalized FGM copula family.

DATA ANALYSIS

The data contained 174 automobile bodily injury claims collected between 1993 and 1998 in Massachusetts and studied by Frees and Wang [43]. The data can be found in the CASdatasets package [42] of the R program under the name 'usmassBI2'. Descriptive statistics of the data are given in Table 5, and a scatterplot of the data is presented in Figure 2. Here, AC represents the average claims per unit of exposure ($), and PPSM represents the population per square mile of the town.

To choose the most appropriate candidate models that can fit the data, an exchangeability test was performed on the data. The results given in Table 6 show that the data did not fit the copula families. The dependency structure of the data is asymmetric because the p-value (=0.01499) is

Table 3. Asymmetry test results of simulated data from symmetric and asymmetric models: Cramer-von Misses statistics and p-values

Symmetric Archimedean models				**Asymmetric KC1 models**			
ρ	**Clayton**	**Frank**	**Gumbel**	ρ	**KC1-Clayton**	**KC1-Frank**	**KC1-Gumbel**
0.1	0.024344	0.02432	0.048299	0.1	0.077899	0.092821	0.089669
	(0.6888)	(0.7218)	(0.1154)		(0.02647)	(0.008492)	(0.004496)
0.2	0.020519	0.013828	0.018285	0.2	0.07341	0.080107	0.42153
	(0.7947)	(0.9805)	(0.8866)		(0.02348)	(0.01449)	(0.0004995)
0.3	0.030133	0.018385	0.016769	0.3	0.092927	0.05489	0.54746
	(0.3921)	(0.8536)	(0.9066)		(0.005495)	(0.04745)	(0.0004995)
0.4	0.023138	0.031925	0.035001	0.4	0.063067	0.076384	0.31175
	(0.6119)	(0.3322)	(0.2143)		(0.02048)	(0.00649)	(0.0004995)
Symmetric Archimedean models				**Asymmetric KC2 models**			
ρ	**Clayton**	**Frank**	**Gumbel**	ρ	**KC2-Clayton**	**KC2-Frank**	**KC2-Gumbel**
0.2	0.0237	0.0230	0.0462	0.2	0.0866	0.0739	0.0787
	(0.7008)	(0.7068)	(0.1234)		(0.0085)	(0.0245)	(0.0145)
0.4	0.0170	0.0255	0.0305	0.4	0.1329	0.0816	0.0815
	(0.8417)	(0.504)	(0.3072)		(0.0015)	(0.0025)	(0.0035)
0.6	0.0174	0.0234	0.0101	0.6	0.0648	0.0805	0.0663
	(0.6838)	(0.4241)	(0.9945)		(0.0025)	(0.0015)	(0.0015)
0.8	0.0097	0.0214	0.0153	0.8	0.0372	0.0339	0.0338
	(0.9276)	(0.1993)	(0.515)		(0.0195)	(0.0245)	(0.0335)

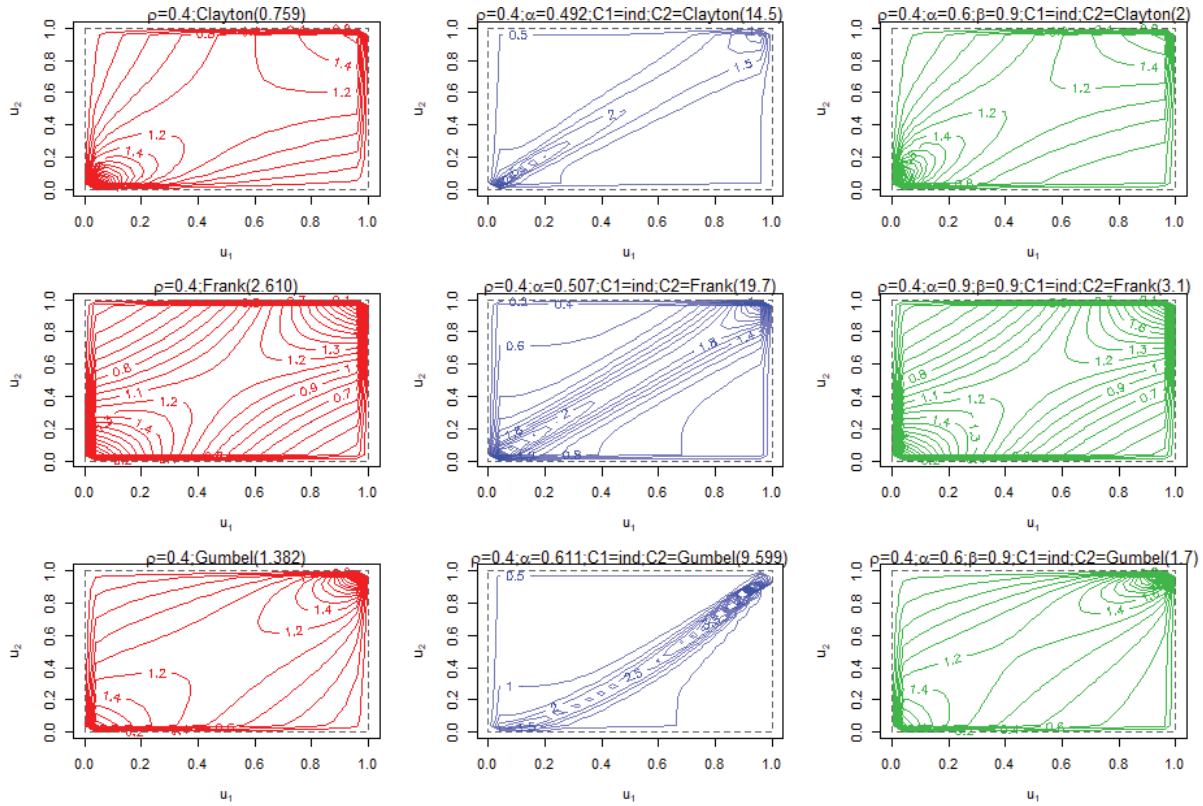


Figure 1. Contour plots for the Archimedean copula (red), KC1 (blue), and KC2 (green) models (p=0.4 and N=1000).

Table 4. Directional dependence coefficients of the asymmetric KC1 and KC2 models

KC1-Clayton						
ρ	ρ_S	ρ_S^2	ρ_C	ρ_C^2	$\rho_{u\to v}^{(2)}$	$\rho_{v\to u}^{(2)}$
0.1	0.0907	0.0082	0.1066	0.0114	0.0357	0.0161
0.2	0.2007	0.0400	0.1951	0.0380	0.0584	0.0241
0.3	0.2993	0.0896	0.3343	0.1117	0.1742	0.1449
0.4	0.4052	0.1642	0.3954	0.1563	0.1430	0.1244
KC1-Frank						
ρ	ρ_S	ρ_S^2	ρ_C	ρ_C^2	$\rho_{u\to v}^{(2)}$	$\rho_{v\to u}^{(2)}$
0.1	0.0913	0.0083	0.1082	0.1170	0.0225	0.0376
0.2	0.2001	0.0400	0.2111	0.0445	0.0634	0.0100
0.3	0.2970	0.0882	0.2981	0.0889	0.0775	0.0867
0.4	0.3935	0.1548	0.4018	0.1614	0.0614	0.1058
KC1-Gumbel						
ρ	ρ_S	ρ_S^2	ρ_C	ρ_C^2	$\rho_{u\to v}^{(2)}$	$\rho_{v\to u}^{(2)}$
0.1	0.0935	0.0087	0.0813	0.0066	0.0240	0.0259
0.2	0.2033	0.0413	0.2135	0.0456	0.0211	0.1393
0.3	0.3024	0.0914	0.2909	0.0846	0.0614	0.0895
0.4	0.4033	0.1626	0.3984	0.1587	0.1845	0.1938
KC2-Clayton						
ρ	ρ_S	ρ_S^2	ρ_C	ρ_C^2	$\rho_{u\to v}^{(2)}$	$\rho_{v\to u}^{(2)}$
0.2	0.2044	0.0418	0.2232	0.0498	0.0335	0.0155
0.4	0.3957	0.1566	0.3987	0.1589	0.1964	0.0837
0.6	0.5936	0.3524	0.5783	0.3344	0.4404	0.5771
0.8	0.7897	0.6236	0.7905	0.6249	0.7191	0.8515
KC2-Frank						
ρ	ρ_S	ρ_S^2	ρ_C	ρ_C^2	$\rho_{u\to v}^{(2)}$	$\rho_{v\to u}^{(2)}$
0.2	0.2050	0.0420	0.2120	0.0450	0.0357	0.0708
0.4	0.4183	0.1750	0.4044	0.1636	0.1365	0.0957
0.6	0.6039	0.3647	0.6097	0.3718	0.3584	0.4006
0.8	0.7998	0.6397	0.7965	0.6344	0.5570	0.6732
KC2-Gumbel						
ρ	ρ_S	ρ_S^2	ρ_C	ρ_C^2	$\rho_{u\to v}^{(2)}$	$\rho_{v\to u}^{(2)}$
0.2	0.2074	0.0430	0.2250	0.0506	0.0538	0.0874
0.4	0.3970	0.1576	0.4053	0.1643	0.1897	0.0816
0.6	0.6050	0.3660	0.6126	0.3753	0.5773	0.4822
0.8	0.7934	0.6295	0.8007	0.6411	0.5764	0.6408

Table 5. Descriptive statistics of automobile bodily injury claim data

	Mean	**Median**	**Min**	**Max**	**St. D**	**Skewness**	**Kurtosis**
AC	137.32	136.49	42.74	248.75	35.18	0.1708	0.2445
PPSM	801.74	593.67	119.56	4636.74	815.41	3.3309	12.4009

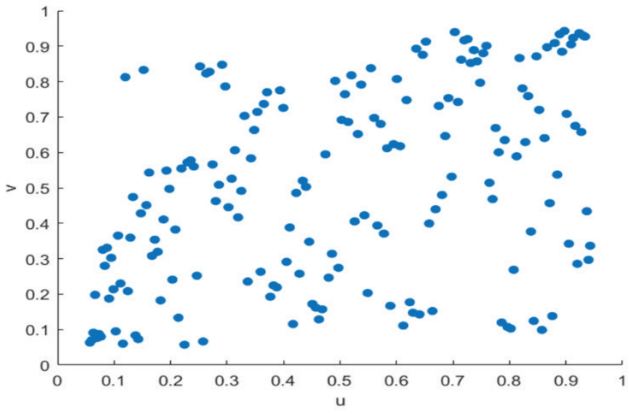


Figure 2. Scatterplot of automobile bodily injury claim data.

smaller than 0.05. The test of goodness of fit for symmetrical Archimedean copula families also supported this result.

Thus, we can choose candidate copula models from among asymmetric copulas. For this purpose, we use the asymmetric KC and LCC models defined by equations (3-4) and (7-8), respectively. We determine the best-fitting model with the AIC value given by

$$AIC = -2LL + 2p$$

where p is the number of parameters and LL is the maximized log-likelihood. The model selection process according to the directional dependency is as follows: First, the model with the smallest AIC value is considered. The model selection is then continued with the directional dependency measures with the smallest and largest values of $\rho^{(2)}_{u\to v}$ and $\rho^{(2)}_{v\to u}$. Finally, the two selected asymmetric models were tested with GOF (Sn-Cramer-von Mises) and the one with the lowest AIC value was selected as the best-fit model.

The parameter estimations, log-likelihood, and the values of Sn, p, and AIC for the asymmetric KC and LCC models were calculated and are shown in Tables 7 and 8. Here, the maximum pseudo-likelihood (MPL) method is used to obtain parameter estimates.

KC models are considered first because the AIC values of the KC models are approximately the same among themselves, and they are insignificant compared to the LCC models. The GOF test was performed on selected KC models, and it is seen in Table 7 that all models fit the data (p >0.05).

The model selection process is then continued according to the directional dependence coefficients. Models with the highest and lowest $\rho^{(2)}_{u\to v}$ and $\rho^{(2)}_{v\to u}$ values are determined, and it is seen in Table 9 that they correspond to the $\mathrm{KC}_1^{\mathrm{G}}$ and $\mathrm{KC}_2^{\mathrm{G}}$ models, respectively. Since more than one copula fits the data, $\mathrm{KC}_2^{\mathrm{G}}$ the model with the smallest AIC value (= -52.7114 from Table 7) is selected as the best-fit model.

A comparison scatterplot between the original data and the simulated data from $\mathrm{KC}_2^{\mathrm{G}}$ the model is made to further check the suitability. For this purpose, their scatterplots are shown in Figure 3. It can be seen from the scatterplots that the simulated data and the original data fit each other very well.

Table 6. Parameter estimation summary of symmetric copula families for automobile bodily injury claim data (S_n: Cramer-von Mises test statistics)

Copula Family	Parameter	*LL*	S_n	*p*-value
Clayton	0.74011	20.53453	0.080170	0.00150
Frank	2.98366	18.34867	0.051319	0.00649
Gumbel	1.42846	20.28379	0.045947	0.02747

Table 7. Estimated parameters, LL, Sn, p, and AIC values of KC models for automobile bodily injury claim data.

Model	$\widehat{\theta}$	$\widehat{\alpha}$	$\widehat{\beta}$	*LL*	S_n	p value	AIC
$\mathrm{KC}_1^{\mathrm{C}}$	25.9738	0.2853	0.7147	32.5540	0.02829	0.2486	-61.1081
$\mathrm{KC}_1^{\mathrm{F}}$	33.9303	0.2828	0.7173	32.3130	0.03527	0.1286	-60.6260
$\mathrm{KC}_1^{\mathrm{G}}$	2.98295	0.3188	0.6812	26.1288	0.02479	0.3114	-48.2575
$\mathrm{KC}_2^{\mathrm{C}}$	19.1733	0.3341	0.8289	33.5462	0.02829	0.2543	-61.0924
$\mathrm{KC}_2^{\mathrm{F}}$	30.4350	0.3164	0.8013	32.7218	0.03190	0.1400	-59.4436
$\mathrm{KC}_2^{\mathrm{G}}$	2.2574	0.4511	0.9987	29.3557	0.02479	0.3171	-52.7114

Table 8. Estimated parameters, LL, and AIC values of LCC models for automobile bodily injury claim data

Model	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_3$	$\widehat{p}_1$	$\widehat{p}_2$	$\widehat{p}_3$	*LL*	AIC
LCC_1^C	2.9103	3.2858	-	0.2800	0.7200	-	-165.8824	377.6481
LCC_1^F	4.1575	5.1835	-	0.1476	0.8524	-	-178.8822	363.7643
LCC_1^G	2.0266	2.9032	-	0.3216	0.6784	-	-160.7671	327.5343
LCC_2^C	2.3716	2.9671	3.2964	0.2889	0.2628	0.4483	-167.8495	345.6991
LCC_2^F	4.6796	5.3969	5.3969	0.1571	0.0291	0.8138	-170.3059	350.6118
LCC_2^G	1.7356	2.6574	3.0947	0.2919	0.4277	0.2804	-171.6882	353.3764

Additionally, the values of ρ_S^2, ρ_C^2, and $\rho_{u\to v}^{(2)}$ the automobile bodily injury claim data in Tables 7 and 9 can be interpreted as follows for the best model chosen (KC_2^G). AC was considered the dependent variable in the data, and PPSM was the independent variable. When the relationship between them was accepted as linear, the coefficient of determination was found to be 0.1843 (ρ_S^2) based on the known Spearman correlation. In other words, the explanation rate of AC with PPSM was 18.43%. When the dependency structure of these variables is modeled with copulas, this explanation rate is 18.975% (ρ_C^2). When modeled with asymmetric copulas, the disclosure rate according to the directional dependency measures was found to be 21.725% ($\rho_{v\to u}^{(2)}$). These results show that the explanation ratio found according to the standard Spearman coefficient without considering dependence and asymmetricity is lower. Thus, the decision-maker will calculate a lower premium than it should be in the case of premium pricing, ignoring the asymmetric model for automobile claim data with an asymmetric dependency.

As a result, the advantage of the model selection method based on directional dependence is that it gives a decision-maker who wants to make actuarial calculations, such as premium pricing in the insurance field, the opportunity to make a more accurate calculation by considering the directional dependence coefficient.

Table 9. Directional dependence coefficients of asymmetric copula models for automobile bodily injury claim data ($\boldsymbol{\rho_s = 0.4293, \rho_s^2 = 0.1843}$)

Models	ρ_C	ρ_C^2	$\rho_{u\to v}^{(2)}$	$\rho_{v\to u}^{(2)}$
KC_1^C	0.33027	0.10908	0.11718	0.13188
KC_1^F	0.32924	0.10840	0.11550	0.13131
KC_1^G	0.32404	0.10500	**0.10619**	**0.12085**
KC_2^C	0.38979	0.15194	0.16472	0.18070
KC_2^F	0.37266	0.13888	0.14915	0.16970
KC_2^G	0.43560	0.18975	**0.19208**	**0.21725**

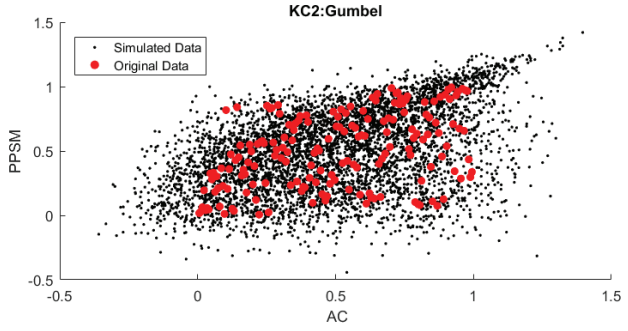


Figure 3. Comparison of the scatterplot between the original data and simulated data.

CONCLUSION

Real data are not always symmetrical. In such cases, it is necessary to use asymmetric models for modeling. Using these asymmetric models, obtaining the desired probabilities and statistical inferences will provide more accurate results. For our data, the first asymmetric tests were performed, and it was found that the dataset was asymmetric. Then, using Clayton Frank and Gumbel copulas, asymmetric models KC1, KC2, LCC1, and LCC2 were created. We used directional dependency coefficients to determine the model that best fits the dataset from among asymmetric models. According to the directional dependency coefficients and AIC, we concluded that the best model that fits our dataset is KC2-Gumbel. This modeling can have important practical implications for the insurance industry, because it can provide more accurate estimates of risk factors and inform risk management and policy decisions. This is particularly important given the increasing importance of data-driven decision-making in the insurance industry and the need to accurately model risk factors to effectively manage risk. In addition, the use of Khoudraji copulas can be applied to other areas of insurance and risk management, where asymmetrically dependent data are

common, such as modeling the joint distribution of insurance claims across different lines of business.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Sklar A. Functions de repartition an dimensions at leurs marges. Publ Inst Statist Univ Paris 1959;8:229-231.
[2] Nelsen RB. An Introduction to Copulas. 2nd ed. New York: Springer; 2006.
[3] Salvadori G, De Michele C. On the use of copulas in hydrology: theory and practice. J Hydrol Eng 2007;12:369-380. [CrossRef]
[4] Genest C, Favre A-C. Everything you always wanted to know about copula modeling but were afraid to ask. J Hydrol Eng 2007;12:347-368. [CrossRef]
[5] Joe H. Dependence Modeling with Copulas. Boca Rotan, Florida: CRC Press; 2014. [CrossRef]
[6] Durante F, Sempi C. Principles of copula theory. Boca Rotan, Florida: CRC Press; 2015. [CrossRef]
[7] Hong Y, Wang JP, Li DQ, Cao ZJ, Ng CWW, Cui P. Statistical and probabilistic analyses of impact pressure and discharge of debris flow from 139 events during 1961 and 2000 at Jiangjia Ravine, China. Eng Geol 2015;187:122-134. [CrossRef]
[8] Salvadori G, De Michele C, Kottegoda NT, Rosso R. Extremes in nature: an approach using copulas. Water Science and Technology Library. Berlin/Heidelberg: Springer Science &Business Media; 2017.
[9] Kara EK, Yıldız O. Bivariate analysis of precipitation and runoff in the Hirfanli Dam Basin, Turkey, using copulas. Istatistik J Turkish Stat Assoc 2014;7:63-70.
[10] Callau Poduje AC, Belli A, Haberlandt UA. Dam risk assessment based on univariate versus bivariate statistical approaches: A case study for Argentina. Hydrol Sci J 2014;59:2216-2232. [CrossRef]
[11] Kara EK. The earthquake risk analysis based on copula models for Turkey. Sigma J Eng Nat Sci 2017;35:187-200.
[12] Kim JM, Jung YS, Sungur EA. Truncation invariant copulas for modeling directional dependence: Application to foreign currency exchange data. Model Assist Stat Appl 2014;9:309-324. [CrossRef]
[13] Kemaloglu SA, Kara EK. Modeling dependent financial assets by dynamic copula and portfolio optimization based on CVaR. Commun Fac Sci Univ Ankara Ser A1 Math Stat 2015;64:1-13. [CrossRef]
[14] Kara EK, Kemaloglu SA. Portfolio optimization of dynamic copula models for dependent financial data using change point approach. Commun Fac Sci Univ Ankara Ser A1 Math Stat 2016;65:175-188. [CrossRef]
[15] Kara EK, Kemaloglu SA, Evkaya ÖO. Modeling Currency Exchange Data with Asymmetric Copula Functions. In: Terzioğlu, M.K. (eds) Advances in Econometrics, Operational Research, Data Science and Actuarial Studies. Contributions to Economics. Springer, Cham., 2022, p. 49-62. [CrossRef]
[16] Moradian S, Olbert AI, Gharbia S, Iglesias G. Copula-based projections of wind power: Ireland as a case study. Renew Sustain Energy Rev 2023;175:113147. [CrossRef]
[17] Corbella S, Stretch DD. Simulating a multivariate sea storm using Archimedean copulas. Coastal Eng 2013;76:68-78. [CrossRef]
[18] Ma P, Zhang Y. Modeling asymmetrically dependent multivariate ocean data using truncated copulas. Ocean Eng 2022;244:110226. [CrossRef]
[19] Li X, Babovic V. Multi-site multivariate downscaling of global climate model outputs: an integrated framework combining quantile mapping, stochastic weather generator and Empirical Copula approaches. Clim Dyn 2019;52:5775-5799. [CrossRef]
[20] Kim JM, Jung YS, Sungur EA, Han KH, Park C, Sohn IA. Copula method for modeling directional dependence of genes. BMC Bioinform 2008;9:225. [CrossRef]
[21] Kim JM, Jung YS, Soderberg T. Directional dependence of genes using survival truncated FGM type modification copulas. Commun Stat Simul Comput 2009;38:1470-1484. [CrossRef]
[22] Sungur EA. A note on directional dependence in regression setting. Commun Stat Theor Methods 2005a;34:1957-1965. [CrossRef]
[23] Sungur EA. Some observations on copula regression functions. Commun Stat Theor Methods 2005b;34:1967-1978. [CrossRef]
[24] Rodríguez-Lallena JA, Úbeda-Flores MA. A new class of bivariate copulas. Stat Probab Lett 2004;66:315-325. [CrossRef]
[25] Klement EP, Mesiar R. How non-symmetric can a copula be? Commen Math Univ Carolinae 2006;47:141-148.

[26] Grimaldi S, Serinaldi F. Asymmetric copula in multivariate flood frequency analysis. Adv. Water Resour 2006;29:1155-1167. [CrossRef]
[27] Mesiar R, Najjari V. New families of symmetric/asymmetric copulas. Fuzzy Sets Syst 2014;252:99-110. [CrossRef]
[28] Mazo G, Girard S, Forbes F. A class of multivariate copulas based on products of bivariate copulas. J Multivariate Anal 2015;140:363-376. [CrossRef]
[29] Khoudraji A. Contributions'al' 'etude des copules et'al-amod 'elisation des valeurs extrˆemes bivari 'ees. PhD thesis, Canada: Universit 'ede Laval, Qu 'ebec; 1995.
[30] Liebscher E. Construction of asymmetric multivariate copulas. J Multivariate Anal 2008;99:2234-2250. [CrossRef]
[31] Durante F. Construction of non-exchangeable bivariate distribution functions. Stat Papers 2009;50:383-391. [CrossRef]
[32] Quessy JF, Kortbi O. Minimum-distance statistics for the selection of an asymmetric copula in Khoudraji's class of models. Stat Sinica Preprint. doi: 10.5705/ss.202014.0082. [CrossRef]
[33] Siburg KF, Stehling K, Stoimenov PA, Weiß GN. An order of asymmetry in copulas, and implications for risk management. Insur Math Econ 2016;68:241-247. [CrossRef]
[34] Bezak N, Rusjan S, Kramar Fijavž M, Mikoš M, Šraj M. Estimation of suspended sediment loads using copula functions. Water 2017;9:628. [CrossRef]
[35] Zhang Y, Kim CW, Beer M, Dai H, Soares CG. Modeling multivariate ocean data using asymmetric copulas. Coastal Eng 2018;135:91-111. [CrossRef]
[36] Zhang Y, Gomes, AT, Beer M, Neumann I, Nackenhorst U, Kim CW. Modeling asymmetric dependences among multivariate soil data for the geotechnical analysis-The asymmetric copula approach. Soils Found 2019;59:1960-1979. [CrossRef]
[37] Lin Y, Dong S, Tao S. Modelling long-term joint distribution of significant wave height and mean zero-crossing wave period using a copula mixture. Ocean Eng 2020;197:106856. [CrossRef]
[38] Bai X, Jiang H, Li C, Huang L. Joint probability distribution of coastal winds and waves using a log-transformed kernel density estimation and mixed copula approach. Ocean Eng 2020;216:107937. [CrossRef]
[39] Huang W, Dong S. Joint distribution of significant wave height and zero-up-crossing wave period using mixture copula method. Ocean Eng 2021;219:108305. [CrossRef]
[40] Wu S. Construction of asymmetric copulas and its application in two-dimensional reliability modeling. Eur J Oper Res 2014;238:476-485. [CrossRef]
[41] Shih JH, Emura T. On the copula correlation ratio and its generalization. J Multivariate Anal 2021;182:104708. [CrossRef]
[42] Dutang C, Charpentier A. Package 'CASdatasets', 2020. http://cas.uqam.ca/pub/web/CASdatasets-manual.pdf
[43] Frees EW, Wang P. Credibility using copulas. North Am Act J 2005;9:31-48. [CrossRef]
[44] Jung YS, Kim JM, Kim J. New approach of directional dependence in exchange markets using generalized FGM copula function. Commun Stat Simul Comput 2008;37:772-788. [CrossRef]
[45] Kim D, Kim JM. Analysis of directional dependence using asymmetric copula-based regression models. J Stat Comput Simul 2014;84:1990-2010. [CrossRef]

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1194–1201, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2023.00032

Research Article

The effect of photobioreactor height/diameter ratio on *Chlorella variabilis* microalgae growth and oil production efficiency

Necla ALTIN[1,*], Başar UYAR[1]

[1]*Department of Chemical Engineering, Kocaeli University, Kocaeli, 41001, Türkiye*

ARTICLE INFO

Article history
Received: 30 January 2023
Revised: 12 March 2023
Accepted: 12 April 2023

Keywords:
Chlorella Variabilis; Microalgae; Photobioreactor

ABSTRACT

This study aims to reveal how height/diameter ratio of column photobioreactors affect the growth and lipid content of microalgae. For this purpose, *Chlorella variabilis* cells were grown in aerated column photobioreactors with height/diameter ratio of 1, 2, and 3 in defined (BG11) culture medium. Results obtained showed that maximum microalgae biomass concentration, cell productivity, cell doubling time, and lipid productivity were found to increase as the height/diameter of photobioreactor increased. After 15 days of cultivation, the highest cell productivity (0.139 gdw/L.day), cell lipid content (21.1%) and lipid productivity (29.33 mg/L.day) were obtained in the photobioreactor with the highest height/diameter ratio (3), whereas the highest specific growth rate (0.045 h^{-1}) was obtained in the photobioreactor with the smallest height/diameter ratio (1). These findings contribute to the knowledge on photobioreactor design and pave way for more efficient use of column type photobioreactors in producing microalgae.

Cite this article as: Altın N, Uyar B. The effect of photobioreactor height/diameter ratio on *Chlorella variabilis* microalgae growth and oil production efficiency. Sigma J Eng Nat Sci 2024;42(4):1194–1201.

INTRODUCTION

In today's world, where oil prices have increased in recent years and fossil resources are decreasing, the fact that these resources will not be able to meet the needs in the near future has led people to turn to renewable energy sources. In addition, due to the negative effects of these resources on the environment, the search for alternative resources has become inevitable. The basis of the damage caused by fossil fuels to the environment is the increase in the amount of greenhouse gases released when they burn in the atmosphere. Greenhouse gases mainly consist of carbon dioxide (CO_2), methane (CH_4), nitrogen oxide (N_2O), water vapour (H_2O), ozone (O_3) and fluorinated compounds (CFC, HFC, PFC, SF_6). These gases come from the use of fossil fuels, industry, transportation, energy production, various wastes and agricultural activities. With the effects of global climate change showing more and more each passing day, studies to reduce the emission of greenhouse gases, which play the most important role in this change, have also gained importance.

***Corresponding author.**
*E-mail address: nejla_altin@hotmail.com

This paper was recommended for publication in revised form by Editor in Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey

In this period of increasing environmental awareness such as global climate change, reduction of CO_2 emission and efficient use of water resources, interest in microalgae has increased considerably in recent years and it is seen as one of the promising microorganisms in biofuel production.

The concept of biological fixation of carbon dioxide is a long-lasting technology to manage the sustainability of the environment. The substantial need to reduce carbon dioxide emissions, microalgae have proven to be the best fit for carbon fixation and production of diverse biofuels by converting atmospheric carbon dioxide into biomass via autotrophs. [1].

Microalgae are primary producers of synthesizing organic matter in the aquatic environment. Microalgae take in nutritional salts and combine them into complex molecules necessary for their vital activities with the help of a light source [2,3].

Microalgae are very important living things due to the protein, fatty acid, pigment substances, vitamins and valuable metabolites they have accumulated in their content [3–5].

For this reason, microalgae are used in many areas such as nutritional support for humans and animals, fertilizer in agriculture, wastewater treatment, cosmetics, and biofuel production. [6–9].

There are many reasons why microalgae are preferred as an energy source. These can be listed as being able to reproduce quickly, having a high oil content, being easily adaptable to various environmental conditions, etc. [10]. In addition, microalgae use sunlight 10 times more efficiently than terrestrial plants [11].

Biofuels derived from diverse microalgal species have grabbed a lot of attention from all over the globe due to rapid growth rate, high efficacy, no restriction of feedstock supply, high photosynthetic efficiency, less cost of production, high content of lipids, carbohydrates, etc. [12]. Significantly, microalgal strains are transforming nutrients into appropriate biomass and different cellular components [13,14].

Microalgae are photosynthetic creatures rich in protein, fat and carbohydrate derivatives. In general, they have oil content between 20-50% of their dry weight [6]. The main inorganic form of carbon required by microalgae is CO_2. This situation ensures that the storage materials of microalgae are mainly oil and starch [15]. Although it varies according to the species, microalgae can be produced quickly in open ponds or indoor bioreactors. They are usually autotrophic and perform photosynthesis.

In recent years, there have been significant developments in studies on closed photobioreactors. Therefore, the use of closed systems in the production of microalgae is becoming increasingly important. These studies aim to reduce the light path length and increase the light intensity reaching each cell. In addition, these systems should have features that allow a good mix of cultures [16].

Effects of height to diameter ratio of bioreactors are multifold: As the H/D increases, air residence time thus the mass transfer coefficient of oxygen and carbon dioxide also increases, mixing time which indicates homogenization efficiency of culture decreases, surface area per volume increases which increases manufacturing cost, lateral heat transfer rate and external illumination efficiency. On the other hand, as the photobioreactor height increases, liquid pressure at the bottom also increases which increases air solubility, air pump cost and power usage. All these factors create a complex problem which is a hurdle in design of photobioreactors. Moreover, their combined effect on microalgae growth needs to be shown identified for an efficient bioprocess operation. Therefore, studies that investigate H/D ratio on culture performance are important to ensure an efficient bioprocess and to provide basis for scale up studies.

Studies have been carried out to increase the oil accumulation of microalgae produced in photobioreactor systems with high production efficiency, and it has been concluded that the oil rate of the species produced in the environment where the nutrient is limited, and the production cost is lower [17].

In this study, the effect of photobioreactor height/diameter ratio on growth and oil content of *Chlorella variabilis* microalgae was investigated under laboratory conditions. In this scope, column photobioreactors with a diameter of 9 cm were operated at liquid levels of 9, 18, 27 cm corresponding to height/diameter ratios of 1, 2, 3, and *Chlorella variabilis* was cultured in BG11 defined medium under continuous illumination. Growth and lipid data were collected; specific growth rates, cell productivities and doubling times, lipid contents and productivities were calculated and reported. To the best of our knowledge, there is no similar research published on the literature. Thus, the findings of this study help fill the knowledge gap in a neglected aspect of photobioreactor design (height/diameter ratio), and pave way for more efficient use of column type photobioreactors in producing microalgae.

MATERIALS AND METHODS

Microalgae Strain and Nutrient Medium

Chlorella variabilis, belonging to the green algae group used in the research, was obtained from the Molecular Biology and Genetics department of Istanbul Medeniyet University. Blue-Green (BG-11) medium was used as the nutrient medium. Chemical composition of this medium was (g/L): $NaNO_3$ 1.5000, KH_2PO_4 0.0400, $MgSO_4.7H_2O$ 0.0750, $CaCl_2.2H_2O$ 0.0360, H_3BO_3 0.0029, Na_2CO_3 0.0200, Citric acid 0.0060, Iron citrate 0.0060, Vitamin 0.0200. Chemicals were analytical grade and were purchased from Merck.

EXPERIMENTAL SETUP AND OPERATING CONDITIONS

Microalgae were cultured in cylindrical bioreactors with different height/diameter (H/D) ratios. The bioreactors used are 9 cm in diameter and 9, 18 and 27 cm in

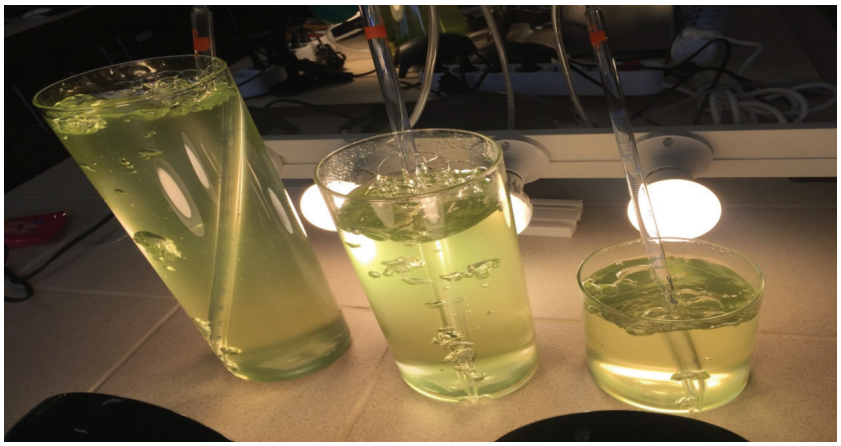


Figure 1. The photobioreactors with different height/diameter ratio.

height. Accordingly, the photobioreactor height/diameter ratios are 1, 2, and 3, respectively (Figure 1). The volumes of the bioreactors are 573, 1145 and 1717 mL, respectively. An average of 6% of the stock culture was inoculated. The illumination of the cultures was provided continuously (24 hours) by OSRAM brand-led bulbs with a 7W 2400K value, which would provide a surface light intensity of 3.80 Klux. The temperature was kept constant at 25 ± 2° C. The aeration of the samples was provided by an air pump at a rate of 300 L/h. The air from the air pump was sterilized by filtration (0.45μm) before being fed into the photobioreactors.

The Geometrical Interpretations of the Effect of Photobioreactor Height

For a typical cylindrical (column) photobioreactor, surface area shows the material amount needed to construct the bioreactor. Here, the shape that has the smallest surface area and therefore requires the least material to construct for a given volume has a height to diameter ratio of 1, and the material cost increases as this ratio increases. Moreover, for externally illuminated photobioreactors (almost all of the indoor photobioreactors, since internal illumination has its own challenges and not usually preferred), surface area per volume of the photobioreactor needs to be high to improve illumination efficiency which means height to diameter ratio should be high. Similarly, for jacketed vessels, heat transfer occurs through lateral surface of the cylinder, therefore high surface area (high aspect ratio) increase heat transfer rate and should be high. On the other hand, as the photobioreactor height increases, liquid pressure at the bottom also increases which increases air solubility, air pump cost and power usage. All these factors create a complex problem and combined effect on microalgae growth should be shown, which the objective of this study is.

Analyses

The growth of microalgae was monitored spectrophotometrically throughout the studies. A 3mL sample was taken from the photobioreactors to determine the optical density. Jenway 6800 UV-VIS spectrophotometer was used for analysis, samples were placed in quartz cuvettes and measurements were made at 600 nm wavelength. The pH (Mettler Toledo Seven Easy model) values of the culture medium were measured. After the studies were completed, the photobioreactor output was centrifuged at 4000 rpm for 30 minutes to determine the dry weight of the microalgae biomass. It was then dried at 60 °C until constant weight.

Lipid analysis was performed according to the method applied by Bligh and Dyer (1959) [18]. 120 mL of methanol-chloroform (1/2 ratio) mixture was added to 0.2 grams of homogenized algae samples. Then, 0.4% $CaCl_2$ solution was added to these samples. After these samples were mixed in a mechanical mixer at 200 rpm for 3 hours, they were filtered into tared flasks with the help of filter paper Figure 2 and Figure 3.

These flasks were kept for one day by closing their mouths in such a way that their mouths did not get airtight, and the top layer consisting of the methanol-water mixture was separated the next day. The chloroform was evaporated from the chloroform-lipid part remaining in the flasks and the remaining lipid was cooled to room temperature and the lipid ratio was calculated.

Calculations

The specific growth rates (μ) of the grown microalgae were calculated according to equation (1)

$$\mu = \frac{\ln X_{2-} \ln X_1}{t_{2-}t_1} \quad (1)$$

In the formula, X_2 and X_1 indicate the biomass concentrations at t_2 and t_1 times during the exponential (logarithmic) growth phase.

The doubling time (τ D) of the cells was found with the help of equation (2)

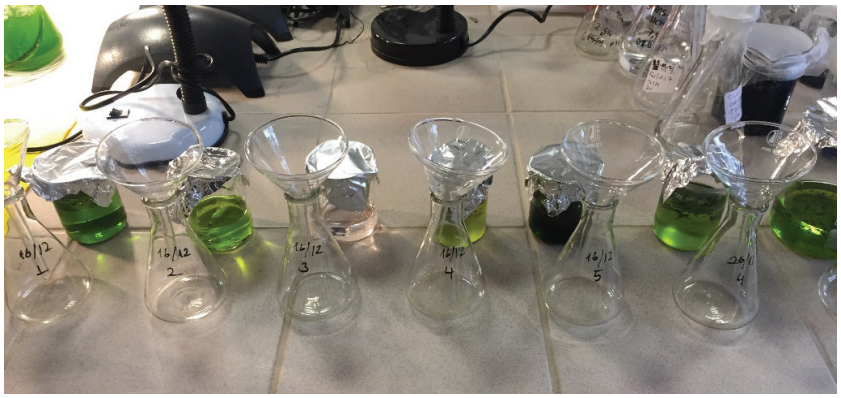


Figure 2. Oil analysis.

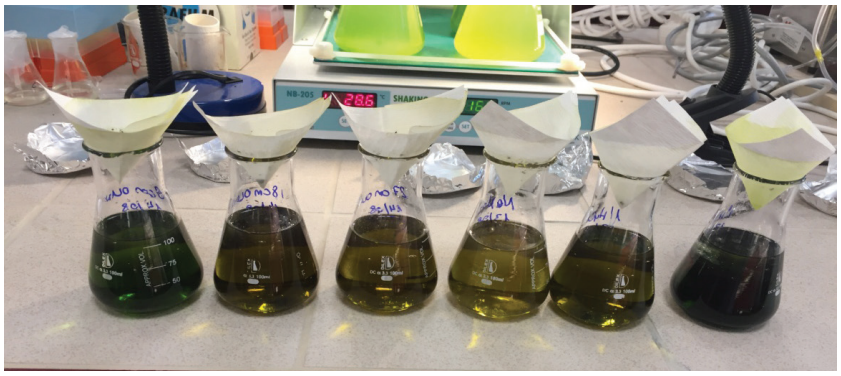


Figure 3. Filtration of solvent-added algae for oil analysis.

$$\text{Doubling time} = \frac{\ln 2}{\mu_{max}} \quad (2)$$

Cell productivity (P_x) is given as the amount of dry cells produced per liter per day (g_{dw}/L.day).

Oil productivity (P_{lipid}) was calculated after 15 days of incubation. Using Equation 3, the amount of lipid recovered (% of lipid) was calculated gravimetrically. Microalgae lipid productivity (mg/l/day) is the product of the lipid content value and the biomass productivity value, as shown in Equation 4.

$$\text{Lipid Amount (\%)} = [\text{Beaker Dare (g)} + \text{Lipid (g)}] - [\text{Beaker Dare (g)}] \times 100 \quad (3)$$

$$\text{Lipid Productivity (mg/l/day)} = \text{Lipid content (\%)} \times \text{Biomass Productivity} \quad (4)$$

Where P_x is cell productivity (gdw/L.day); P_{lipid}, lipid efficiency g/L.day); and % lipid indicates the lipid content (%) by mass of the cell.

RESULTS AND DISCUSSION

The effects of different height/diameter ratios on the growth of microalgae in the bioreactor were studied in detail. The design of a reactor for microalgae cultivation is a very important issue. Because mass transfer, light-dark region areas, and mixing factor affecting oxygen concentration are closely related. The height/diameter ratio (H/D) is closely related to the working efficiency of the reactor due to mixing in the photobioreactor. In this study, the reactor diameter was set to 9 cm, and the reactor heights were set to 9, 18 and 27 cm. The relationship between the maximum growth rate and height/diameter ratio of *Chlorella variabilis* cultured in bioreactors with different heights was investigated. The studies were carried out in 2 sets. The effects of photobioreactor height/diameter ratio (H/D) on the growth of *Chlorella variabilis* microalgae are shown in Figure 4 and pH changes are shown in Figure 5. The graph is shown as the average of two sets.

As can be seen from the graph, the specimens in all height/diameter ratios grew until the 69th hour and continued to grow at a steady rate after this hour. After the 284th hour in all samples, it was determined that the microbial activities of the microorganisms began to decrease and the death phase began.

The pH levels of microalgae are a critical parameter of growth conditions in bioreactors. The pH values, which were initially measured around 8.50, showed a tendency to increase for a certain period of time. Then, they varied between 10-10.5 during the study period. pH changes can

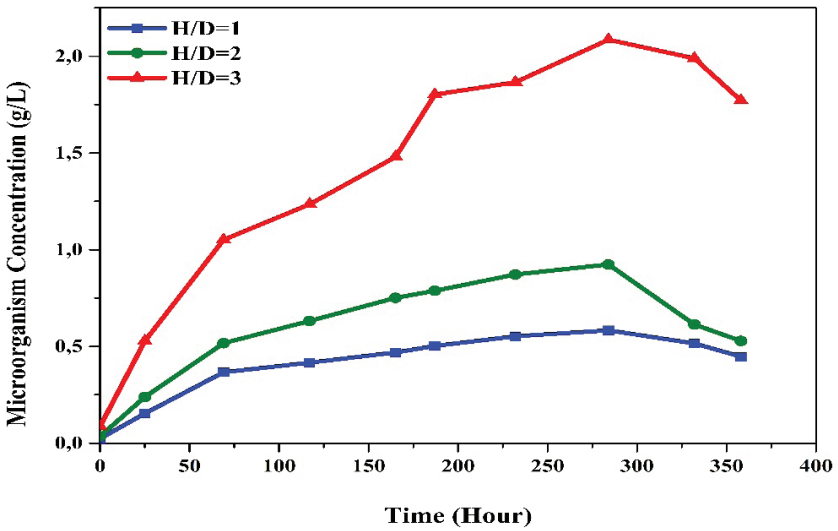


Figure 4. Effect of photobioreactor H/D ratio on the growth of *Chlorella variabilis*.

H: Bioreactor height D: Bioreactor diameter

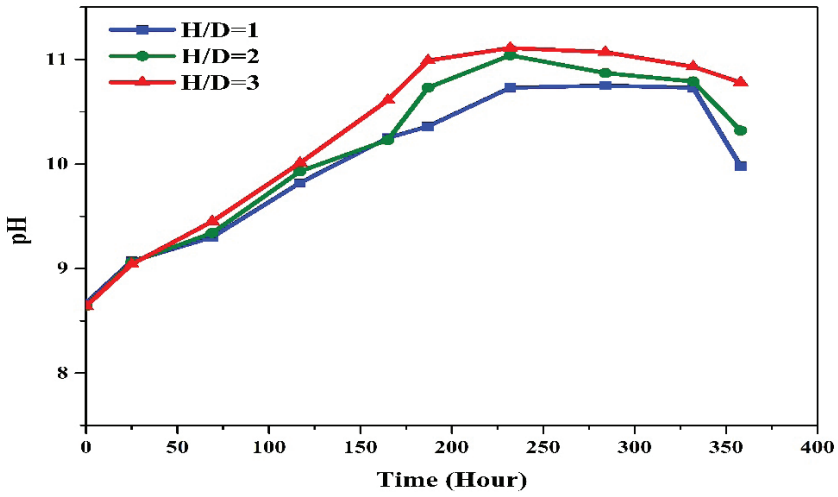


Figure 5. Effect of photobioreactor H/D ratio on pH change of *Chlorella variabilis*.

also affect the biochemical composition and lipid production of microalgae, an important factor in industrial applications such as biofuel production.

The effect of length-diameter ratio on microalgae growth is shown in Table 1. The values in the table were calculated as the average of the two sets.

The highest biomass amount was reached in the bioreactor with H/D=3 with 1.67 g/L. The biomass amounts of microorganisms in other bioreactors were 0.43 and 0.75 g/L, respectively. In this study, it was found that as the height/diameter ratio increased, the amount of biomass increased approximately 2 times. The maximum specific growth rate

Table 1. Effect of photobioreactor height/diameter ratio on the growth of *Chlorella variabilis* microalgae

H/D ratio	Dry weight (g L^{-1})	Px (gdw/L.day)	Growth rate μ ($hour^{-1}$)	Doubling time (hour)
1	0.43	0.039	0.045	15
2	0.75	0.062	0.039	17
3	1.67	0.139	0.037	19

Table 2. Effect of photobioreactor height/diameter ratio on the lipid content of *Chlorella variabilis* microalgae

H/D ratio	Fat Content (g/L)	Fat content (%)	Plipid ($mgL^{-1}day^{1}$)
1	0.046	18	6.98
2	0.087	17	10.6
3	0.22	21	29.33

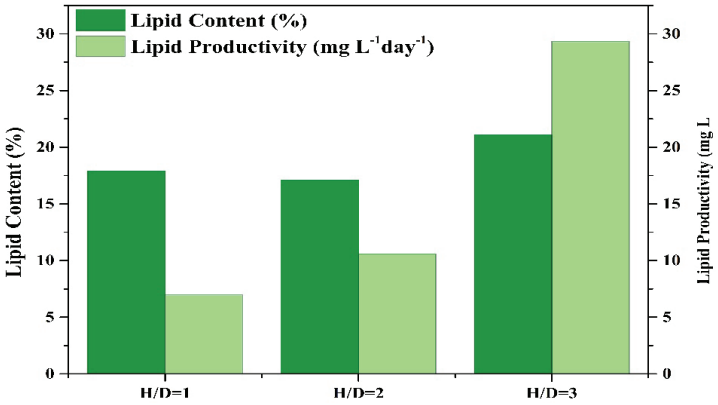


Figure 6. Lipid content and productivity of Chlorella variabilis microalgae grown at different height/diameter ratios.

was determined in the bioreactor with a growth rate of 0.045 h^{-1} and H/D=1. The doubling times were respectively; It is determined as 15, 17 and 19 h. These findings emphasize the importance of bioreactor design. In particular, it shows that the effect of H/D ratio on cell concentration is decisive. Higher H/D ratios generally provide more surface area, which facilitates better diffusion and transfer of nutrients and oxygen needed for cells. At the same time, less mechanical stress and lower shear forces can be applied compared to smaller diameter bioreactors, promoting healthy growth of cells. High H/D ratios also provide a more homogeneous distribution of nutrient substrates within the bioreactor and can contribute to the reduction of the gravity effect, which increases cell productivity and thus oil production efficiency.

The lipid content of microalgae grown in bioreactors with different H/D ratios is shown in Figure 6. Different H/D ratios favored lipid accumulation by microalgae (Table 2). Lipid contents were H/D=3 >H/D=1>H/D=2, respectively. At the end of the experiment, the highest amount of lipid was obtained at H/D=3 ratio (21%). This is 15% more than the amount of lipid in H/D=1 ratio.

Lipid productivity is as shown in Figure 6. The findings in this study showed that lipid productivity ranged from 6.98 to 29.33 mg L^{-1} day^{-1}. Lipid productivity was found to be higher at H/D=3 ratio. These differences in lipid productivity may be mainly due to differences in biomass as well as different H/D ratios. Because the deviations in lipid content were smaller.

Those findings are generally on a par than the ones reported in the literature previously; Rajapitamahuni et al., (2019) reported highest *C. variabilis* cell concentration as 0.76 g/L with a productivity of 0.022 mg/L/day in a continuous culture [19], Tran et al., (2020) determined maximum biomass concentrations of 1.52 - 1.72 g/L, with specific growth rates of 0.009 - 0.014 h^{-1} when *C. variabilis* was cultivated in domestic wastewater [20], Altın et al., (2018) reported *C. variabilis* maximum biomass concentration and specific growth rate as 1.30 g/L and 0.038 h^{-1}, respectively [21], and Uyar et al., (2016) calculated highest biomass concentration and specific growth rates as 0.30 g/L and 0.037 h^{-1} in a compartmentalized photobioreactor system [22].

Studies that focus on growing microalgae in larger scales are necessary to assess the performance and economy of the process. Notable literature with *C. variabilis* include a pilot scale study that used 35L closed panel photobioreactor in which 11 gdw biomass was obtained after 24 days [23], and industrial scale cultivation of *C. variabilis* in open solar pans that covers a vast area of 772 m^2 with a total cultivation volume of 360 m^3, where an average biomass productivity of 34.6 $g/m^2/d$ was achieved and through the application of solar the authors stated that the entire process was cost effective and energy efficient leading to the sustainable development of microalgae-based biofuel for future commercialization [24]. Other studies that employed different microalgae species in pilot scale photobioreactors, such as *Chlorococcu*m sp in 50 L outdoor photobioreactor [25], and *Tetraselmis* sp. in 80 L indoor photobioreactors [26], were also conducted to define and address issues related to design and technology of photobioreactors. In addition, scientific publications dealt with the analysis of production costs, energy profitability and product life cycle, which are especially important for the production of biofuels, and research results suggest that the economic

viability of producing microalgae solely for biofuel purposes is not cost-effective, highly uncertain and risky, whereas a financially viable process is possible if energy products are generated as by-products in a multifunctional biorefinery system, combined with, for example, carbon dioxide capture and wastewater treatment [27].

Stirred tank bioreactors usually have a height to diameter ratio in the range of 1 to 4, whereas airlift and bubble column bioreactors have higher ratio (4 to 8) to accommodate lower mass transfer of O_2/CO_2 due to lack of mechanical stirrer. This study was limited to H/D ratios of 1 to 3, thus higher ratios should be tested for a more comprehensive understanding of the effects. Studies with other microalgae strains are also required to complement these results.

Another problem that challenges large scale and low cost production of microalgae is the low production efficiency; microalgae require specific conditions for growth, including appropriate light, nutrients, and temperature. Scaling up the production of microalgae to an industrial level while maintaining optimum conditions can also be challenging. However, ongoing research and development is continually improving our understanding of microalgae's potential and as technology advances and new applications are discovered, algae may become a more viable option for technological applications in the future [6,13,14,28,29]

CONCLUSION

In this study, the relationship between the maximum growth rate and height/diameter ratio of *Chlorella variabilis* microalgae cultured in bioreactors with different heights was investigated. No such study has been found in the literature for this species. It was found that as the height/diameter ratio increased, the amount of biomass (and productivity) increased approximately four times, and the oil yield increased approximately five times. This methodology, described in short-term experiments, can be easily tested for other microalgae species. Because there is variation between species. It helps to optimize reactor designs to increase the performance and economic efficiency of microalgae-based industries.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Kumar A, Ergas S, Yuan X, Sahu A, Zhang Q, Dewulf J, et al. Enhanced CO_2 fixation and biofuel production via microalgae: Recent developments and future directions. Trends Biotechnol 2010;28:371-380. [CrossRef]

[2] Göksan T. The Growth of spirulina platensis in different culture systems under greenhouse condition. Turkish J Biol 2007;31:47-52.

[3] Spolaore P, Joannis-Cassan C, Duran E, Isambert A. Commercial applications of microalgae. J Biosci Bioeng 2006;101:87-96. [CrossRef]

[4] Del Campo JA, García-González M, Guerrero MG. Outdoor cultivation of microalgae for carotenoid production: Current state and perspectives. Appl Microbiol Biotechnol 2007;74:1163-1174. [CrossRef]

[5] Ward OP, Singh A. Omega-3/6 fatty acids: Alternative sources of production. Process Biochem 2005;40:3627-3652. [CrossRef]

[6] Chisti Y. Biodiesel from microalgae. Biotechnol Adv 2007;25:294-306. [CrossRef]

[7] Gladue RM, Maxey JE. Microalgal feeds for aquaculture. J Appl Phycol 1994;6:131-141. [CrossRef]

[8] Guerin M, Huntley ME, Olaizola M. Haematococcus astaxanthin: Applications for human health and nutrition. Trends Biotechnol 2003;21:210-216. [CrossRef]

[9] Olguín EJ. Phycoremediation: Key issues for cost-effective nutrient removal processes. Biotechnol Adv 2003;22:81-91. [CrossRef]

[10] Altın N. Chlorella varıabilis türü mikroalgin büyümesine ve yağ içeriğine etki eden parametrelerin belirlenmesi. (Master thesis). Kocaeli: Kocaeli University; 2017.

[11] Skjånes K, Lindblad P, Muller J. BioCO_2 - A multidisciplinary, biological approach using solar energy to capture CO_2 while producing H_2 and high value products. Biomol Eng 2007;24:405-413. [CrossRef]

[12] Peng L, Zhang Z, Cheng P, Wang Z, Lan CQ. Cultivation of Neochloris oleoabundans in bubble column photobioreactor with or without localized deoxygenation. Bioresour Technol 2016;206:255-263. [CrossRef]

[13] Bouabidi ZB, El-Naas MH, Zhang Z. Immobilization of microbial cells for the biotreatment of wastewater: A review. Environ Chem Lett 2019;17:241-257. [CrossRef]

[14] Mata TM, Martins AA, Caetano NS. Microalgae for biodiesel production and other applications: A review. Renew Sustain Energy Rev 2010;14:217-232. [CrossRef]

[15] Posten C, Schaub G. Microalgae and terrestrial biomass as source for fuels-A process view. J Biotechnol 2009;142:64-69. [CrossRef]
[16] Schenk PM, Thomas-Hall SR, Stephens E, Marx UC, Mussgnug JH, Posten C, et al. Second generation biofuels: High-efficiency microalgae for biodiesel production. BioEnergy Res 2008;1:20-43. [CrossRef]
[17] Sharma KK, Schuhmann H, Schenk PM. High lipid induction in microalgae for biodiesel production. Energies 2012;5:1532-1553. [CrossRef]
[18] Bligh, E.G. and Dyer WJ. Canadian Journal of Biochemistry and Physiology. Can J Biochem Physiol 1959;37:911–917. [CrossRef]
[19] Rajapitamahuni S, Bhayani K, Bachani P, Vamsi VB, Mishra S. An effective approach of bacterial siderophore as nitrogen source triggering the desired biochemical changes in microalgae Chlorella variabilis ATCC 12198. Algal Res 2019;43:101610. [CrossRef]
[20] Tran DT, Van Do TC, Nguyen QT, Le TG. Simultaneous removal of pollutants and high value biomaterials production by Chlorella variabilis $TH0_3$ from domestic wastewater. Clean Technol Environ Policy 2021;23:3-17. [CrossRef]
[21] Altın N, Kutluk T, Uyar B, Kapucu N. Effect of Different Nitrogen Sources on the Growth and Lipid Accumulation of Chlorella variabilis. J Appl Biol Sci 2018;12:38-40.
[22] Kutluk T, Kapucu N, Uyar B. Effect of light ıntensity on the growth of chlorella variabilis. Deu Muhendis Fak Fen Muhendis 2016;18:49. [CrossRef]
[23] Uyar B, Kutluk T, Özer Uyar GE, Kapucu N. Growth and lipid production of two microalgae strains in pilot scale (35 L) panel photobioreactors. J Adv Phys 2019;7:527-529. [CrossRef]
[24] Bhattacharya S, Maurya R, Mishra SK, Ghosh T, Patidar SK, Paliwal C, et al. Solar driven mass cultivation and the extraction of lipids from Chlorella variabilis: A case study. Algal Res 2016;14:137-142. [CrossRef]
[25] Gonzalez-Hita L, Tienza MB. (Determination of river-bed transport with iridium-192 in the upper reaches of the River Lerma). Isot Hydrol 1983 Proc Symp Vienna, (IAEA; STI/PUB/650) 1984:753-769.
[26] Pérez-López P, González-García S, Ulloa RG, Sineiro J, Feijoo G, Moreira MT. Life cycle assessment of the production of bioactive compounds from Tetraselmis suecica at pilot scale. J Clean Prod 2014;64:323-331. [CrossRef]
[27] Borowiak D, Krzywonos M. Bioenergy, biofuels, lipids and pigments-research trends in the use of microalgae grown in photobioreactors. Energies 2022;15:5357. [CrossRef]
[28] Rodolfi L, Zittelli GC, Bassi N, Padovani G, Biondi N, Bonini G, et al. Microalgae for oil: Strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. Biotechnol Bioeng 2009;102:100-112. [CrossRef]
[29] Lardon L, Hélias A, Sialve B, Steyer JP, Bernard O. Life-cycle assessment of biodiesel production from microalgae. Environ Sci Technol 2009;43:6475-6481. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00027

Research Article

Optimization of ultrasound-assisted extraction of protein from the by-product of the hazelnut oil industry using reverse micelles

Elif Meltem İŞÇİMEN[1,*], Mehmet HAYTA[1]

[1]*Department of Food Engineering, Faculty of Engineering, Erciyes University Kayseri, 38280, Türkiye*

ARTICLE INFO

Article history
Received: 25 January 2023
Revised: 03 March 2023
Accepted: 01 April 2023

Keywords:
AOT; Hazelnut Waste; FTIR; Optimization; Ultrasound

ABSTRACT

In the current research, an ultrasound-assisted extraction (UAE) procedure was established for the protein extracted employing a reverse micelles system (RMS) from the by-product of the hazelnut oil industry. The optimum extraction circumstances in the UAE were identified as a dioctyl sodium sulfosuccinate (AOT) an amount of 0.05 g/mL, a water content of 25.2 (W_0), a 0.02 g/mL solid-to-liquid ratio, ultrasound time of 17.52 min, ultrasound cycle 1, and ultrasound power 80% using the response surface approach. Under ideal circumstances, the maximum yield was recorded as 44.84 mg BSA/g of hazelnut meal protein (HMP) for RMS, it had a higher yield obtained by alkaline solution (AS). α-helix, β-turn, and β-sheet structures of HMP increased while the random coil decreased as evidenced by FTIR and SEM images proving that the cell walls were destructed and had more cracks in RMS. Overall, the findings indicated that UAE combined with RMS might be an effective approach for extracting protein from HMP and is likely to lead to an alternative evaluation possibility of an industrial by-product.

Cite this article as: İşçimen EM, Hayta M. Optimization of ultrasound-assisted extraction of protein from the by-product of the hazelnut oil industry using reverse micelles. Sigma J Eng Nat Sci 2024;42(4):1202–1213.

INTRODUCTION

Extraction from food waste has been increasingly popular as a means of decreasing environmental concerns and resources from waste plants [1]. It is becoming increasingly common to employ hazelnut by-products as natural antioxidants and functional food ingredients [2]. Turkey leads the globe in hazelnut production with 665 TMT in 2020 [3]. Hazelnut meal is produced as a feed ingredient that is obtained following the pressing and extraction of oil production from hazelnuts. However, it might be used as a nutritional raw material, because of its rich nutritional content [4]. The rheological properties and functional qualities of hazelnut meal products [5], microwave-assisted antioxidant compound extraction from hazelnut [2], edible film-forming potentials of hazelnut meal protein (HMP) obtained after hot extraction, acetone washing, or a combination of methods [6] have been reported. Although there are many investigations on HMP extraction,

*Corresponding author.
*E-mail address: eliferen@erciyes.edu.tr

This paper was recommended for publication in revised form by Editor in Chief Ahmet Selim Dalkilic

ultrasound-assisted extraction (UAE) by the reverse micelles system (RMS) has not been reported.

The UAE has recently gained importance as a method for extracting bioactive from plant materials and increasing extraction efficiency and decreasing extraction time [7]. The matrix is mechanically affected by ultrasound, which causes it to break apart and form smaller particles, giving the extraction fluid more surface area [8]. A study employed UAE with RMS due to the obtained protein from wheat germ [9].

In nonpolar solvents, surfactant molecules aggregate into nanometer-sized aggregates with water molecules as inner cores. RM are biotechnologically essential due to their capacity to saturate their polar cores with water and hydrophilic molecules like proteins [10]. Surfactant-mediated extraction might be used to extract bioactive components such as polyphenols, lectins, and proteins [11-13]. RMS divides protein extraction into two stages: forward and backward extraction. Proteins solubilize into the RMS in the forward extraction process, whereas during the backward extraction phase, the solubilized proteins recover from the RMS [14]. RMS has various benefits over other protein processing methods, including the retention of natural functional characteristics, little interfacial tension, high yield, simplicity of scaling up, and the ability to process continuously [15]. The characteristic properties of protein change with RMS extraction especially the secondary structure of the protein [16, 17]. When compared to proteins obtained in an aqueous buffer, a reduction in the percentage of β-turn whereas a rise in the β-sheet, α-helix, and random coil proportion have been reported [18].

Therefore, in this research, UAE and RMS were utilized to improve the forward and backward extraction efficiency of HMP. The six-factor (AOT concentration, water content, ratio, ultrasound pulse and power, and time) were evaluated by one-factor analysis before the optimization process. One-factor findings based on ultrasonic power, application time, and the ratio was applied to response surface methodology (RSM). Additionally, the secondary structure and protein content of HMP extracted at its peak using RMS and alkaline solution (AS) were studied. The objective of the current investigation was to employ a novel extraction method that would offer better protein yields and greater activity.

MATERIALS AND METHODS

Materials

Dioctyl sodium sulfosuccinate (AOT) (D201170), potassium chloride (KCl) (44675), and isooctane (104727) were bought from Sigma-Aldrich (St. Louis, MO, USA). The substances utilized in the current investigation were of analytical grade. Hazelnut meals were obtained from the oil industry.

Preparation Of Reverse Micelles System (RMS)

Isooctane, AOT, and KCl phosphate buffer (pH=7.5) were used to prepare RMS. Firstly, AOT was dissolved in the magnetic stirrer at 25°C and the phosphate buffer containing 0.1 M KCl was added according to the molar ratio of water to RM (W_0= [H2O]/[AOT]) [19].

One Factor Tests

For UAE with RMS of HMP, there are a few key aspects to consider; AOT concentration, W_0, ratio, ultrasound pulse, power, and application time. Therefore, the AOT concentration of 0.02-0.1 g/mL, water content (W_0) of 5-30, the ratio of 0.01-0.1 g/mL, ultrasound pulse of 0.1-1, ultrasound power of 20-100% amplitude, and application time of 5-20 min. were optimized by RSM of Design Expert (Trial Version 7.0.0, Stat-Ease Inc., USA) with one-factor analysis [9,16].

Ultrasound-Assisted Extraction (UAE) of HMP

HM was added directly to the extraction process in 20 mL of surfactant solvent into the beaker. The probe of the ultrasound device (UP400S, Hielscher, Germany) was submerged center point of the mixture. After the extraction process, for 10 min, the fluid was centrifuged at 4000 g [19]. After centrifugation, the supernatant of the RM phase loaded with HMP of forward extraction was gathered and the equivalent amount of (pH 7.5) aqueous phase KCl in the amount of 1 mol/L was added. The backward extraction procedure was performed in a magnetic stirrer for 1 h at 25 °C. The protein-rich phase was gathered after centrifugation (4000g,10 min). The Bradford technique was used to determine the protein concentration of the HM extract [20]. A standard curve was established using the BSA.

Optimization by Response Surface Methodology (RSM)

The surfactant ratio, W_0, and ultrasound pulse were selected according to a one-factor experiment for further Box-Behnken design (BBD) of RSM of Design-Expert software. The solid-to-liquid ratio, ultrasonic power, and application duration were all optimized to perfection. The findings of the preliminary studies were used to define these optimal ranges, which included application time (X_1) 10-20 min., ratio (X_2) 0.02-0.06 g/mL, and ultrasound power (X_3) 40-80 % amplitude. The protein content of the forward extract was used response factor. In the design, five central points were chosen. As a result, extraction was carried out at a total of seventeen separate points (Table 1). Samples were prepared as 20 mL in beakers with a capacity of 100 mL. After UAE, centrifugation at 4000 g for 10 min separated the undissolved residue, and the content of protein in the supernatant was assessed using the Bradford test.

The suitability and fitness of the model were assessed using ANOVA. Based on the results of the RSM optimization inquiry, the impacts of quadratic effects were

determined using a second-order polynomial model [21].

Backward Extraction of HMP

The supernatant of the RM phase from the forward extraction that was loaded with HMP was collected, and an equivalent volume of the (pH 7.5) aqueous phase with 1 mol L^{-1} of KCl was introduced. The backward transport method was carried out for one hour at 25°C in a magnetic stirrer. Following centrifugation of the mixture (4000 g, 10 min) [22]. The protein-rich solution was collected and the protein was precipitated with the ternary liquid system at 25°C. Acetone, deionized water, and isooctane were combined in a ternary liquid system with a volume ratio of 15:5:1. To eliminate any remaining surfactant, the HMP precipitate was washed with a 65% ethanol solution [23]. Finally, the HMP underwent freeze-drying (Christ Alpha, 1 2 LD plus, Germany), and stored at -20°C.

Alkaline Extraction

The alkaline solution (AS) (pH 12.0) was used for conventional extraction [21]. The protein was extracted using optimal extraction conditions that had been determined by BBD. A ternary liquid system was then used to precipitate the protein in the supernatant after the mixture had been centrifuged at 4000 g for 10 min. The HMP was freeze-dried in a lyophilizer (Christ Alpha, 1 2 LD plus, Germany) and stored at -20°C for further analysis [16].

Scanning Electron Microscopy (SEM)

The microstructure properties of lyophilized HMP obtained by UA-RMS and UA-AS were investigated by an SEM (Leica 231 LEO: S-440, Cambridge, USA). A thin gold coating was applied to the HMP. A 25 kV excitation voltage was used for visualization.

Fourier Transform Infrared (FTIR) Spectroscopy

The secondary structure of HMP samples was determined by FTIR spectroscopy (Spectrum 400, PerkinElmer Instruments, Waltham, USA). ATR unit was used with 4 cm^{-1} resolutions. The measurements were made using a 100 scan speed in the wavenumber range of 450-4000 cm^{-1}.

X-Ray Diffraction (XRD)

A diffractometer (D8 Advanced, Bruker AXS GmbH, Karlsruhe, Germany) was used for the analysis. 45 kV and 40 mA were chosen as the voltage and current, respectively. Cu-K nickel was used to filter the radiation, and the wavelength was 1.5406. The scanning speed was 5°/min. Diffraction data were gathered in the 2θ range of 5° to 70° [24].

Statistical Analysis

Design Expert (Trial Version 7, Stat-Ease Inc., Minneapolis, MN) was used for the optimization process. The suitability of the model was evaluated using the coefficient of determination (R^2) of ANOVA, lack of fit-derived parameters, and the F-test.

Table 1. Protein extraction from hazelnut by-products using the Box-Behnken Design by UAE-RMS

Run	Time (min.) (X1)	Solid/liquid (g/mL) (X2)	Power (amplitude %) (X3)	Protein (mg BSA/g)
1	15.00	0.02	40.00	30.78
2	15.00	0.04	60.00	22.14
3	20.00	0.04	80.00	20.33
4	15.00	0.04	60.00	16.28
5	10.00	0.04	80.00	18.39
6	10.00	0.06	60.00	10.98
7	15.00	0.04	60.00	18.92
8	10.00	0.04	40.00	17.84
9	20.00	0.06	60.00	10.83
10	20.00	0.04	40.00	14.61
11	20.00	0.02	60.00	33.53
12	15.00	0.06	80.00	11.37
13	15.00	0.02	80.00	46.14
14	15.00	0.04	60.00	17.34
15	15.00	0.06	40.00	14.42
16	15.00	0.04	60.00	15.97
17	10.00	0.02	60.00	30.66

RESULTS AND DISCUSSION

Optimization of Ultrasound-Assisted Reverse Micelle by One Factor

The water content and AOT concentration are significant parameters affecting HMP yield. Firstly, the AOT concentration was optimized by one-factor analysis under the following condition: W_0 20, ultrasound power 50%, time 10 min., ratio 0.03 g/mL, ultrasound cycle 1. The most effective AOT concentration was found at 0.05 g/mL according to Figure 1a. The HMP yield was increased with an increasing AOT ratio at 0.06 g/mL however, it decreased after 0.06 g/mL. Similarly, in a study, AOT concentration which ranged from 0.04 to 0.08g/mL was optimized for protein extraction from wheat germ. Protein content increased quickly as concentration went from 0.02 to 0.06 g/mL. and as the concentration of AOT rose, the efficiency of forward extraction began to decrease. [19]. These findings concurred with the in another investigation [25]. A faster protein extraction yield of watermelon seed was reported when surfactant concentrations changed from 0.04 to 0.10 mol/L. However, protein yields insignificantly increased after 0.10 mol/L AOT [16].

Subsequently, different water ratios (W_0: 5-30) have been optimized by keeping the AOT ratio at optimum. It was seen the optimum water content was found as 25.2 (Figure 1b). Firstly, the increase in RMS water content increased HMP yield. The increments of protein yields were insignificant after then 25 W_0. High surfactant concentrations and water content have been demonstrated to enhance the number of surfactant aggregations and modify the surfactant size, hence improving protein transfer into RMS [26]. The optimum W_0 was reported as 20 for protein extraction from watermelon seed [16] and 25 from the wheat germ [19]. The increase in W_0 correlated with an increase in micelle size, and the micelle size highly influenced protein solubilization. The capacity of a protein micelle to be solubilized depended critically on its size. [27]. Larger RM was generated when W_0 was increased, allowing for the inclusion of several protein molecules [28].

The ultrasound power was between 20% and 100% amplitude. The other factor was selected as the concentration of AOT 0.05 g/mL, W_0 25.2, time 10 min., ratio 0.03 g/mL, ultrasound cycle 1. The HMP yield was increased with increasing power up to 80% amplitude. The optimum power was selected as 65.9 % amplitude according to a one-factor design (Figure 1c). Higher protein content at higher ultrasonic amplitudes might be related to structural cell damage caused by ultrasound [29]. However, it has been reported that the power set above 363 W, and the efficiency of UAE did not improve significantly [30]. The application time ranged from 5 to 20 min (Figure 1c). The optimum HMP extraction time was determined as 15.4 min. and the cycle was selected 1 according to Figure 1d.

The ratio was an optimized one-factor test in the present study as a crucial independent variable that affects the dissolving balance of the sample. The effects of the ratio were investigated in the following conditions: concentration of AOT 0.05 g/mL, W_0 25.2, time 10 min., power 65.9%, and ultrasound cycle 1. According to Figure 1e, in the RMS, HMP dissolution approached equilibrium when the ratio was 0.02 g/mL. As a result, The ideal operating of the surfactant process by UAE was found to be 0.02 g/mL. A similar result was reported for protein extraction from watermelon seeds [16].

Optimization of UAE-RMS

Three parameters; time (X_1) solid to liquid ratio (X_2), and power (X_3) were enhanced via the BBD of RSM. The RSM is a popular method for determining parameters and analyzing response factors [31]. The experimental settings and relevant response levels based on the experimental plan are listed in Table 1. For the recovery of protein from HM, the quadratic model was proposed by the model. The following quadratic regression model illustrates the connection between these factors and overall HMP extraction yields.

$$HMP= 18.13 + 0.18X_1 - 11.69X_2 + 2.32X_3 - 0.75X_1X_2 + 1.29X_1X_3 - 4.60X_2X_3\ 2.26X_1^2 + 5.63X_2^2 + 1.92X_3^2$$

The ideal HMP extraction parameters were found to be 17.52 min, 0.02 g/mL, and 80% amplitude. As the determined predicted protein yield by RSM was 44.84 mg BSA/g. The model worked well because there wasn't a big gap between experimental and projected values.

Using the p-value, the statistical significance of the interaction of independent variables was evaluated. The significance and suitability of the quadratic model were assessed using ANOVA data and the results of the ANOVA were shown in Table 2. While the p-values of a model, solid-to-liquid ratio, and power were found as <0.05, time had a p-value greater than 0.05. The lack of fit test, which determines if models are relevant to entirely predict variance, was used to verify the validity of models [32]. The lack of fit value of the model was determined insignificant as >0.05. The determination coefficients (R^2) value, adj. R^2 value and pred. The R^2 value of the sample was 0.97, 0.93, and 0.79, respectively. The optimization model of BBD for HMP extraction by UAE-RMS was statistically significant and the experimental data were appropriately fitted by the model.

Figure 2 depicts the interaction of the variables as well as the impact of independent variables on response. The 3-D plot in Figure 2a demonstrated that as ultrasonic power increased, so did the yield of protein extraction. However, the increasing time does not significantly affect extraction yield. The extraction yield suffered as the ratio increased (Figure 2b). Similarly, the ratio showed a similar trend (Figure 2c).

COMPARISON: UAE-RMS AND UAE-AS AND KINETICS

The best-operating conditions for one factor were: AOT concentration, 0.05 g/mL; $W_{0,}$ 25.2; ultrasound cycle, 1. According to the optimization process of BBD, the ratio is 0.02 g/mL; ultrasound power is 80% amplitude; and time is 17.52 min. UAE- RMS was used to extract the protein under ideal circumstances and the protein obtained UAE-AS at 0.02 g/mL; ultrasound power, 80% amplitude; time, 17.52

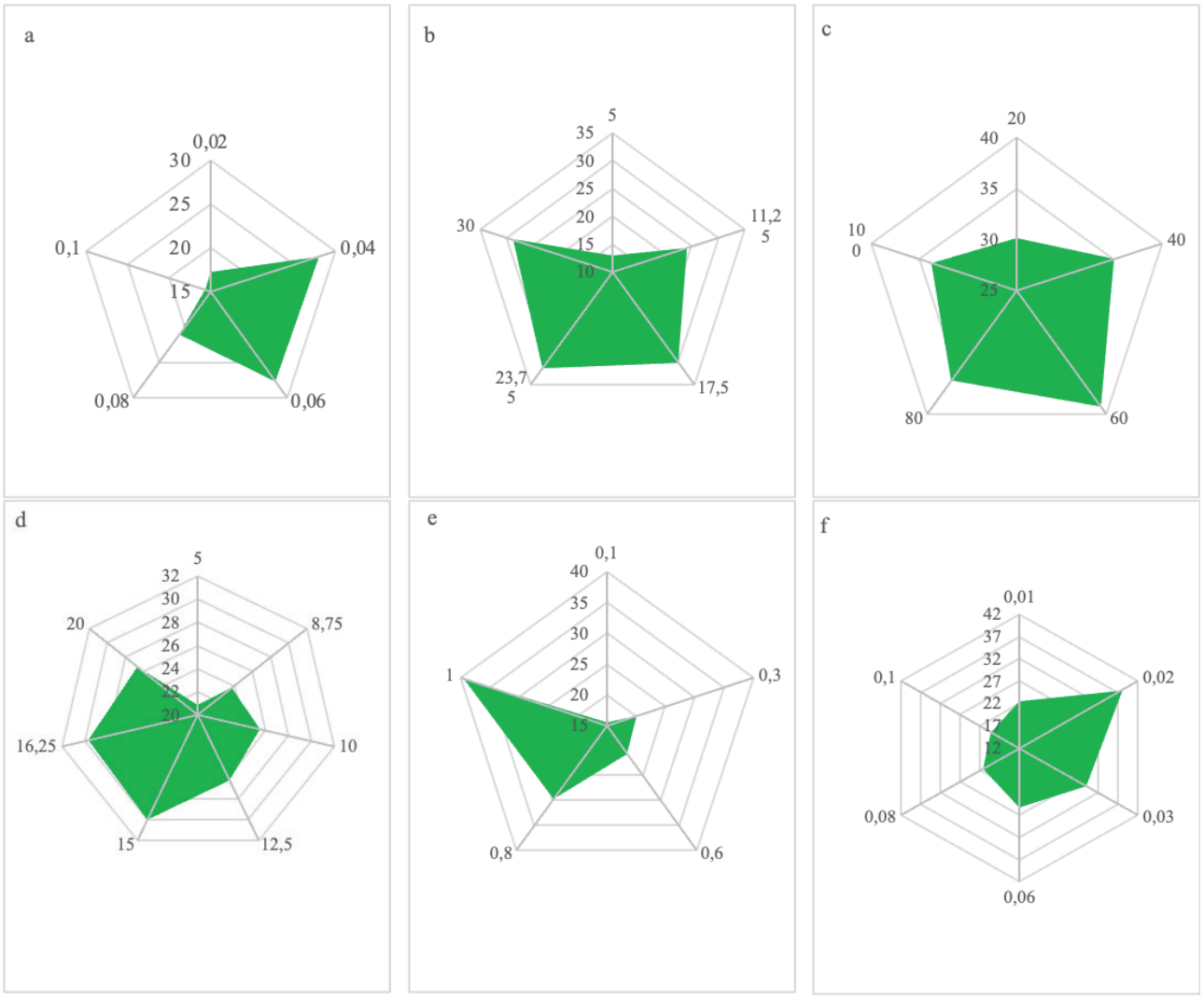


Figure 1. The one-factor test of ultrasound-assisted extraction (UAE)-reverse micellar solution (RMS).

The impact of AOT ratio on protein production (a); the effect of water (W0) on protein production (b); the effect of ultrasound power on protein production (c); the effect of time on protein production (d); the effect ultrasound cycle on protein production (e); the impact of the solids-to-liquids ratio on protein production (f).

One Factor	p-value				
	Model	Lack of fit	Independent variable	Point prediction	R^2
AOT (g/ mL)	0.00	0.41	0.01	0.05	0.99
W0	0.00	0.99	0.00	25.2	0.96
Power (% Amplitude)	0.01	0.38	0.04	65.9	0.91
Time (min)	0.01	0.24	0.00	15.4	0.93
Cycle	0.00	0.10	0.00	1.00	0.98
Solid/liquid (g/mL)	0.00	0.42	0.00	0.02	0.10

Table 2. ANOVA for the recovery of protein from the by-product of hazelnut

	Sum of squares	DF	Mean square	F-value
Model	1398.54	9	155.39	26.18****
X_1: Time (min)	0.25	1	0.25	0.043ns
X_2: Solid/liquid (g/mL)	1092.87	1	1092.87	184.12****
X_3:Power (%, amplitude)	43.17	1	43.17	7.27*
X_1X_2	2.27	1	2.27	0.38ns
X_1X_3	6.67	1	6.67	1.12ns
X_2X_3	84.81	1	84.81	14.29**
X_1^2	21.49	1	21.49	3.62ns
X_2^2	133.43	1	133.43	22.48**
X_3^2	15.52	1	15.52	2.61ns
Residual	41.55	7	5.94	
Lack of fit	16.14	3	5.38	0.85ns
Pure error	25.41	4	6.35	
Core total	1440.09	16		
R-Squared 0.97 Adj R-Squared 0.93 Pred R-Squared 0.79 Adeq Precision 18.76				

*p≤ 0.05; **p≤ 0.01; ***p≤ 0.001; ****p≤ 0.0001; ns p>0.05

min. When the extraction yield was compared, it was seen that the RMS was more effective than AS. The protein content was found for RMS and AS as 53.58±0.4 and 22.62±1.00 mg BSA/g, respectively. The variation of protein yield over time for both systems was given in Figure 3. The second-order kinetic model (Eq. 1) was used to evaluate the extraction's kinetics [33]. Moreover, the initial extraction rate is specified by Eq. 2 as time t approaches zero and the root means square error (RMSE) was determined according to Eq. 3 using the Statistica (Tibco Software Inc. USA).

$$CL(t) = \frac{Ce^2kt}{1+Ce\ kt} = \frac{Cet}{\left(\frac{1}{Cek}\right)+t} \quad (1)$$

$$h = kCe^2 \quad (2)$$

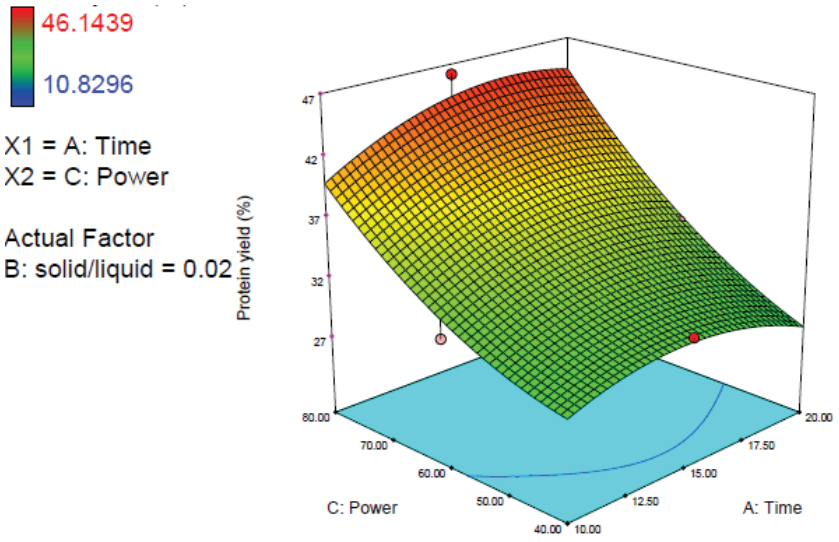
$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{[CL,p(t)-CL,e(t)]^2}{n}} \quad (3)$$

Time, t (min), second-order release rate constant, k (g mg^{-1} min^{-1}), initial releasing rate, h (mg g^{-1} min^{-1}), the concentration of protein, Ce (mg/g), CL, p model-predicted protein content (mg L^{-1}) CL, e protein content obtained experimentally (mg L^{-1}).

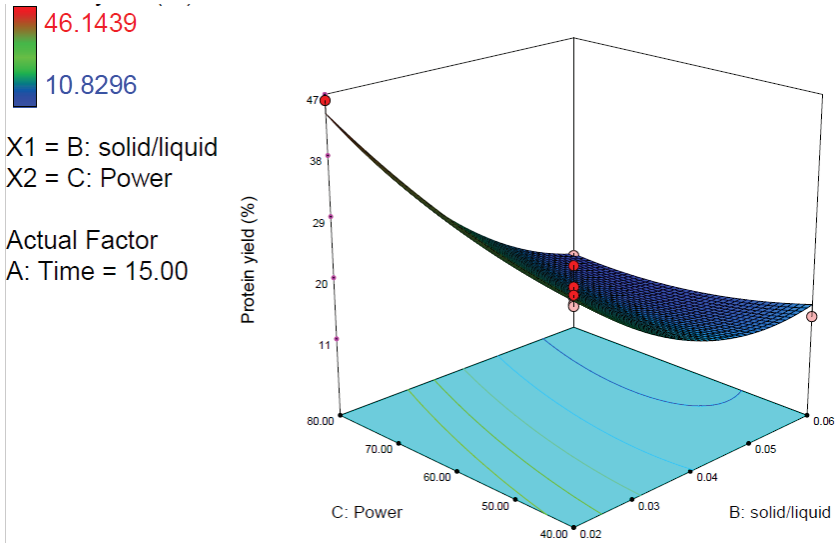
The h and k value was found as 54.15 mg g^{-1} min^{-1} and 0.024 g mg^{-1} min^{-1} for RMS, respectively, and higher than AS. This verified that the UAE-RMS could greatly improve the extraction rates of protein from HM. A similar study reported that the RM extraction of AOT-SDS/isooctane was an efficient and successful method for separating full-fat peanut powder sources of peanut protein [34]. In addition, researchers have examined the dynamics of protein partitioning for batch protein extraction using RM phases containing AOT in isooctane [10]. In another comparison study, the protein yield of the RMS was found more effective than the AS [16].

SEM Analysis of HMP

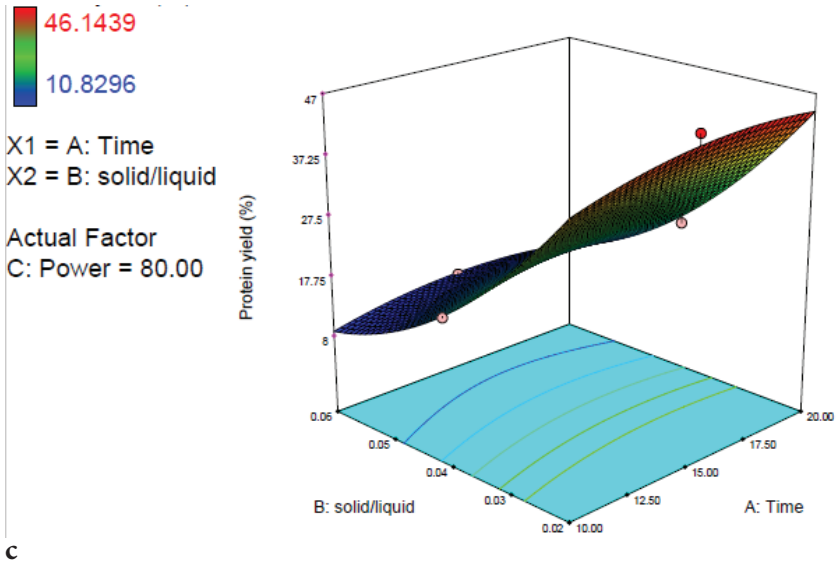
Figure 4 shows the SEM micro-images of HMP after UAE-RMS (Figure 4 a, b) and alkaline extraction (Figure 4 c, d). The cell structure of the HMP extracted by AS was more intact and smooth according to the HMP extract obtained from RMS. This finding suggested that the ultrasound application with RMS might cause cell walls to be destroyed. The cell walls were destructed and had more cracks in RMS. A similar result was reported for a watermelon seed protein extraction [16]. Literature suggested that the RM approach might alter walnut microstructure [24]. Similarly, a study reported that peanut proteins obtained by RMS had more holes on their surfaces, which were made up of a discontinuous and loose network according to proteins abstained by the aqueous buffer [24]. When the persistent network transitions to the discontinuous phase, RM may produce hydrophobic interactions, electrostatic contacts, and, hydrogen bonds acting as a junction zone and providing additional structural support [35]. It has been known that protein microstructure might



a

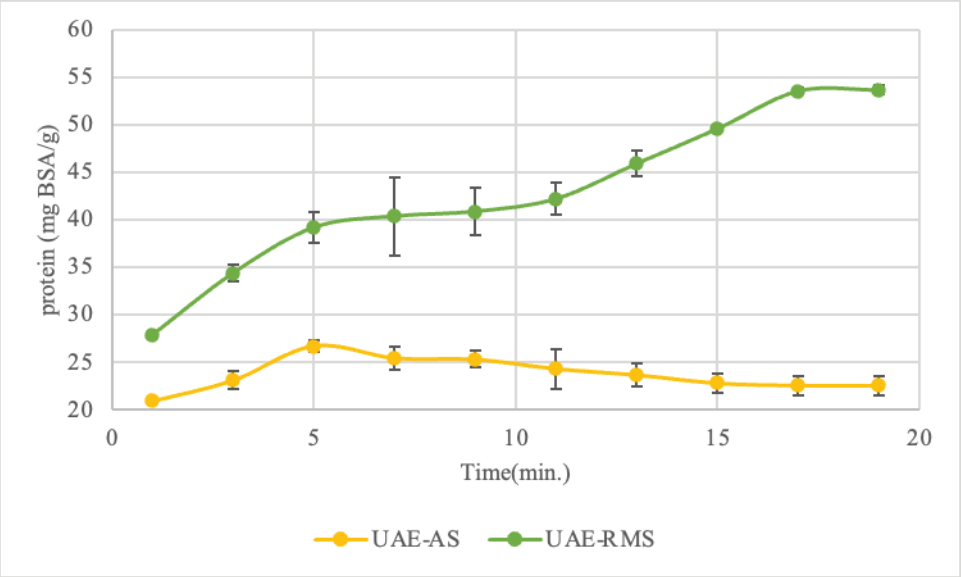


b



c

Figure 2. 3D plots optimization by Box Behnken Design (BBD)



	h	k	Ce	RMSE
UAE-RMS	54.15	0.024	47.50	3.62
UAE-AS	40.03	0.015	51.66	1.36

Figure 3. The comparative study and kinetics of ultrasound-assisted extraction-reverse micelles system (UAE-RMS) and ultrasound-assisted extraction-alkaline system (UAE-AS).

h: initial releasing rate (mg g^{-1} min^{-1}); k: second-order release rate constant (g mg^{-1} min^{-1}); Ce: concentration of protein (mg/g), RMSE: root mean square error (mg/g).

influence the physicochemical properties leading to alterations in and functional performance of proteins [36, 37].

FTIR Spectroscopy

The FTIR approach is effective in determining the secondary structure of the protein. The functional groups in the two proteins were identified using FTIR spectra in the 400-4000 cm^{-1} range. (Figure 5 a). RMS and AS were used to quantify the secondary structure of proteins that the UAE had retrieved, and the data from curve fitting was standardized for the amide I region. (1700-1600 cm^{-1}) (Figure 5 b,c). The amide I band was superimposed using a Gaussian curve fitting approach by Origin, and the results showed that the detection method was suitable. The compositions of HMP secondary structures were also shown in Table 3.

The four main parts of the protein secondary structure are α-helix, β-sheet, β-turn, and random coil structures. The internal structure of a protein which is α-helix and β-sheet is determined by intermolecular hydrogen bonds, which indicates the protein's high order. The flexibility of protein is linked to the random coil [16]. For the α-helix structure, bands between 1650-1663 cm^{-1} were used. The region of the β-sheet was determined according to bands 1612-1640 cm^{-1}, 1670-1694cm^{-1}. When the β-turn was investigated according to 1664-1684 cm^{-1},1694-1696 cm^{-1}, the random coil was determined as stated by 1640-1650cm^{-1} [24, 38]. In Figure 3a, it was seen that the sharp peak was observed between 1612-1640 cm^{-1}. However, the peak of UAE-AS was weaker than UAE-RMS. These results indicated that a higher β-sheet structure of HMP was obtained via RMS.

When RMS was utilized as a solvent, the quantity of α-helix, β-sheet, and β-turn structure increased and the amount of random coil structure decreased (Table 3). The α-helix, β-sheet, and β-turn structure may convert into a random coil shape due to the varying microenvironment for protein extraction [39]. The increase in random coil structure might be due to protein denaturation/unfolding after standard AS extraction [40-42]. In the present study, Variable degrees of random coil structures were introduced into the β-sheet, α-helix, and β-turn structures., implying that HMP was denatured by alkaline extraction. Similar to the current study, an increase in the α-helix and β-turn structures in RMS was reported. Watermelon seed protein found an increase in the β-sheet and β-turn in RMS, and a decrease in the α-helix structure [16]. The findings of the current study suggest that the protein function or activity of HMP extracted by UAE-RMS maintains, however in various microenvironment settings, and the secondary structure compositions of HMP fractions might alter.

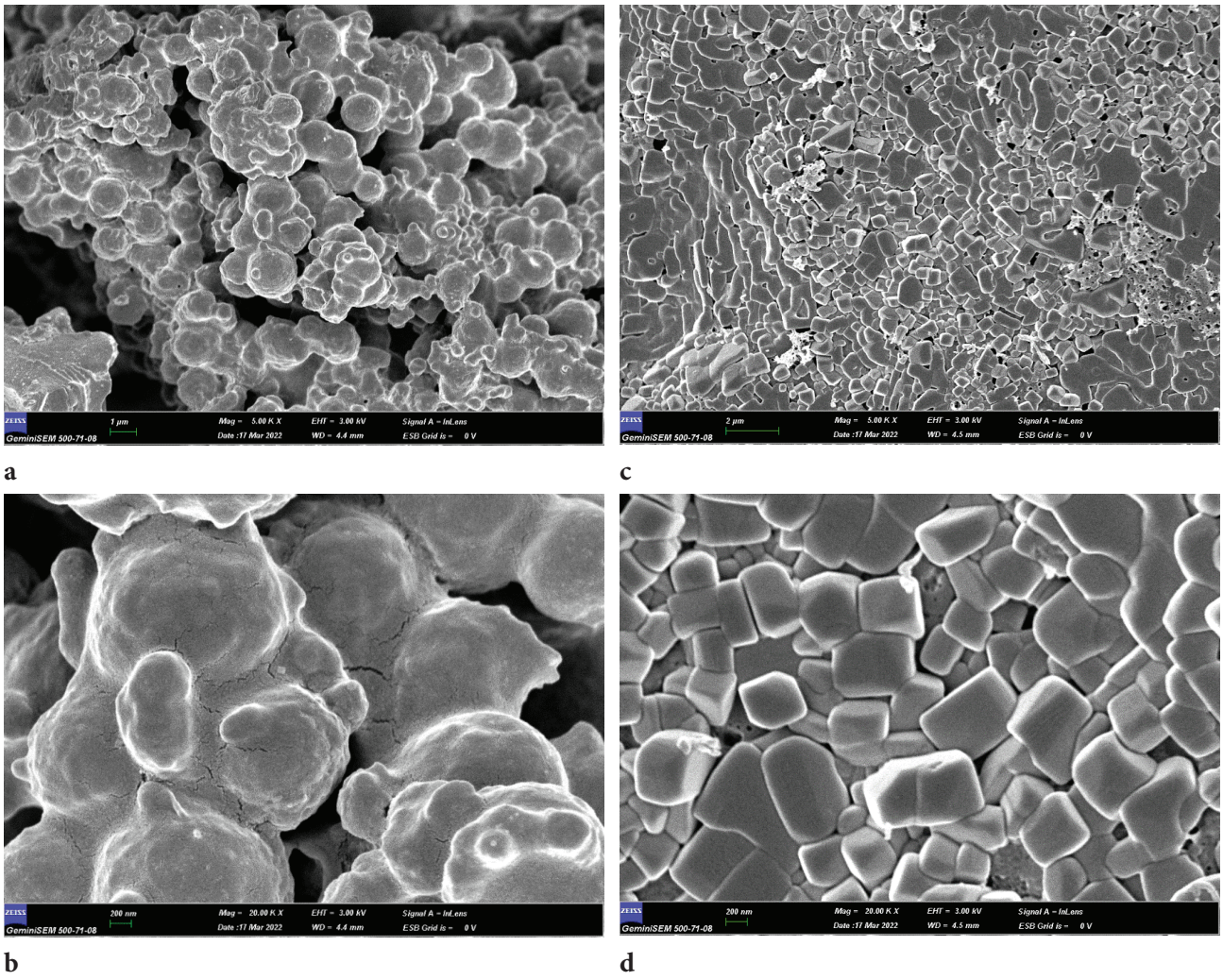


Figure 4. Scanning electron microscopy (SEM) spectra hazelnut meal protein (HMP) obtained by ultrasound-assisted extraction (UAE)-reverse micellar solution (RMS) (a,b); UAE-alkaline solution (AS) (c,d).

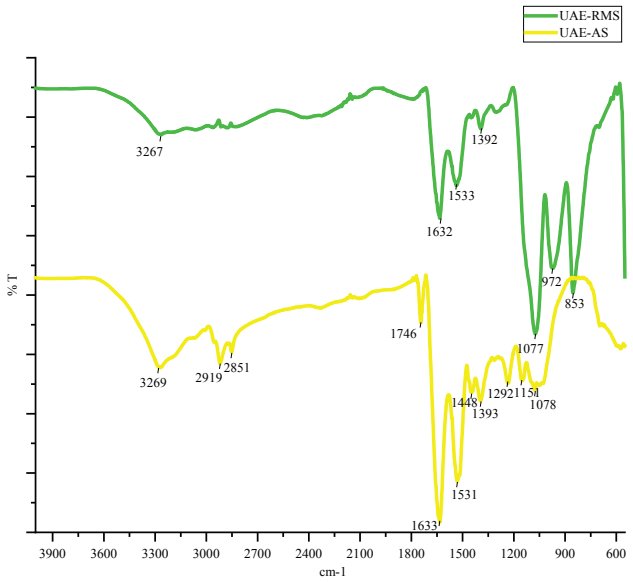
Table 3. FTIR analysis

	α-helix (%) (1650-1663 cm^{-1})	β-Sheet (%) (1670-1694cm^{-1} 1612-1640 cm^{-1})	β-turn (%) (1694-1696 cm^{-1} 1664-1684 cm^{-1})	Random coil (%) (1640-1650cm^{-1})
UAE-RMS	14.74	31.66	21.45	32.13
UAE-AS	7.28	27.2	17.3	48.21

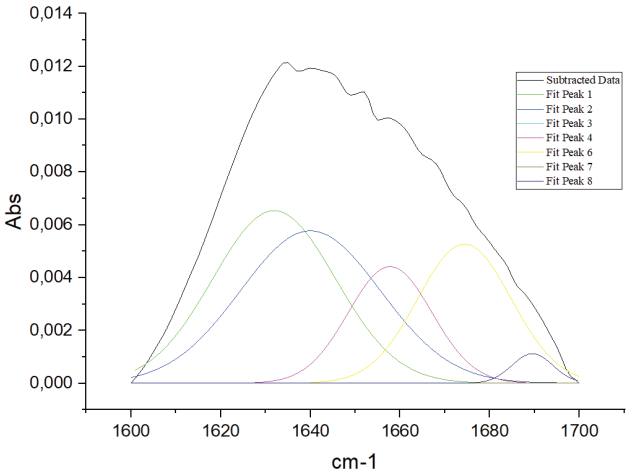
UAE-RMS: Ultrasound assisted extraction- reverse micelles system.
UAE-AS: Ultrasound-assisted extraction- alkaline solution.

XRD

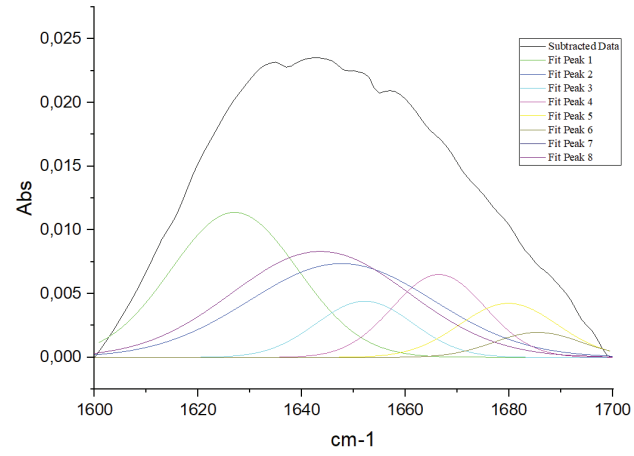
The use of XRD patterns to determine the crystallinity of inorganic polymer materials is a useful approach [43]. Figure 6 shows the XRD pattern of HMP extracted by UAE-RMS and UAE-AS. The overall spectra shape is much similar to HMP obtained with UAE-RMS and UA-AS. A significant crystalline peak was visible in the HMP recovered using AS and RMS, with a 2 value between 28.40° and 40.50°. The peak of UAE-AS was higher than the peak of UAE-RMS. Furthermore, the intensity of a peak might show changes in the structure of the protein [24]. The findings revealed that the backbone structures of HMP are almost identical. The tiny peaks were observed at a 2θ value of about 50.30°, 58.70°, and 66.50°. When pure HMP failed to show



a



b



c

Figure 5. Fourier transform infrared (FTIR) spectra HMP extracted by UAE-RMS and UAE-AS (a); The curve-fitting amide I band spectra (1700–1600 cm-1) of hazelnut meal protein (HMP) obtained by ultrasound-assisted extraction (UAE)-reverse micellar solution (RMS) (b); UAE-alkaline solution (AS) (c).

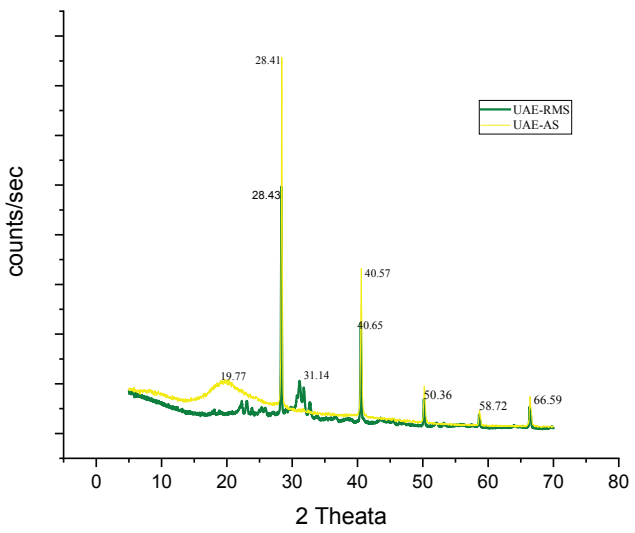


Figure 6. X-ray diffraction (XRD) spectra.

UAE-RMS: ultrasound-assisted extraction- reverse micellar solution
UAE-AS: ultrasound-assisted extraction-alkaline solution

any discernible intensity peaks in the XRD diffraction pattern spectrum, the presence of gold nanoparticles could be detected in the visible range of the spectrophotometer [43]. The results of this investigation are consistent with previous research [37, 44].

CONCLUSION

In the present work, the utilization of AOT-RMS which is a new technique to obtain HMP was successfully performed in this study. Throughout the UAE, RMS effectively damaged the hazelnut meal cells. Because the hydrogen bonds and electrostatic and hydrophobic interactions of the RMS aggregates facilitated the transport of HMP into RMS. In addition, since surfactant and organic solvent recycling were ecologically benign, protein is obtained by an environmentally friendly method. The optimization study was conducted due to determine the most effective extraction parameters. Firstly, one-factor analysis has been applied to set optimization limits. According to a one-factor analysis, BBD was applied. The protein isolate was produced at the optimum point for UAE-RMS. In order to test the effectiveness of RMS, the protein was produced at optimum points by using AS instead of RMS. In a comparative study, it was found that the UAE-RMS had greater extraction yields than the UAE-AS. The evaluation of the secondary structure of protein revealed that α-helix, β- turn, and β-sheet structures of HMP obtained by UAE-RMS were 14.74%, 31.66%, and 21.45%, respectively while the α-helix, β- turn, and β-sheet structure of HMP obtained by UAE-AS were 7.28%, 27.2%, and 17.3%, interpreted as changing the functional properties of the protein. UAE-RMS proved very

useful in obtaining HMP indicating a potential for scaling up the RMS and UAE approach into a commercial process. However, to ascertain the effect of UAE-RMS on the technologically useful features of HMP, more investigation is required.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] El-Adawy TA, Taha KM. Characteristics and composition of watermelon, pumpkin, and paprika seed oils and flours. J Agric Food Chem 2001;49:1253-1259. [CrossRef]

[2] Bener M, Şen FB, Önem AN, Bekdeşer B, Çelik SE, Lalikoglu M, et al. Microwave-assisted extraction of antioxidant compounds from by-products of Turkish hazelnut (Corylus avellana L.) using natural deep eutectic solvents: Modeling, optimization and phenolic characterization. Food Chem 2022;385:132633. [CrossRef]

[3] FAOSTAT. 2020: https://www.fao.org/faostat/en/#data/QCL/visualize.

[4] Sen D, Kahveci D. Production of a protein concentrate from Hazelnut meal obtained as a hazelnut oil industry by-product and its application in a functional beverage. Waste Biomass Valori 2020;11:5099-5107. [CrossRef]

[5] Tatar F, Tunç MT, Kahyaoglu T. Turkish Tombul hazelnut (Corylus avellana L.) protein concentrates: Functional and rheological properties. J Food Sci Technol 2015;52:1024-1031. [CrossRef]

[6] Aydemir LY, Gökbulut AA, Baran Y, Yemenicioğlu A. Bioactive, functional and edible film-forming properties of isolated hazelnut (Corylus avellana L.) meal proteins. Food Hydrocoll 2014;36:130-142. [CrossRef]

[7] Kumar K, Srivastav S, Sharanagat VS. Ultrasound assisted extraction (UAE) of bioactive compounds from fruit and vegetable processing by-products: A review. Ultrason Sonochem 2021;70:105325. [CrossRef]

[8] Rostagno MA, Palma M, Barroso CG. Ultrasound-assisted extraction of soy isoflavones. J Chromatogr A 2003;1012:119-128. [CrossRef]

[9] Zhu KX, Sun XH, Zhou HM. Optimization of ultrasound-assisted extraction of defatted wheat germ proteins by reverse micelles. J Cereal Sci 2009;50:266-271. [CrossRef]

[10] Lye GJ, Asenjo JA, Pyle DL. Protein extraction using reverse micelles: kinetics of protein partitioning. Chem Eng Sci 1994;49:3195-3204. [CrossRef]

[11] He S, Shi J, Walid E, Ma Y, Xue SJ. Extraction and purification of a lectin from small black kidney bean (Phaseolus vulgaris) using a reversed micellar system. Process Biochem 2013;48:746-752. [CrossRef]

[12] Liu F, Wang X, Zhao X, Hu H, Chen F, Sun Y. Surface properties of walnut protein from AOT reverse micelles. Int J Food Sci Technol 2014;49:626-633. [CrossRef]

[13] Li F, Raza A, Wang YW, Xu XQ, Chen GH. Optimization of surfactant-mediated, ultrasonic-assisted extraction of antioxidant polyphenols from rattan tea (Ampelopsis grossedentata) using response surface methodology. Pharmacogn Mag 2017;13:446-453. [CrossRef]

[14] Leser ME, Mrkoci K, Luisi PL, Reverse micelles in protein separation: the use of silica for the back-transfer process. Biotechnol Bioeng 1993;41:489-492. [CrossRef]

[15] Chuo SC, Mohd-Setapar SH, Mohamad-Aziz SN, Starov VM. A new method of extraction of amoxicillin using mixed reverse micelles. Colloids Surf A Physicochem Eng Asp 2014;460:137-144. [CrossRef]

[16] Liu L, Xi J. Mechanochemical-assisted extraction of protein from watermelon seeds with surfactant. LWT 2021;142:111025. [CrossRef]

[17] Naoe K, Noda K, Kawagoe M, Imai M. Higher order structure of proteins solubilized in AOT reverse micelles. Colloids Surf B Biointerfaces 2004;38:179-185. [CrossRef]

[18] Chen X, Ru Y, Chen F, Wang X, Zhao X, Ao Q. FTIR spectroscopic characterization of soy proteins obtained through AOT reverse micelles. Food Hydrocoll 2013;31:435-437. [CrossRef]

[19] Sun XH, Zhu KX, Zhou HM. Protein extraction from defatted wheat germ by reverse micelles: Optimization of the forward extraction. J Cereal Sci 2008;48:829-835. [CrossRef]

[20] Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 1976;72:248-254. [CrossRef]

[21] Wani AA, Sogi DS, Grover L, Saxena DC. Effect of temperature, alkali concentration, mixing time and meal/solvent ratio on the extraction of watermelon seed proteins - A response surface approach. Biosyst Eng 2006;94:67-73. [CrossRef]

[22] Chen J, Chen F, Wang X, Zhao X, Ao Q. The forward and backward transport processes in the AOT/hexane reversed micellar extraction of soybean protein. J Food Sci Technol 2014;51:2851-2856.

[23] Sun XH, Zhu KX, Zhou HM. Optimization of a novel backward extraction of defatted wheat germ protein from reverse micelles. Innov Food Sci Emerg Technol 2009;10:328-333. [CrossRef]

[24] Zhao X, Liu H, Zhang X, Zhu H, Ao Q. Surface structure and volatile characteristic of peanut proteins obtained through AOT reverse micelles. Colloids Surf B Biointerfaces 2019;173:860-868. [CrossRef]

[25] Lye GJ, Asenjo JA, Pyle DL. Extraction of lysozyme and ribonuclease-a using reverse micelles: Limits to protein solubilization. Biotechnol Bioeng 1995;47:509-519. [CrossRef]

[26] Wolbert RB, Hilhorst R, Voskuilen G, Nachtegaal H, Dekker M, Riet KVT, et al. Protein transfer from an aqueous phase into reversed micelles. Eur J Biochem 1989;184:627-633. [CrossRef]

[27] Cöklen KE, Hatton TA. Protein extraction using reverse micelles. Biotechnol Progress 1985;1:69-74. [CrossRef]

[28] Ichikawa S, Imai M, Shimizu M. Solubilizing water involved in protein extraction using reversed micelles. Biotechnol Bioeng 1992;39:20-26. [CrossRef]

[29] Geow CH, Tan MC, Yeap SP, Chin NL. Application of osmotic dehydration and ultrasound to enhance hazelnut oil extraction. Food Anal Methods 2021;14:411-421. [CrossRef]

[30] Wei F, Gao GZ, Wang XF, Dong XY, Li PP, Hua W, et al. Quantitative determination of oil content in small quantity of oilseed rape by ultrasound-assisted extraction combined with gas chromatography. Ultrason Sonochem 2008;15:938-942. [CrossRef]

[31] Xu Y, Zhang H, Xu X, Wang X. Numerical analysis and surrogate model optimization of air-cooled battery modules using double-layer heat spreading plates. Int J Heat Mass Transf 2021;176:121380. [CrossRef]

[32] Quanhong L. Caili F. Application of response surface methodology for extraction optimization of germinant pumpkin seeds protein. Food Chem 2005;92:701-706. [CrossRef]

[33] Tao Y, Zhang Z, Sun DW. Experimental and modeling studies of ultrasound-assisted release of phenolics from oak chips into model wine. Ultrason Sonochem 2014;21:1839-1848. [CrossRef]

[34] Guo Z, Chen F, Yang H, Liu K, Zhang L. Kinetics of protein extraction in reverse micelle. Int J Food Prop 2015;18:1707-1718. [CrossRef]

[35] Leser ME, Luisi PL, Paimieri S. The use of reverse micelles for the simultaneous extraction of oil and proteins from vegetable meal. Biotechnol Bioeng 1989;34:1140-1146. [CrossRef]

[36] Wang Z, Li H, Liang M, Yang L. Glutelin and prolamin, different components of rice protein, exert differently in vitro antioxidant activities. J Cereal Sci 2016;72:108-116. [CrossRef]

[37] Zhao X, Zhu H, Zhang B, Chen J, Ao Q, Wang X. XRD, SEM, and XPS analysis of soybean protein powders obtained through extraction involving reverse micelles. J Am Oil Chem Soc 2015;92:975-983. [CrossRef]

[38] Zhang Y, Yang R, Zhang W, Hu Z, Zhao W. Structural characterization and physicochemical properties of protein extracted from soybean meal assisted by steam flash-explosion with dilute acid soaking. Food Chem 2017;219:48-53. [CrossRef]

[39] Carbonaro M. Nucara A. Secondary structure of food proteins by Fourier transform spectroscopy in the mid-infrared region. Amino Acids 2010;38:679-690. [CrossRef]

[40] Baltacıoğlu H, Bayındırlı A, Severcan F. Secondary structure and conformational change of mushroom polyphenol oxidase during thermosonication treatment by using FTIR spectroscopy. Food Chem 2017;214:507-514. [CrossRef]

[41] Güler G, Vorob'ev MM, Vogel V, Mäntele W. Proteolytically-induced changes of secondary structural protein conformation of bovine serum albumin monitored by Fourier transform infrared (FT-IR) and UV-circular dichroism spectroscopy. Spectrochim Acta A Mol Biomol Spectrosc 2016;161:8-18. [CrossRef]

[42] Zhang L, Pan Z, Shen K, Cai X, Zheng B, Miao S. Influence of ultrasound-assisted alkali treatment on the structural properties and functionalities of rice protein. J Cereal Sci 2018;79:204-209. [CrossRef]

[43] Jayaramudu T, Raghavendra GM, Varaprasad K, Sadiku R, Raju KM. Development of novel biodegradable Au nanocomposite hydrogels based on wheat: For inactivation of bacteria. Carbohydr Polym 2013;92:2193-2200. [CrossRef]

[44] Zhao X, Chen F, Xue W, Lee L. FTIR spectra studies on the secondary structures of 7S and 11S globulins from soybean proteins using AOT reverse micellar extraction. Food Hydrocol 2008;22:568-575. [CrossRef]

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1214–1238, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00097

Research Article

Reliability estimation of a fault coverage distributed system with replacement options under four different scenarios

Jinbiao WU[1,*], Muhammad SALIHU ISA[1,2]

[1]*School of Mathematics and Statistics, Central South University Changsha, China*
[2]*Department of Mathematics, Yusuf Maitama Sule University, Kano, Nigeria*

ARTICLE INFO

Article history
Received: 23 January 2023
Revised: 04 April 2023
Accepted: 18 April 2023

Keywords:
Availability; Expected Profit; Fault Tolerant Factor; Reliability; Sensitivity

ABSTRACT

In this present study, series-parallel system composed of five subsystems with the following specifications were analyzed: subsystem 1 consists of two dissimilar clients that are connected to a single unit load balancer I which made up subsystem 2, whereas subsystem 3 consist of two active fog node working in parallel, subsystem 4 comprises of a load balancer II and subsystem 5 is made up of two similar units/components of cloud server. Cloud server, load balancer, fog node and clients failure and repair rate are assumed to be exponentially distributed. The system is under four different scenarios as follows: Scenario 1 system with replacement at complete, scenario 2 system with replacement at partial failure and complete, scenario 3 system without failure detection and replacement repair at complete and lastly, scenario 4 system with undetected failure and replacement at complete. This system is susceptible under first order differential difference equation to formulate the expression of availability and MTTF. The steady state availability, MTTF, sensitivity and expected profit based on general were compared and presented. This study is important to system engineers, designers, plant management, developers and maintenance personnel in the suitable designing and analysis of maintenance policy or processes and also in the assessment of performance and safety of the systems in general during and after the burn-in period.

Cite this article as: Wu J, Salihu Isa M. Reliability estimation of a fault coverage distributed system with replacement options under four different scenarios. Sigma J Eng Nat Sci 2024;42(4):1214–1238.

INTRODUCTION

In many scenarios, computer system utilizes number of distributed networks to provide available and optimal network to the clients. The study of computer network system present its economic and technical feasibility as the best choice for the multipurpose network. However, with the advancement in technology, availability of computer network happens to be subject of research and discussion. Meeting optimal level of availability is of paramount important in information, communication, military and institutional sector. Moreover, reliability could not attend its maximum level, computer network will be very poor.

***Corresponding author.**
*E-mail address: wujinbiao@csu.edu.cn

This paper was recommended for publication in revised form by Editor in-Chief Ahmet Selim Dalkilic

High computer system reliability is vital to industrial growth due to the fact that revenue mobilization is proportional to system performance. Due to its importance in industrial, domestic, institutional and manufacturing sector, literature study on dependability, reliability, maintainability and availability modelling of different computer network were developed. However, the developed models are used to address the computer network performance, subject to system failure. The technique of redundancy is thoroughly used to enhance reliability, dependability and availability of the system. In some computer network, the availability and dependability rely on the design of the system and strength of the units. To retain availability and dependability of complex computer network to an optimal level, the structure of the system and its components of optimal availability are required. Generally, system designers can develop technologies in a serial network to improve network availability, dependability and reliability. [1] Explore on performance analysis on computer network system that comprises of centralized database server, load balancer and distributed database server, [2] discuss on reliability metrics of network communication system having receiver, relay and transmitter. [3] Writes on reliability and dependability assessment of complex system having two subsystems on k-out-of-n working under G policy in subsystem 1 and four identical units in active parallel in subsystem 2. [4] Studied reliability of computer network base on genetic algorithm and the optimization technique were developed for better reliability. [5] Presents stochastic performance of computer based test having four subsystems arranged in series namely load balancer, clients, centralized server and database server. [6] Explore on reliability analysis of computer network which comprises of three subsystems: router, workstation and hub. [7] Investigate the performance measures of network with transparent bridge as follows 1-out-of-2: G, 2-out-of-3: F, a bridge unit and D 3-out-of-5: G schemes. [8] Dealt with an article on computer networking systems.

[9] Publish an article on reliability enhancement in intuitionistic fuzzy space, [10] discuss on heterogeneity using rpc in client. [11] Estimate the coliform values of the Tekkekoy deep sea discharge system, which is chosen as an application area, by using a radial-based artificial neural network structure, [12] writes on production-distribution network system for a company, which is active in producing bottled natural spring water was established.

[13] Develop models for the strength and performance analysis of computer network under different maintenance scenario. [14] Writes on reliability measures of database cluster, virtual router redundancy protocol and load balancer, the article analyzed availability by comparing the reliability if load balancer, virtual router redundancy protocol and high availability proxy were put in place, [15] investigated the impact of structure of the system reliability measures of software agent and client server. [16] Defined a Secure Simple Epidemic Algorithm (SSEA) for PSN where a security condition controls the traffic. [17] Classification algorithms were used to classify electromyography and depth sensor data, [18] optimum CW size is defined through meta-heuristic optimization algorithms. [19] Published on OLTP applications with incremental repartitioning of shared-nothing distributed databases, [20] investigates the implication of load balancing of distributed system. [21] Writes on analysis of the FANET TCAs currently in use, along with a brand-new taxonomy of TCAs based on the FANET topology architectures and underlying mathematical models.

[22] Explore on reliability analysis of computer network which comprises of three subsystems: router, workstation and hub. [23] Dealt with repairable system with reboot delay, one repair policy and imperfect coverage, [24] present work on the reliability measures of coverage factor with a standby system. [25] Investigate parallel system with three types of failure namely human failure, unit failure and major failure. [26] Consider a distributed system with five standby subsystems A (the clients), B (two load balancers), C (two distributed database servers), D (two mirrored distributed database serves) and E (centralized database server) is considered arranged as series-parallel system. [27] Analyzes the advantage of data center network topology by taking reliability and profit requirements into account, with distributed data center network topology having three components as follows: client applications, directory proxy server, and master servers were considered.

This research work further improved the work of previous researchers were five subsystems were considered. Subsystem A consist of 2-clients, subsystem B comprises of a load balancer I, 2-fog node are in subsystem C, subsystem D comprises of load balancer II and lastly, subsystem E consist of 2-cloud server. However, analysis of the model in terms of fault tolerant, general repair and copula were thoroughly investigated. Reliability analysis measures such as availability, MTTF, sensitivity, cost analysis was carried out for different scenarios to check optimality of the entire system with respect both failure and repair rate. Moreover, some practical applications were considered.

This work is structured as follows. Description, assumptions and nomenclatures on the system are presented in section 2, model formulation were discussed in section 3, section 4 consist of results and discussion and lastly section 5 which comprises of the conclusion.

According to the literature review, little research articles on performance estimation of a fault coverage distributed system with replacement options under four different scenarios have been published. Motivated by this fact, we are interested to conducting a research on performance estimation of a fault coverage distributed system with replacement options under four different scenarios in this present work. The impact of the fault tolerance factor, in conjunction with the different scenarios, on the system availability, MTTF, sensitivity and profit were captured. The primary goal of this work is to determine how different scenarios will

improve the availability and profit of the system under consideration, followed by a discussion and references, where the paper is concluded.

NOTATIONS, ASSUMPTIONS AND MODEL DESCRIPTION

Notations

v_0 failure rate of load balancer I
v_1 failure rate of clients
v_2 failure rate of fog nodes
v_3 failure rate of cloud servers
v_4 failure rate of load balancer II
ξ_0 Repair rate of load balancer
ξ_1 Replacement rate of clients
ξ_2 Replacement rate of fog nodes
ξ_3 Replacement rate cloud servers
ξ_4 Repair rate of load balancer II
c Fault tolerant (probability of withstanding fault)
$\delta_0 = 1 - c$
$\omega_i(t)$ Probability that a system is in a certain state at a given time.
A_{vk} At time t, the system is available

Assumptions

a. Failure of client is independent to the failure of fog node, load balancer and cloud server and vice vasa.
b. Repair / Replacement is immediate.
c. It is assumed that all the clients are active.
d. Each failure is repairable.
e. Rate of failure and repair obeys exponential distribution.
f. Systems have redundant standby units
g. Switching from standby to operation is perfect

Model Description

Subsystem A is made up of 2-clients in active parallel, subsystem B made up of load balancer I. 2-fog nodes in active parallel made up subsystem C, subsystem D made up of load balancer 2 and lastly, 2-cloud server in active parallel made up subsystem E. Moreover, the entire structure of the system, that is: Client, load balancer, fog node and cloud servers were configured as series-parallel, clients send request to the cloud server which in turn process the result and respond to the request. However, the two load balancers helps in utilization of the information required from the server, in Figure 1 (block diagram of the system), fog node serves as an intermediate between the clients, load balancers and cloud server. Table 1 provides a brief description of the states, while Figure 2 depicts all possible state transition for the model.

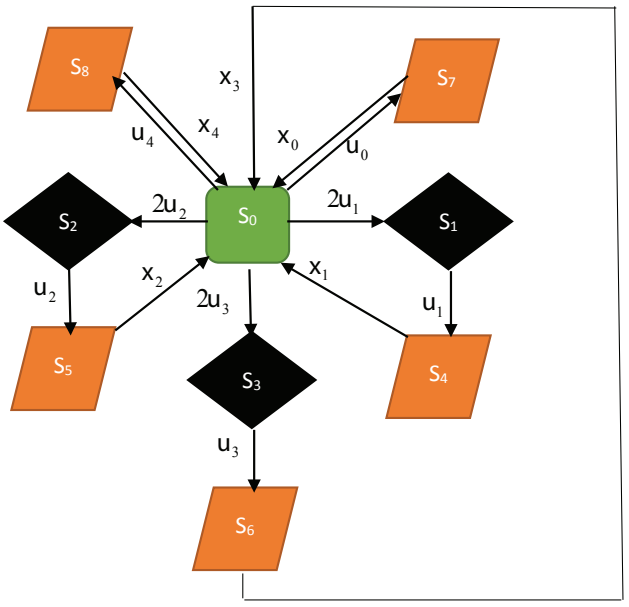


Figure 2. Transition diagram of scenario 1 system with replacement at complete.

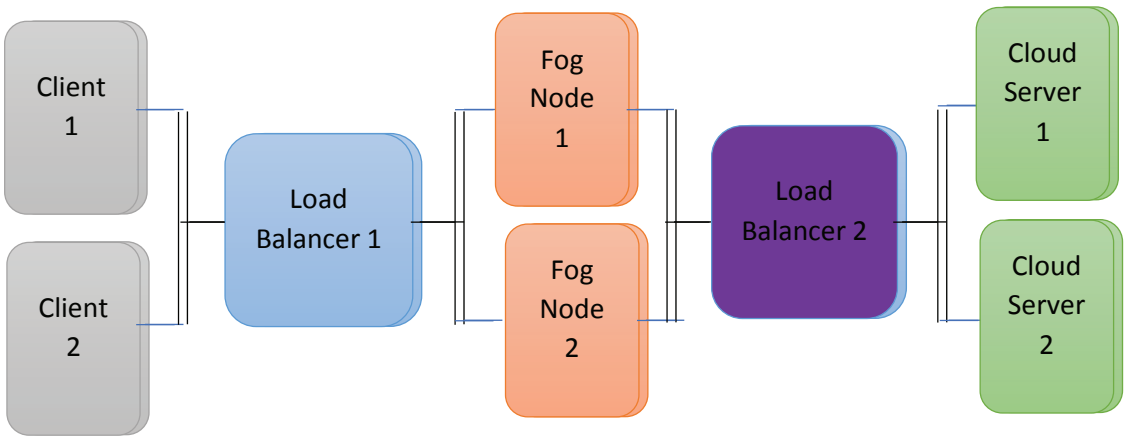


Figure 1. Performance block diagram of the system.

Table 1. State description of the system

State	Description	System Status
S_0	The clients, fog nodes, load balancer and cloud servers are working.	System is operational
S_1	Two fog nodes, One client failed, another client, load balancer and two cloud servers are working.	System is operational
S_2	One fog node, another fog node, two clients, load balancer and two cloud servers are working.	System is operational
S_3	One cloud server failed, another cloud server, two clients, two fog nodes are working	System is operational
S_4	One client failed, another client has failed.	System is down
S_5	One fog node failed, another fog node failed	System is down
S_6	One cloud server failed, another cloud server failed	System is down
S_7	Load balancer I failed	System is down
S_8	Load balancer II failed	System is down

Subsystem A: System with Replacement at Complete Failure State

Maintenance staff performs a perfect repair (repair as new) when a cloud server or client experiences a partial hardware failure. In the event of a complete failure over time, the component will be completely replaced. The system's Markov chain-based state transitions are shown in Figure 2 below.

Subsystem B: System with Replacement at Partial Failure and Complete State

The analysis is carried out as follows: in the event a system component fails due to hardware failure, maintenance personnel are charged with the responsibility of replacing the problematic part of the system to ensure that the system can still function. Figure 3 below shows the markov chain transition diagram

Subsystem C: System without Failure Detection and Replacement Repair at Complete State

The underlying premise is that whenever a fault manifests itself, whether at the cloud server or client side, the failure detection device were not in place to verify the failed component, as a result the failed component is therefore being replaced in order to avoid the failure occurring again anytime soon. Figure 4 below shows a diagram of a Markov chain transition.

Subsystem D: System with Detected Failure and Replacement at Complete State

In this subsystem, the units were considered fault tolerant in the sense that even when a fault occurs in one or more host components, they continue to operate without malfunctioning. Fault tolerance device is the property that allows a system to continue operating properly on the

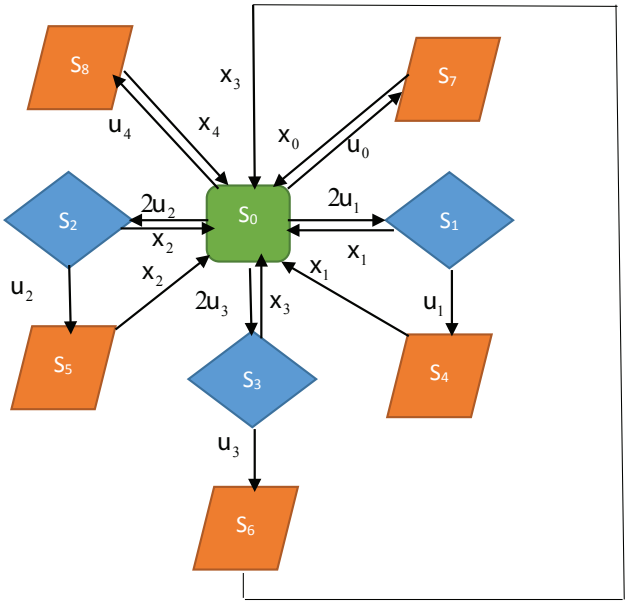


Figure 3. Transition diagram of scenario 2 system with replacement at partial failure and complete.

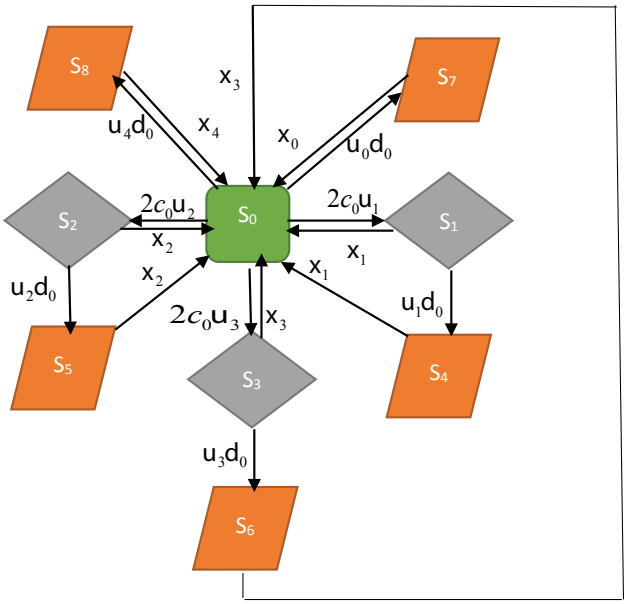


Figure 4. Transition diagram of scenario 3 system without failure detection and replacement repair at complete.

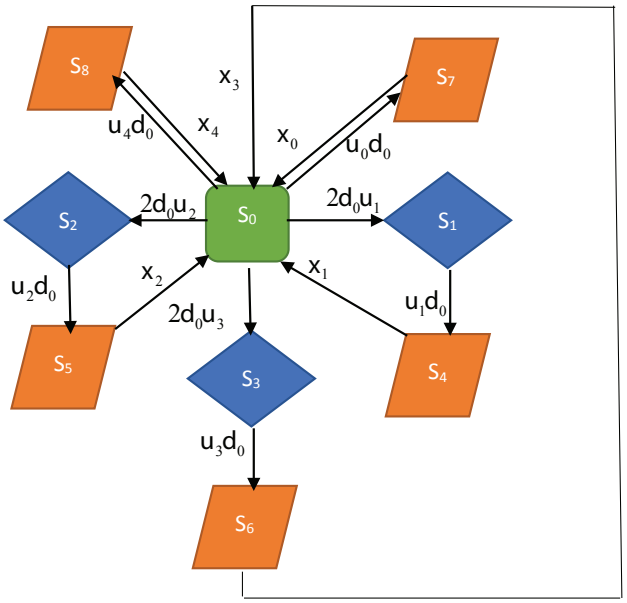


Figure 5. Transition diagram of scenario 4 system with detected failure and replacement at complete.

occurrence of a failure. The fault tolerant system, however, cannot withstand catastrophic failures, which results in system failure and requires replacement. The system's state transitions are shown in Figure 5 below using the Markov chain model.

MODEL FORMULATION

From Figure 1 above to derive the system of linear differential equation, the explicit expression of system availability can be obtained by solving the equations below. The results of the state probability equations for the system's operational states can then be used to determine the system availability. In order to analyse the system availability of the system, we define the $\omega_i(t)$ to be the probability that the system is in state i at time t and that we have $\omega(t)=\left[\omega_1(t),\omega_2(t),...,\omega_8(t)\right]$ be the probability row vector with initial conditions.

$$\omega_k(0)=\begin{cases}1, & k=0\\ 0, & k=1,2,3,4,5,6,7,8\end{cases} \tag{1}$$

The steady state probability of systems availability can be obtained from the solutions for $\omega_i(t), i=0,1,2,3,4,5,6,7,8$ State 0,1,2 and 3 are the only working states of all the scenarios in Figure 1, thus the steady state availability $A_{vi}(\infty)$ at time bility that the system is in state i at time t and that we have is

$$A_{Ti}(\infty)=\omega_0(\infty)+\omega_1(\infty)+\omega_2(\infty)+\omega_3(\infty) \tag{2}$$

From Figure 2, the corresponding set of differential difference equations for Subsystem 1 are

$$\begin{pmatrix}\omega_0'(t)\\ \omega_1'(t)\\ \omega_2'(t)\\ \omega_3'(t)\\ \omega_4'(t)\\ \omega_5'(t)\\ \omega_6'(t)\\ \omega_7'(t)\\ \omega_8'(t)\end{pmatrix}=\begin{pmatrix}-\eta_1 & 0 & 0 & 0 & \xi_1 & \xi_2 & \xi_3 & \xi_0 & \xi_4\\ 2\nu_1 & -\nu_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 2\nu_2 & 0 & -\nu_2 & 0 & 0 & 0 & 0 & 0 & 0\\ 2\nu_3 & 0 & 0 & -\nu_3 & 0 & 0 & 0 & 0 & 0\\ 0 & \nu_1 & 0 & 0 & -\xi_1 & 0 & 0 & 0 & 0\\ 0 & 0 & \nu_2 & 0 & 0 & -\xi_2 & 0 & 0 & 0\\ 0 & 0 & 0 & \nu_3 & 0 & 0 & -\xi_3 & 0 & 0\\ \nu_0 & 0 & 0 & 0 & 0 & 0 & 0 & -\xi_0 & 0\\ \nu_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\xi_4\end{pmatrix}\begin{pmatrix}\omega_0(t)\\ \omega_1(t)\\ \omega_2(t)\\ \omega_3(t)\\ \omega_4(t)\\ \omega_5(t)\\ \omega_6(t)\\ \omega_7(t)\\ \omega_8(t)\end{pmatrix}$$

Availability of Subsystem 1 is

$$A_{T1}(\infty)=\omega_0(\infty)+\omega_1(\infty)+\omega_2(\infty)+\omega_3(\infty) \tag{3}$$

Setting $\omega_k'(t)=0$ as $t\to\infty$ in steady state, to obtained

$$\begin{pmatrix}-\eta_1 & 0 & 0 & 0 & \xi_1 & \xi_2 & \xi_3 & \xi_0 & \xi_4\\ 2\nu_1 & -\nu_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 2\nu_2 & 0 & -\nu_2 & 0 & 0 & 0 & 0 & 0 & 0\\ 2\nu_3 & 0 & 0 & -\nu_3 & 0 & 0 & 0 & 0 & 0\\ 0 & \nu_1 & 0 & 0 & -\xi_1 & 0 & 0 & 0 & 0\\ 0 & 0 & \nu_2 & 0 & 0 & -\xi_2 & 0 & 0 & 0\\ 0 & 0 & 0 & \nu_3 & 0 & 0 & -\xi_3 & 0 & 0\\ \nu_0 & 0 & 0 & 0 & 0 & 0 & 0 & -\xi_0 & 0\\ \nu_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\xi_4\end{pmatrix}\begin{pmatrix}\omega_0(t)\\ \omega_1(t)\\ \omega_2(t)\\ \omega_3(t)\\ \omega_4(t)\\ \omega_5(t)\\ \omega_6(t)\\ \omega_7(t)\\ \omega_8(t)\end{pmatrix}=\begin{pmatrix}0\\0\\0\\0\\0\\0\\0\\0\\1\end{pmatrix}$$

the normalizing condition is

$$\sum_{k=0}^{8}\omega_k(\infty)=1 \tag{4}$$

Using (4) to give the explicit expressions for the steady-state availability of Subsystem 1 given in (3) is now

$$A_{T1}(\infty)=\frac{7\xi_0\xi_1\xi_2\xi_3\xi_4}{Q_1} \tag{5}$$

where

$Q_1=7\xi_0\xi_1\xi_2\xi_3\xi_4+\nu_4\xi_0\xi_1\xi_2\xi_3+2\nu_3\xi_0\xi_1\xi_2\xi_4+2\nu_3\xi_0\xi_1\xi_2\xi_4$
$+2\nu_2\xi_0\xi_1\xi_3\xi_4+2\nu_1\xi_0\xi_2\xi_3\xi_4+\nu_0\xi_1\xi_2\xi_3\xi_4$ and

$\eta_1=(\nu_0+2\nu_1+2\nu_2+2\nu_3+\nu_4)$.

To evaluate the $MTTF_1$, the rows and columns of the absorbing (failure) states from the above matrix were deleted and transposed to obtain the new matrix L_1.

$$L_1=\begin{pmatrix}-\eta_1 & 2\nu_1 & 2\nu_2 & 2\nu_3\\ 0 & -\nu_1 & 0 & 0\\ 0 & 0 & -\nu_2 & 0\\ 0 & 0 & 0 & -\nu_3\end{pmatrix}$$

$$MTTF_1=\frac{7}{\nu_0+2\nu_1+2\nu_2+2\nu_3+\nu_4} \tag{6}$$

$$PF_1=AT_1C_0-Cr_1Br_1 \tag{7}$$

From Figure 3, the corresponding set of differential difference equations for Subsystem 2 are

$$\begin{pmatrix} \omega_0'(t) \\ \omega_1'(t) \\ \omega_2'(t) \\ \omega_3'(t) \\ \omega_4'(t) \\ \omega_5'(t) \\ \omega_6'(t) \\ \omega_7'(t) \\ \omega_8'(t) \end{pmatrix} = \begin{pmatrix} -\eta_1 & \xi_1 & \xi_2 & \xi_3 & \xi_1 & \xi_2 & \xi_3 & \xi_0 & \xi_4 \\ 2\nu_1 & -\eta_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2\nu_2 & 0 & -\eta_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2\nu_3 & 0 & 0 & -\eta_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & \nu_1 & 0 & 0 & -\xi_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \nu_2 & 0 & 0 & -\xi_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \nu_3 & 0 & 0 & -\xi_3 & 0 & 0 \\ \nu_0 & 0 & 0 & 0 & 0 & 0 & 0 & -\xi_0 & 0 \\ \nu_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\xi_4 \end{pmatrix} \begin{pmatrix} \omega_0(t) \\ \omega_1(t) \\ \omega_2(t) \\ \omega_3(t) \\ \omega_4(t) \\ \omega_5(t) \\ \omega_6(t) \\ \omega_7(t) \\ \omega_8(t) \end{pmatrix}$$

Using the same argument above, availability expression of Subsystem 2 is

$$A_{T2}(\infty) = \omega_0(\infty) + \omega_1(\infty) + \omega_2(\infty) + \omega_3(\infty) \tag{8}$$

Setting $\omega_k'(t) = 0$ as $t \to \infty$ in steady state, to obtained

$$\begin{pmatrix} -\eta_1 & \xi_1 & \xi_2 & \xi_3 & \xi_1 & \xi_2 & \xi_3 & \xi_0 & \xi_4 \\ 2\nu_1 & -\eta_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2\nu_2 & 0 & -\eta_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2\nu_3 & 0 & 0 & -\eta_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & \nu_1 & 0 & 0 & -\xi_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \nu_2 & 0 & 0 & -\xi_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \nu_3 & 0 & 0 & -\xi_3 & 0 & 0 \\ \nu_0 & 0 & 0 & 0 & 0 & 0 & 0 & -\xi_0 & 0 \\ \nu_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\xi_4 \end{pmatrix} \begin{pmatrix} \omega_0(t) \\ \omega_1(t) \\ \omega_2(t) \\ \omega_3(t) \\ \omega_4(t) \\ \omega_5(t) \\ \omega_6(t) \\ \omega_7(t) \\ \omega_8(t) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

the normalizing condition is

$$\sum_{k=0}^{8} \omega_k(\infty) = 1 \tag{9}$$

Using (8) to give the explicit expressions for the steady-state availability of Subsystem 2 given in (8) is now

$$A_{T2}(\infty) = \frac{2(\xi_0\xi_1\xi_2\xi_3\xi_4\nu_2 + \xi_0\xi_2\xi_3\xi_4\nu_1\nu_2)}{Q_2} \tag{10}$$

Where $Q_2 = 2\xi_0\xi_2^2\xi_3\nu_1^2 + \xi_0\xi_4\xi_2^2\xi_3 + 2\nu_2\xi_2\xi_3\xi_4 + 2\nu_3\xi_0\xi_1\xi_2\xi_4 + 2\nu_2^2\xi_0\xi_1\xi_3\xi_4 + \nu_4\xi_0\xi_1\xi_2^2\xi_3 + 2\nu_3\xi_0\xi_1\xi_2^2\xi_4$

and $\eta_1 = (\nu_0 + 2\nu_1 + 2\nu_2 + 2\nu_3 + \nu_4), \eta_2 = (\xi_1 + \nu_1), \eta_3 = (\xi_2 + \nu_2), \eta_4 = (\xi_3 + \nu_3)$

To evaluate the $MTTF_2$, the rows and columns of the absorbing (failure) states from the above matrix were deleted and transposed to obtain the new matrix L_2.

$$L_2 = \begin{pmatrix} -\eta_1 & 2\nu_1 & 2\nu_2 & 2\nu_3 \\ 0 & -\eta_2 & 0 & 0 \\ 0 & 0 & -\eta_3 & 0 \\ 0 & 0 & 0 & -\eta_4 \end{pmatrix}$$

$$MTTF_2 = \frac{2(\nu_3\xi_1\xi_2 + \nu_3\nu_2\xi_1 + \nu_2\nu_3\xi_2 + \nu_1\nu_2\nu_3)}{2(\nu_2\nu_3^2\xi_1 + 2\nu_3\nu_2^2\xi_1 + 2\nu_1\nu_3^2\xi_2 + 2\nu_1\nu_2^2\xi_3 + 2\nu_2\nu_1^2\xi_3)} \tag{11}$$

$$PF_2 = AT_2C_0 - Cr_2Br_2 - Cr_1Br_{22} \tag{12}$$

From Figure 4, the corresponding set of differential difference equations for Subsystem 3 are

$$\begin{pmatrix} \omega_0'(t) \\ \omega_1'(t) \\ \omega_2'(t) \\ \omega_3'(t) \\ \omega_4'(t) \\ \omega_5'(t) \\ \omega_6'(t) \\ \omega_7'(t) \\ \omega_8'(t) \end{pmatrix} = \begin{pmatrix} -\eta_5 & \xi_1 & \xi_2 & \xi_3 & \xi_1 & \xi_2 & \xi_3 & \xi_0 & \xi_4 \\ 2c_0\nu_1 & -\eta_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2c_0\nu_2 & 0 & -\eta_7 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2c_0\nu_3 & 0 & 0 & -\eta_8 & 0 & 0 & 0 & 0 & 0 \\ 0 & \delta_0\nu_1 & 0 & 0 & -\xi_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \delta_0\nu_2 & 0 & 0 & -\xi_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \delta_0\nu_3 & 0 & 0 & -\xi_3 & 0 & 0 \\ \delta_0\nu_0 & 0 & 0 & 0 & 0 & 0 & 0 & -\xi_0 & 0 \\ \delta_0\nu_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\xi_4 \end{pmatrix} \begin{pmatrix} \omega_0(t) \\ \omega_1(t) \\ \omega_2(t) \\ \omega_3(t) \\ \omega_4(t) \\ \omega_5(t) \\ \omega_6(t) \\ \omega_7(t) \\ \omega_8(t) \end{pmatrix}$$

Using the same argument above, availability expression of Subsystem 3 is

$$A_{T3}(\infty) = \omega_0(\infty) + \omega_1(\infty) + \omega_2(\infty) + \omega_3(\infty) \tag{13}$$

Setting $\omega_k'(t) = 0$ as $t \to \infty$ in steady state, to obtained

$$\begin{pmatrix} -\eta_5 & \xi_1 & \xi_2 & \xi_3 & \xi_1 & \xi_2 & \xi_3 & \xi_0 & \xi_4 \\ 2c_0\nu_1 & -\eta_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2c_0\nu_2 & 0 & -\eta_7 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2c_0\nu_3 & 0 & 0 & -\eta_8 & 0 & 0 & 0 & 0 & 0 \\ 0 & \delta_0\nu_1 & 0 & 0 & -\xi_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \delta_0\nu_2 & 0 & 0 & -\xi_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \delta_0\nu_3 & 0 & 0 & -\xi_3 & 0 & 0 \\ \delta_0\nu_0 & 0 & 0 & 0 & 0 & 0 & 0 & -\xi_0 & 0 \\ \delta_0\nu_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\xi_4 \end{pmatrix} \begin{pmatrix} \omega_0(t) \\ \omega_1(t) \\ \omega_2(t) \\ \omega_3(t) \\ \omega_4(t) \\ \omega_5(t) \\ \omega_6(t) \\ \omega_7(t) \\ \omega_8(t) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

the normalizing condition is

$$\sum_{k=0}^{8} \omega_k(\infty) = 1 \tag{14}$$

Using (14) to give the explicit expressions for the steady-state availability of Subsystem 3 given in (14) is now

$$A_{T3}(\infty) = \frac{2(c_0\xi_0\xi_1\xi_2^2\xi_3\xi_4\nu_1 + c_0\xi_0\xi_2\xi_3\xi_4\nu_1\nu_2)}{Q_3} \tag{15}$$

Where $Q_3 = \xi_0\xi_1^2\xi_2^2\xi_3\xi_4 + \nu_2\xi_0\xi_1^2\xi_2\xi_3\xi_4 + 2c_0\nu_2^2\xi_0\xi_1^2\xi_2\xi_3 + 2c_0\nu_3\xi_0\xi_1\xi_2^2\xi_4 + \delta_0\nu_4\xi_0\xi_1^2\xi_2^2\xi_3 + \delta_0\nu_0\xi_4\xi_1^2\xi_2^2\xi_3$

and $\eta_5 = (\delta_0\nu_0 + 2c_0\nu_1 + 2c_0\nu_2 + 2c_0\nu_3 + \delta_0\nu_4), \eta_6 = (\xi_1 + \delta_0\nu_1), \eta_7 = (\xi_2 + \delta_0\nu_2), \eta_8 = (\xi_3 + \delta_0\nu_3)$

To evaluate the $MTTF_3$, the rows and columns of the absorbing (failure) states from the above matrix were deleted and transposed to obtain the new matrix L_3.

$$L_3 = \begin{pmatrix} -\eta_5 & 2c_0\nu_1 & 2c_0\nu_2 & 2c_0\nu_3 \\ \xi_1 & -\eta_6 & 0 & 0 \\ \xi_2 & 0 & -\eta_7 & 0 \\ \xi_3 & 0 & 0 & -\eta_8 \end{pmatrix}$$

$$MTTF_3 = \frac{2c_0(v_3\xi_1\xi_2 + \delta_0 v_3 v_2\xi_1 + \delta_0 v_1 v_3\xi_2 + 2\delta_0 v_1 v_2 v_3)}{2c_0\delta_0^2 v_2 v_3^2\xi_1 + 2c_0\delta_0^2 v_3 v_2^2\xi_1 + 2c_0\delta_0^2 v_1 v_3^2 + 2v_1 v_2 v_3^2 + 2v_1 v_2^2 v_3 + 2v_2 v_1^2 v_3} \quad (16)$$

$$PF_3 = AT_2C_0 - Cr_2Br_3 - Cr_1Br_{33} \quad (17)$$

From Figure 5, the corresponding set of differential difference equations for Subsystem 4 are

$$\begin{pmatrix} \omega_0'(t) \\ \omega_1'(t) \\ \omega_2'(t) \\ \omega_3'(t) \\ \omega_4'(t) \\ \omega_5'(t) \\ \omega_6'(t) \\ \omega_7'(t) \\ \omega_8'(t) \end{pmatrix} = \begin{pmatrix} -\eta_9 & 0 & 0 & 0 & \xi_1 & \xi_2 & \xi_3 & \xi_0 & \xi_4 \\ 2\delta_0 v_1 & -v_1\delta_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2\delta_0 v_2 & 0 & -v_2\delta_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2\delta_0 v_3 & 0 & 0 & -v_3\delta_0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \delta_0 v_1 & 0 & 0 & -\xi_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \delta_0 v_2 & 0 & 0 & -\xi_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \delta_0 v_3 & 0 & 0 & -\xi_3 & 0 & 0 \\ \delta_0 v_0 & 0 & 0 & 0 & 0 & 0 & 0 & -\xi_0 & 0 \\ \delta_0 v_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\xi_4 \end{pmatrix} \begin{pmatrix} \omega_0(t) \\ \omega_1(t) \\ \omega_2(t) \\ \omega_3(t) \\ \omega_4(t) \\ \omega_5(t) \\ \omega_6(t) \\ \omega_7(t) \\ \omega_8(t) \end{pmatrix}$$

Using the same argument above, availability expression of Subsystem 4 is

$$A_{T3}(\infty) = \omega_0(\infty) + \omega_1(\infty) + \omega_2(\infty) + \omega_3(\infty) \quad (18)$$

Setting $\omega_k'(t) = 0$ as $t \to \infty$ in steady state, to obtained

$$\begin{pmatrix} -\eta_9 & 0 & 0 & 0 & \xi_1 & \xi_2 & \xi_3 & \xi_0 & \xi_4 \\ 2\delta_0 v_1 & -v_1\delta_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2\delta_0 v_2 & 0 & -v_2\delta_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2\delta_0 v_3 & 0 & 0 & -v_3\delta_0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \delta_0 v_1 & 0 & 0 & -\xi_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \delta_0 v_2 & 0 & 0 & -\xi_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \delta_0 v_3 & 0 & 0 & -\xi_3 & 0 & 0 \\ \delta_0 v_0 & 0 & 0 & 0 & 0 & 0 & 0 & -\xi_0 & 0 \\ \delta_0 v_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\xi_4 \end{pmatrix} \begin{pmatrix} \omega_0(t) \\ \omega_1(t) \\ \omega_2(t) \\ \omega_3(t) \\ \omega_4(t) \\ \omega_5(t) \\ \omega_6(t) \\ \omega_7(t) \\ \omega_8(t) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

the normalizing condition is

$$\sum_{k=0}^{8} \omega_k(\infty) = 1 \quad (19)$$

Using (19) to give the explicit expressions for the steady-state availability of Subsystem 4 given in (19) is now

$$A_{T4}(\infty) = \frac{7\xi_0\xi_1\xi_2\xi_3\xi_4}{Q_4} \quad (20)$$

Where $Q_4 = 7\xi_0\xi_1\xi_2\xi_3\xi_4 + \delta_0 v_4\xi_0\xi_1\xi_2\xi_3 + 2\delta_0 v_3\xi_0\xi_1\xi_2\xi_4$
$2\delta_0 v_2\xi_0\xi_1\xi_2\xi_4 + 2\delta_0 v_1\xi_0\xi_2\xi_3\xi_4 + \delta_0 v_0\xi_1\xi_2\xi_3\xi_4$
and, $\eta_9 = (\delta_0 v_0 + 2\delta_0 v_1 + 2\delta_0 v_2 + 2\delta_0 v_3 + \delta_0 v_0 + \delta_0 v_4)$

To evaluate the $MTTF_4$, the rows and columns of the absorbing (failure) states from the above matrix were deleted and transposed to obtain the new matrix L_4.

$$L_4 = \begin{pmatrix} -\eta_9 & 2\delta_0 v_1 & 2\delta_0 v_2 & 2\delta_0 v_3 \\ 0 & -v_1\delta_0 & 0 & 0 \\ 0 & 0 & -v_2\delta_0 & 0 \\ 0 & 0 & 0 & -v_3\delta_0 \end{pmatrix}$$

$$MTTF_4 = \frac{7}{\delta_0(2v_0 + 2v_1 + 2v_2 + 2v_3 + v_4)}$$

$$PF_4 = AT_4C_0 - Cr_1Br_4 \quad (22)$$

RESULTS AND DISCUSSIONS

The objective of this section is to express numerical experiment so as to see effect of the parameters on the performance by the use of MATLAB software. The findings of availability, MTTF and profit for all the four (4) scenarios in terms of failure rates v_o, v_1, v_2, v_3, and v_4 with repair rate ξ_o, ξ_1, ξ_2, ξ_3, and ξ_4 as follows: Tables 2, 3 and Figure [6 – 21] visually explain the detailed analysis of the availability,

Table 2. Variation of availability, MTTF and profit with respect to failure rate v_0 for the four scenarios

	Availability				Mean time to Failure (MTTF)				Profit*10^7			
	Scenario				Scenario				Scenario			
v_0	1	2	3	4	1	2	3	4	1	2	3	4
0.0	0.979	0.963	0.988	0.993	152.1	54.07	172.2	507.2	2.449	2.409	2.470	2.484
0.1	0.794	0.392	0.670	0.928	47.94	9.151	28.88	94.85	1.986	0.980	1.675	2.320
0.2	0.668	0.246	0.507	0.870	28.45	4.999	15.76	52.31	1.670	0.615	1.267	2.175
0.3	0.576	0.179	0.407	0.819	20.23	3.438	10.84	36.11	1.441	0.448	1.019	2.048
0.4	0.506	0.141	0.340	0.774	15.69	2.620	8.260	27.58	1.267	0.352	0.852	1.935
0.5	0.452	0.116	0.292	0.733	12.82	2.117	6.672	22.30	1.130	0.290	0.732	1.833
0.6	0.408	0.098	0.256	0.697	10.83	1.775	5.596	18.72	1.020	0.247	0.642	1.742
0.7	0.372	0.086	0.228	0.663	9.383	1.529	4.819	16.13	0.930	0.215	0.571	1.659
0.8	0.341	0.076	0.206	0.633	8.274	1.342	4.231	14.17	0.854	0.190	0.514	1.584
0.9	0.316	0.068	0.187	0.606	7.399	1.197	3.771	12.63	0.790	0.170	0.468	1.516

Table 3. Variation of availability, MTTF and profit with respect to repair rate x_0 for the four scenarios

	Availability				**Mean Time To Failure (MTTF)**				**Profit*10^7**			
	Scenario				**Scenario**				**Scenario**			
x_0	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
0.0	0.000	0.000	0.000	0.000	56.00	20.62	60.83	185.1	0.000	0.000	0.000	0.000
0.1	0.934	0.892	0.962	0.979	56.00	20.62	60.83	185.1	2.337	2.232	2.406	2.448
0.2	0.935	0.895	0.963	0.979	56.00	20.62	60.83	185.1	2.338	2.239	2.409	2.449
0.3	0.935	0.896	0.964	0.979	56.00	20.62	60.83	185.1	2.339	2.242	2.410	2.449
0.4	0.935	0.897	0.964	0.979	56.00	20.62	60.83	185.1	2.339	2.243	2.411	2.449
0.5	0.935	0.897	0.964	0.979	56.00	20.62	60.83	185.1	2.339	2.244	2.411	2.449
0.6	0.936	0.897	0.964	0.979	56.00	20.62	60.83	185.1	2.339	2.244	2.411	2.449
0.7	0.936	0.898	0.964	0.979	56.00	20.62	60.83	185.1	2.340	2.245	2.411	2.449
0.8	0.936	0.898	0.964	0.979	56.00	20.62	60.83	185.1	2.340	2.245	2.412	2.449
0.9	0.936	0.898	0.964	0.979	56.00	20.62	60.83	185.1	2.340	2.245	2.412	2.449

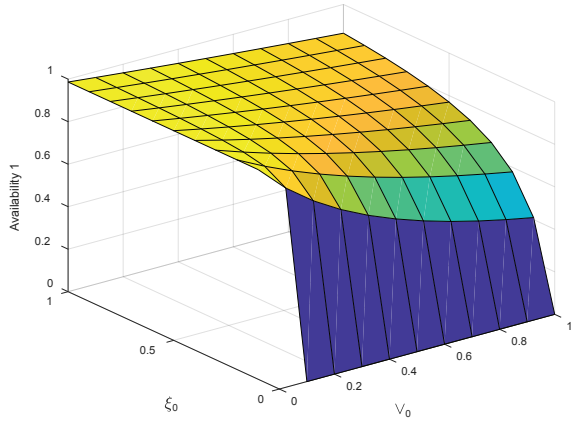


Figure 6. Availability against scenario 1 for v_0 and x_0.

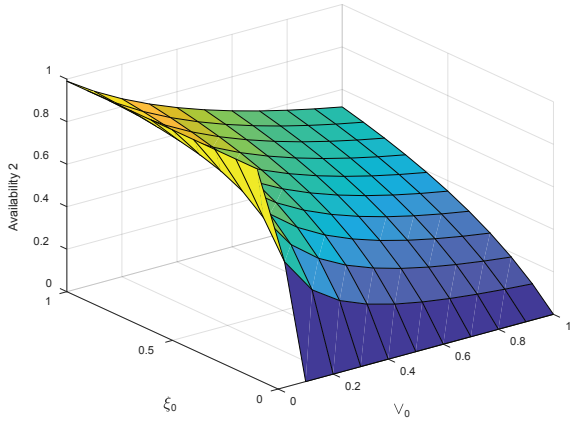


Figure 7. Availability against scenario 2 for v_0 and x_0.

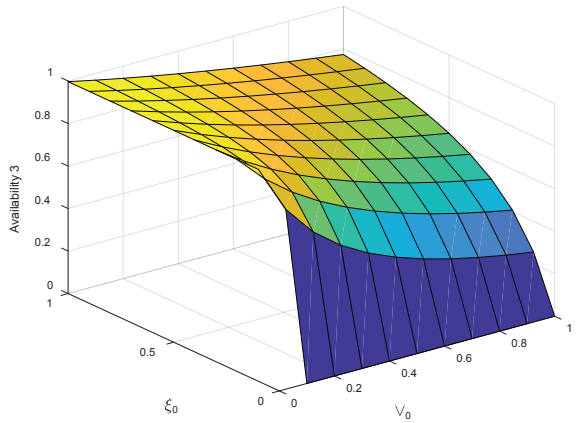


Figure 8. Availability against scenario 3 for v_0 and x_0.

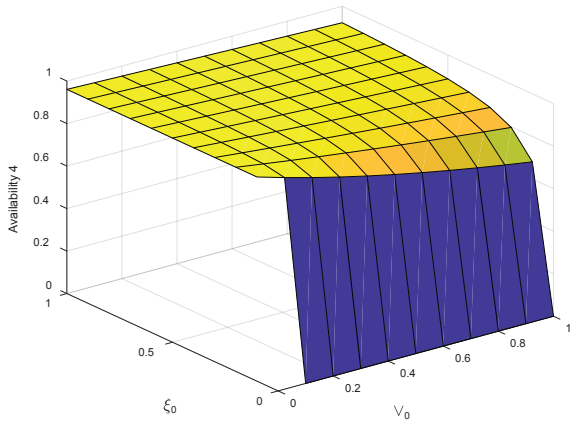


Figure 9. Availability against scenario 4 for v_0 and x_0.

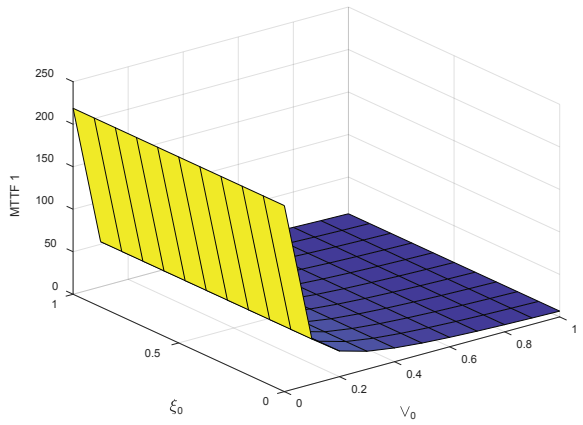


Figure 10. MTTF against scenario 1 for v_0 and x_0.

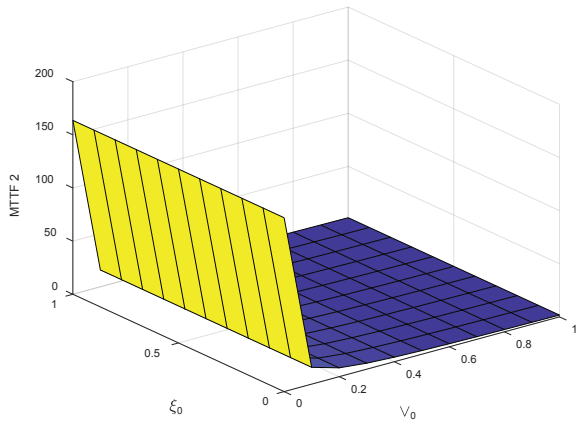


Figure 11. MTTF against scenario 2 for v_0 and x_0.

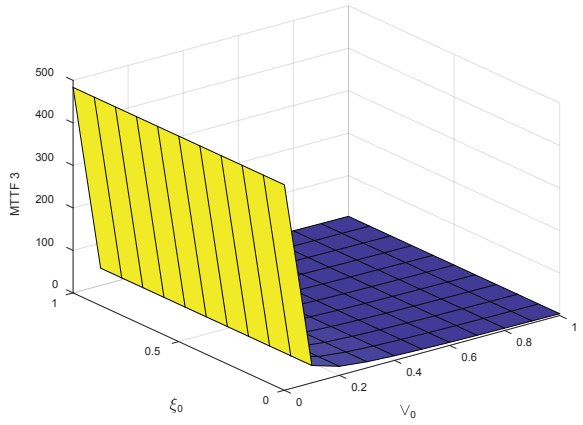


Figure 13. MTTF against scenario 3 for v_0 and x_0.

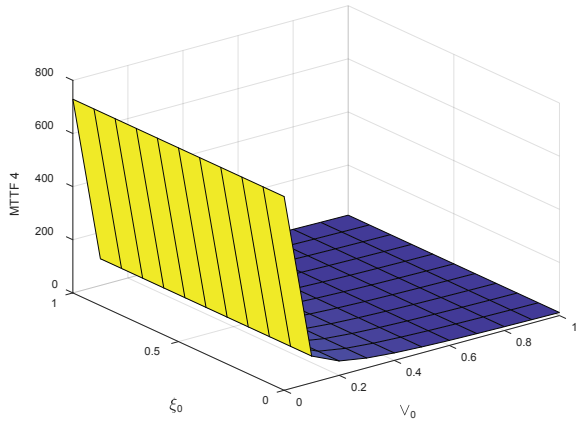


Figure 14. MTTF against scenario 4 for v_0 and x_0.

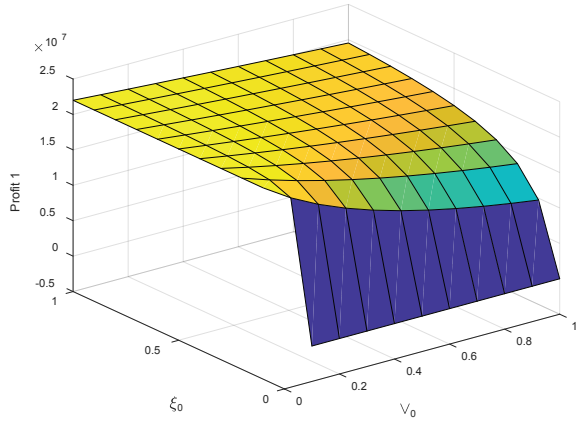


Figure 15. Profit against scenario 1 for v_0 and x_0.

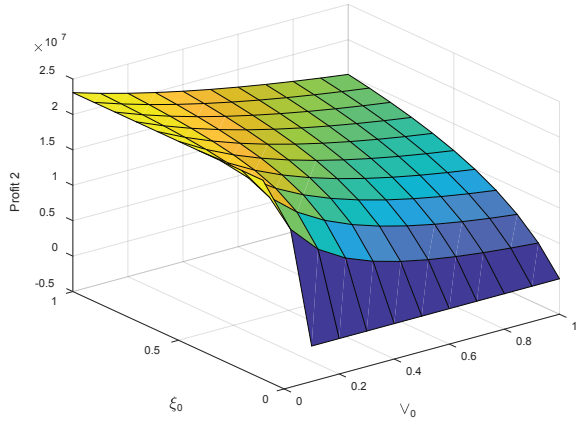


Figure 16. Profit against scenario 2 for v_0 and x_0.

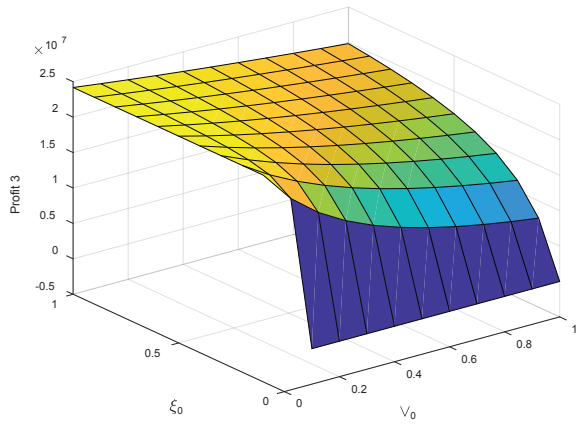


Figure 17. Profit against scenario 3 for v_0 and x_0.

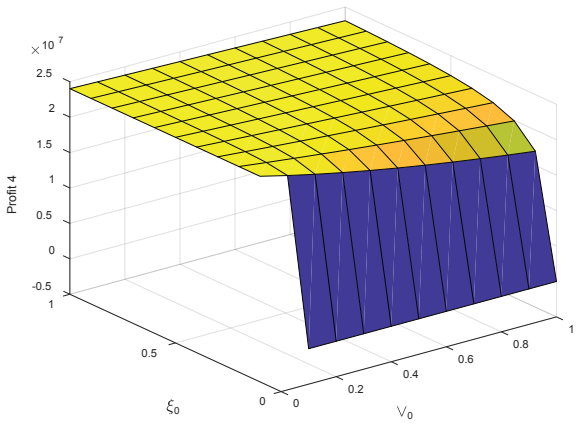


Figure 18. Profit against scenario 4 for v_0 and x_0.

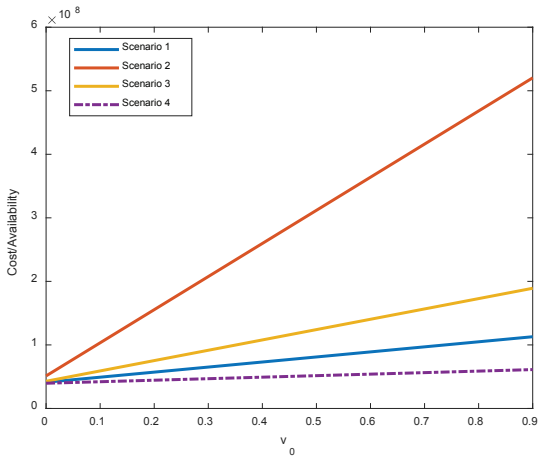
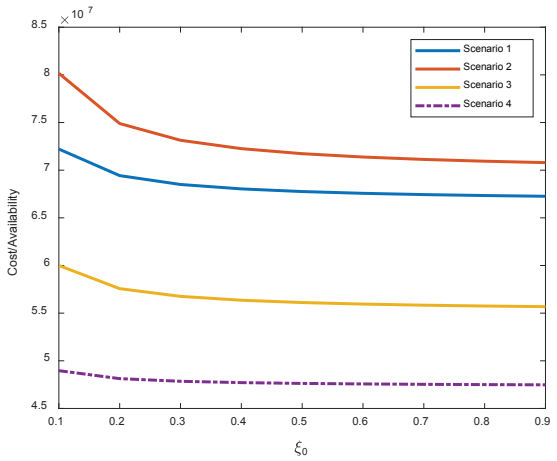


Figure 19. Cost/availability versus failure rate.



20. Cost/availability versus repair rate.

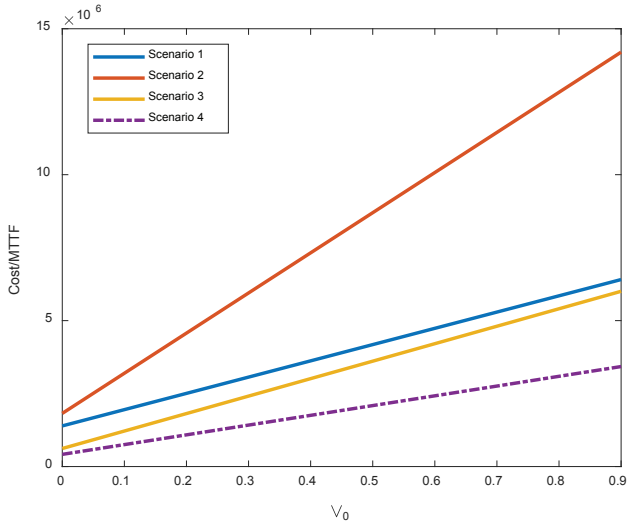


Figure 21. Cost/MTTF versus failure rate.

MTTF, profit and cost benefit respectively in terms of v_o and ξ_o. On the other hand, additional figures show an increasing pattern, highlighting the system's robustness in reaction to variations in failure and repair rates v_1 and ξ_1 as shown in Tables 4, 5 and Figure [22 – 25] in terms of availability, Figure [26 – 29] for MTTF, Figure [30 – 33] in terms of profit and Figure [34 – 36] in terms cost benefit. Tables 6, 7 and Figure [37 -51] are relevant to availability, MTTF, profit and cost benefit in terms of v_2 and ξ_2. The graphical representations encapsulated in Tables 8, 9 and figure [52 – 66] serve as a visual exploration of the intricate dynamics between failure and repair rates v_3 and ξ_3 and their consequential impact on availability, MTTF, profit and cost benefit for four different scenarios. However, availability analysis, MTTF, profit analysis and cost benefit was carried out to the same scenarios in Tables 10, 11 and figure [67 -81] it was observed that availability increases with increase in all repair rates and decreases as the failure rate increases

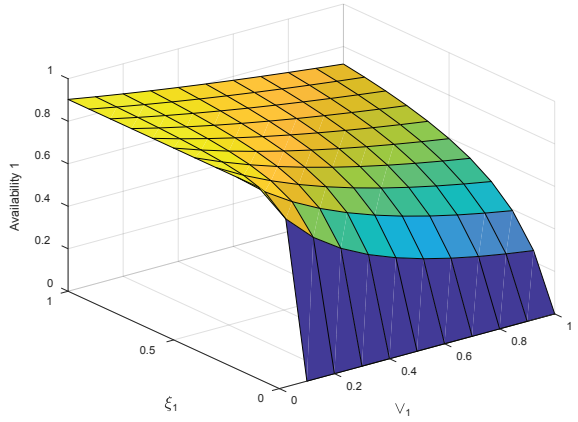


Figure 22. Availability against scenario 1 for v_1 and x_1.

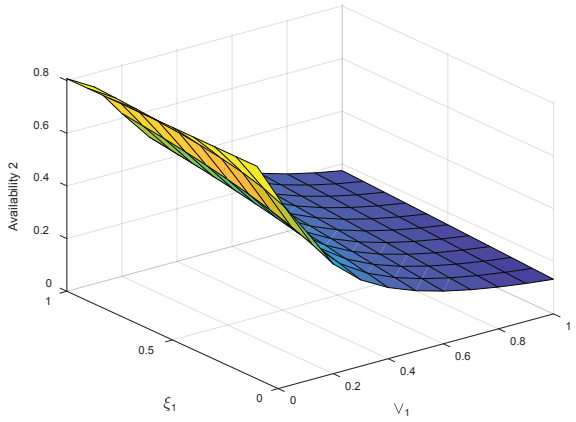


Figure 23. Availability against scenario 2 for v_1 and x_1.

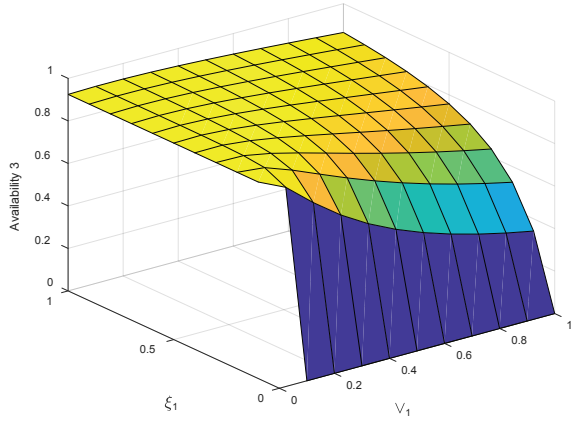


Figure 24. Availability against scenario 3 for v_1 and x_1.

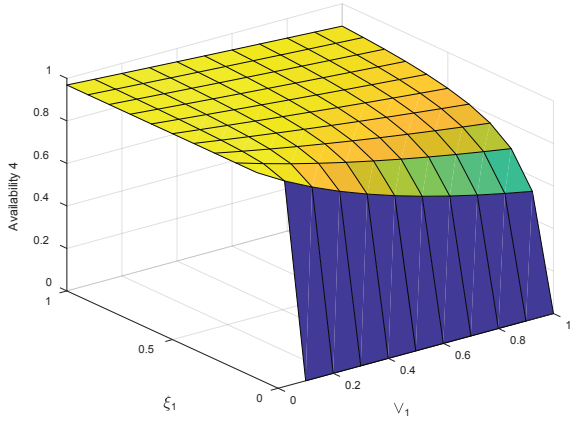


Figure 25. Availability against scenario 4 for v_1 and x_1.

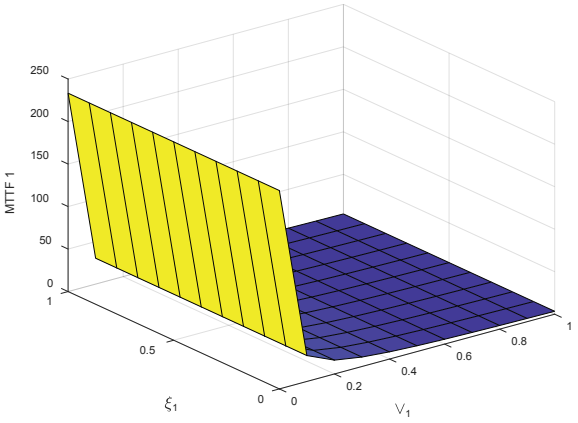


Figure 26. MTTF against scenario 1 for v_1 and x_1.

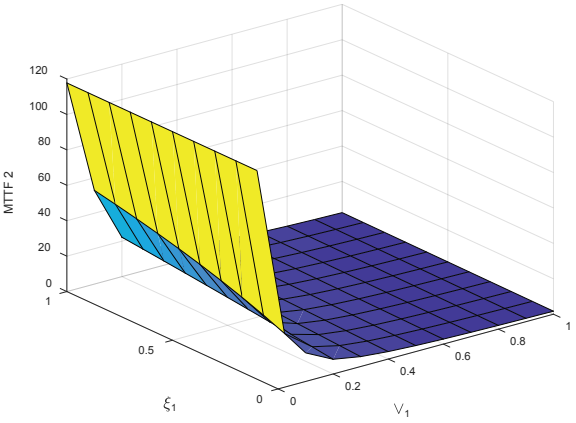


Figure 27. MTTF against scenario 2 for v_1 and x_1.

Table 4. Variation of availability, MTTF and profit with respect to failure rate v_1 for the four scenarios

	Availability				**Mean time to failure (MTTF)**				**Profit*10^7**			
	Scenario				**Scenario**				**Scenario**			
v_1	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
0.0	0.981	0.948	0.983	0.994	155.5	50.29	162.3	507.2	2.453	2.371	2.458	2.486
0.1	0.700	0.884	0.835	0.886	28.57	16.18	60.30	94.85	1.752	2.211	2.088	2.216
0.2	0.544	0.797	0.678	0.799	15.73	8.043	28.19	52.31	1.362	1.992	1.695	1.999
0.3	0.445	0.723	0.565	0.728	10.85	5.307	17.89	36.11	1.114	1.808	1.413	1.820
0.4	0.377	0.661	0.483	0.668	8.284	3.952	13.02	27.58	0.942	1.653	1.207	1.671
0.5	0.326	0.609	0.421	0.618	6.698	3.147	10.20	22.30	0.817	1.523	1.053	1.545
0.6	0.288	0.564	0.373	0.574	5.622	2.613	8.381	18.72	0.720	1.411	0.932	1.436
0.7	0.258	0.526	0.334	0.536	4.844	2.234	7.106	16.13	0.645	1.315	0.837	1.342
0.8	0.233	0.492	0.303	0.503	4.255	1.951	6.166	14.17	0.583	1.231	0.759	1.259
0.9	0.213	0.462	0.277	0.474	3.794	1.731	5.444	12.63	0.532	1.157	0.694	1.186

Table 5. Variation of availability, MTTF and profit with respect to repair rate x_1 for the four scenarios

	Availability				**Mean time to failure (MTTF)**				**Profit*10^7**			
	Scenario				**Scenario**				**Scenario**			
x_1	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
0.0	0.000	0.950	0.000	0.000	56.00	47.96	186.2	185.1	0.000	0.000	0.000	0.000
0.1	0.936	0.889	0.961	0.980	56.00	20.38	60.48	185.1	2.341	2.222	2.403	2.450
0.2	0.939	0.887	0.961	0.980	56.00	20.09	60.07	185.1	2.347	2.219	2.403	2.452
0.3	0.939	0.887	0.961	0.981	56.00	19.99	59.93	185.1	2.349	2.217	2.402	2.452
0.4	0.940	0.886	0.961	0.981	56.00	19.94	59.87	185.1	2.350	2.217	2.402	2.453
0.5	0.940	0.886	0.961	0.981	56.00	19.91	59.83	185.1	2.351	2.216	2.402	2.453
0.6	0.940	0.886	0.961	0.981	56.00	19.89	59.80	185.1	2.351	2.216	2.402	2.453
0.7	0.940	0.886	0.961	0.981	56.00	19.88	59.78	185.1	2.352	2.216	2.402	2.453
0.8	0.941	0.886	0.961	0.981	56.00	19.87	59.76	185.1	2.352	2.216	2.402	2.453
0.9	0.941	0.886	0.961	0.981	56.00	19.86	59.75	185.1	2.352	2.216	2.402	2.453

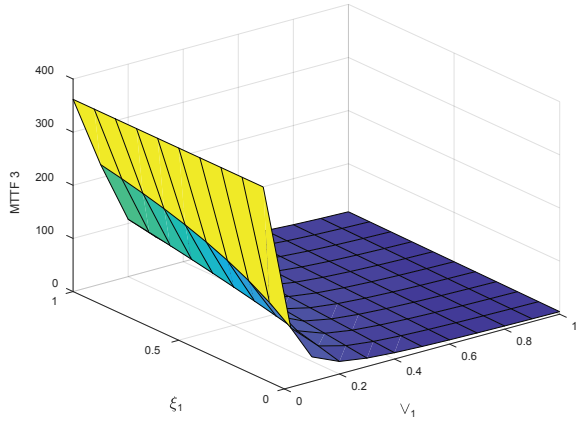
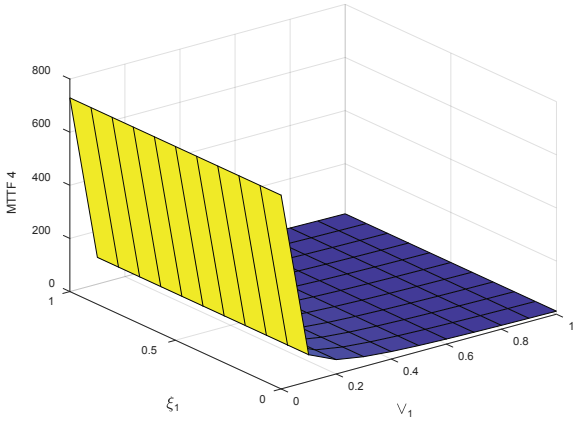


Figure 28. MTTF against scenario 1 for v_1 and x_1.



29. MTTF against scenario 4 for v_1 and x_1.

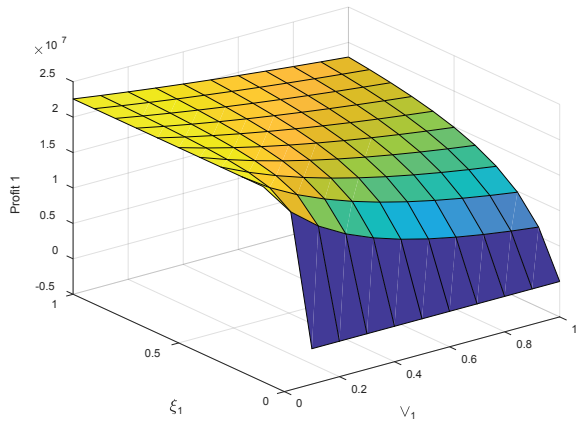


Figure 30. Profit against scenario 1 for v_1 and x_1.

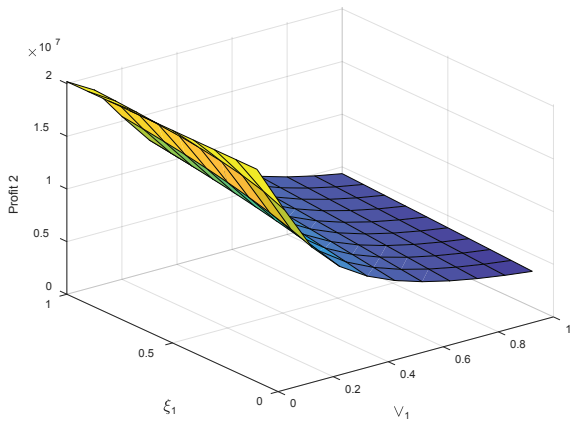


Figure 31. Profit against scenario 2 for v_1 and x_1.

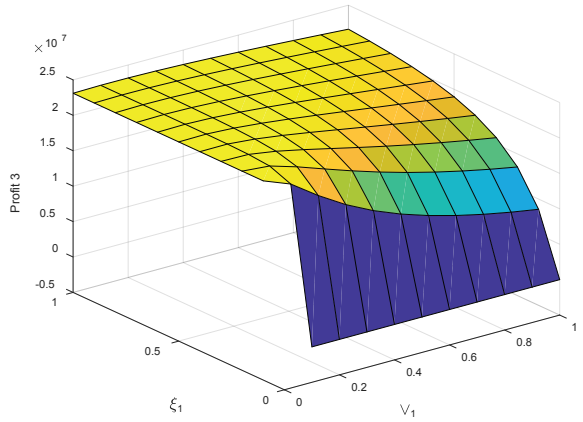


Figure 32. Profit against scenario 3 for v_1 and x_1.

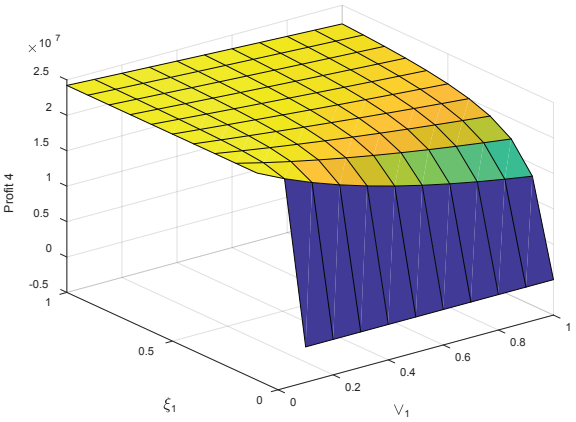


Figure 33. Profit against scenario 4 for v_1 and x_1.

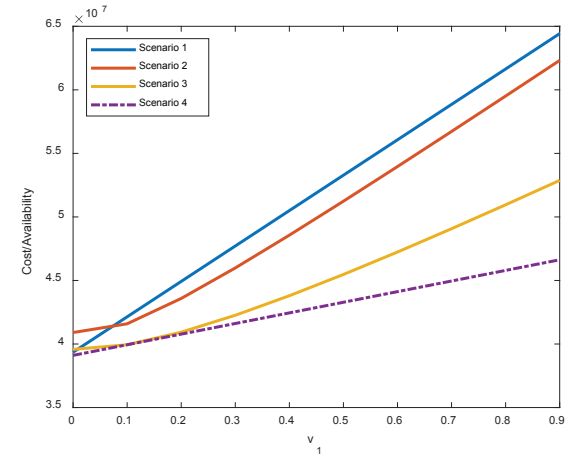


Figure 34. Cost/Availability versus failure rate.

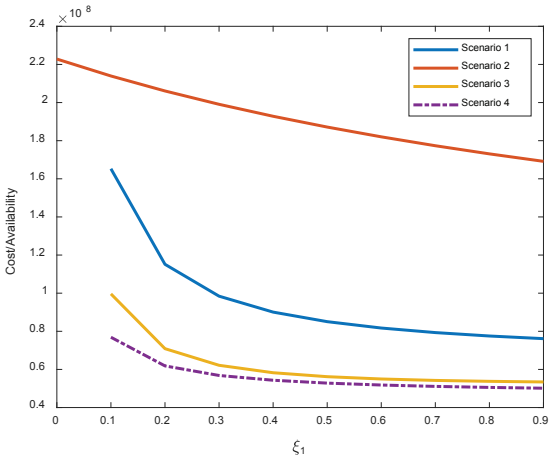


Figure 35. Cost/Availability versus repair rate.

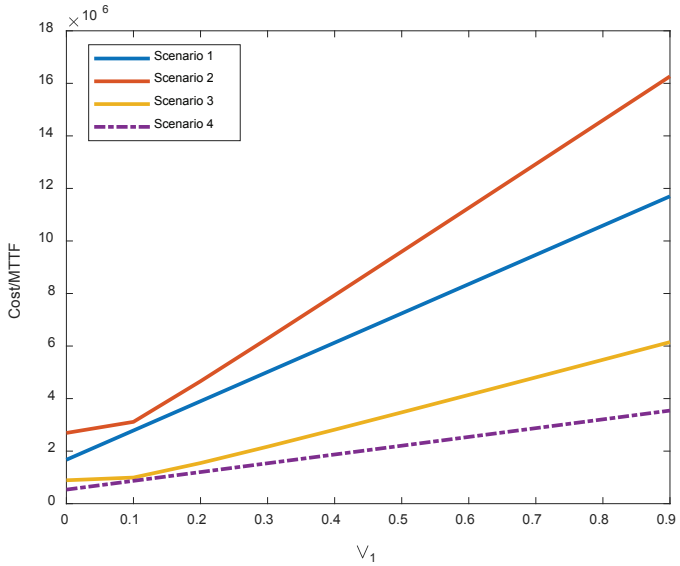


Figure 36. Cost/MTTF versus failure rate.

Table 6. Variation of Availability, MTTF and Profit with respect to failure rate v_2 for the four Scenarios

	Availability				Mean Time To Failure (MTTF)				Profit*10^7			
	Scenario				Scenario				Scenario			
v_2	1	2	3	4	1	2	3	4	1	2	3	4
0.0	0.980	0.949	0.983	0.994	155.5	50.57	162.7	507.2	2.451	2.374	2.459	2.485
0.1	0.726	0.602	0.857	0.898	28.57	16.54	62.49	94.85	1.815	1.506	2.142	2.246
0.2	0.576	0.408	0.725	0.819	15.73	8.170	29.03	52.31	1.441	1.021	1.814	2.048
0.3	0.478	0.307	0.626	0.753	10.85	5.371	18.32	36.11	1.195	0.769	1.565	1.883
0.4	0.408	0.246	0.550	0.697	8.284	3.992	13.27	27.58	1.021	0.616	1.374	1.742
0.5	0.356	0.056	0.489	0.648	6.698	3.174	10.37	22.30	0.891	0.514	1.224	1.621
0.6	0.316	0.176	0.441	0.606	5.622	2.633	8.500	18.72	0.790	0.440	1.104	1.516
0.7	0.284	0.154	0.401	0.569	4.844	2.250	7.195	16.13	0.710	0.385	1.004	1.423
0.8	0.257	0.137	0.368	0.536	4.255	1.963	6.235	14.17	0.644	0.342	0.921	1.341
0.9	0.236	0.123	0.340	0.507	3.794	1.742	5.500	12.63	0.590	0.308	0.851	1.268

Table 7. Variation of Availability, MTTF and Profit with respect to repair rate x_2 for the four Scenarios

	Availability				Mean Time To Failure (MTTF)				Profit*10^7			
	Scenario				Scenario				Scenario			
x_2	1	2	3	4	1	2	3	4	1	2	3	4
0.0	0.000	0.000	0.000	0.000	56.00	37.74	140.2	185.1	0.000	0.000	2.405	2.454
0.1	0.941	0.892	0.962	0.981	56.00	20.02	59.86	185.1	2.354	2.231	2.406	2.464
0.2	0.954	0.894	0.962	0.985	56.00	18.92	58.17	185.1	2.386	2.236	2.406	2.468
0.3	0.958	0.894	0.962	0.987	56.00	18.52	57.59	185.1	2.396	2.234	2.406	2.469
0.4	0.961	0.893	0.962	0.988	56.00	18.32	57.30	185.1	2.402	2.233	2.406	2.471
0.5	0.962	0.892	0.962	0.988	56.00	18.19	57.12	185.1	2.405	2.232	2.406	2.471
0.6	0.963	0.892	0.962	0.988	56.00	18.11	57.00	185.1	2.407	2.231	2.406	2.472
0.7	0.963	0.892	0.962	0.988	56.00	18.05	56.92	185.1	2.409	2.230	2.406	2.472
0.8	0.964	0.891	0.962	0.989	56.00	18.00	56.85	185.1	2.410	2.229	2.406	2.472
0.9	0.964	0.891	0.962	0.989	56.00	17.97	56.80	185.1	2.411	2.229	2.406	2.472

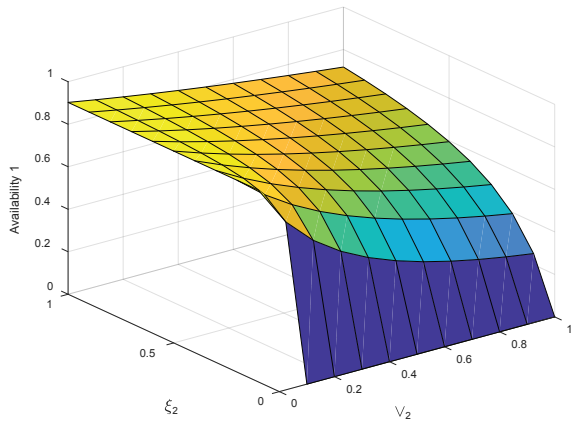


Figure 37. Availability against scenario 1 for v_2 and x_2.

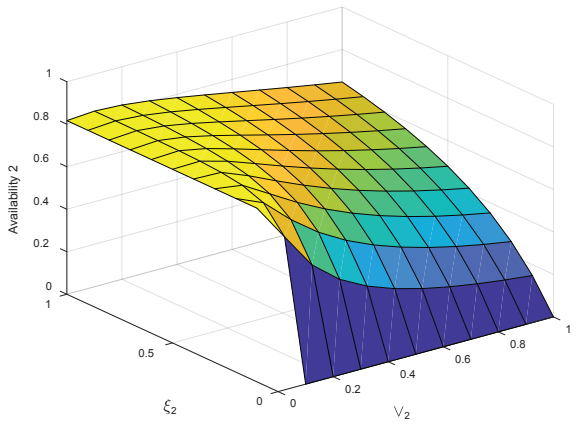


Figure 38. Availability against scenario 2 for v_2 and x_2.

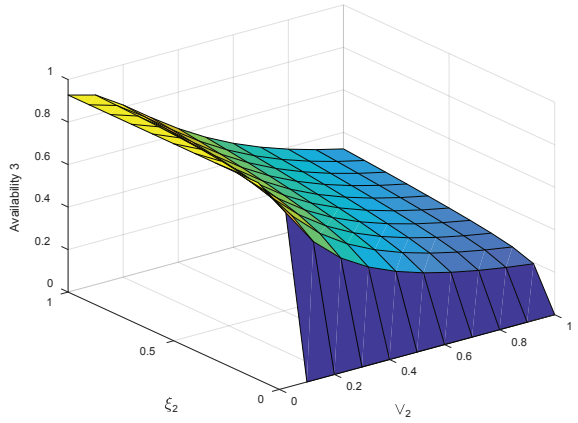


Figure 39. Availability against scenario 3 for v_2 and x_2.

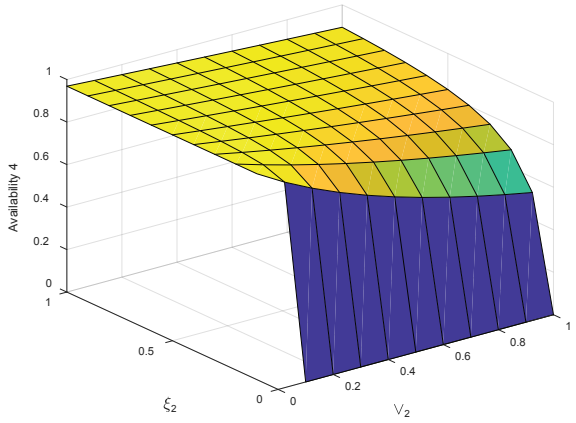


Figure 40. Availability against scenario 4 for v_2 and x_2.

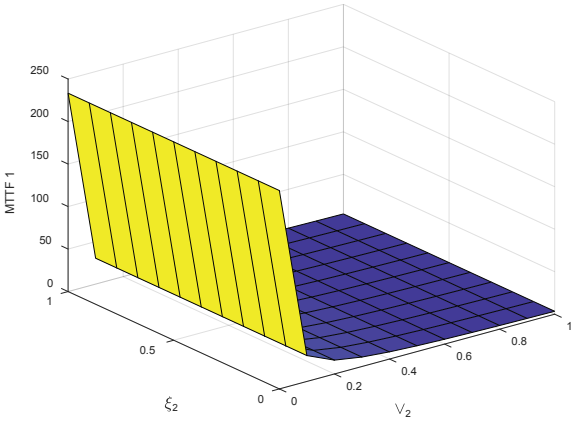


Figure 41. MTTF against scenario 1 for v_2 and x_2.

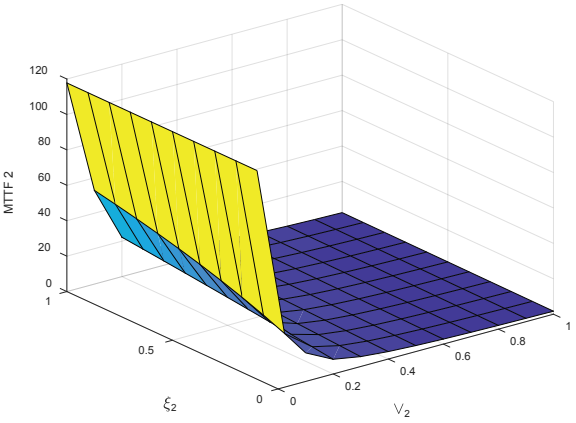


Figure 42. Availability against scenario 2 for v_2 and x_2.

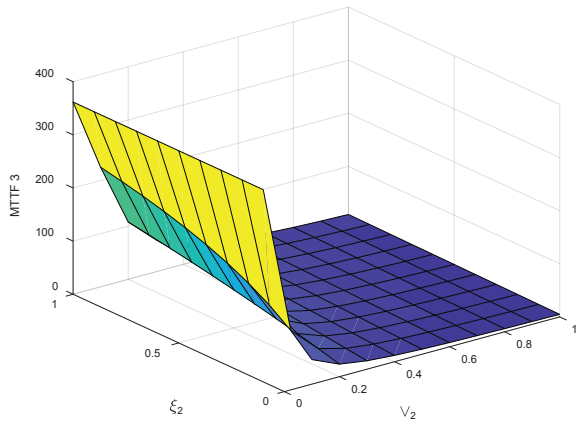


Figure 43. MTTF against scenario 3 for v_2 and x_2.

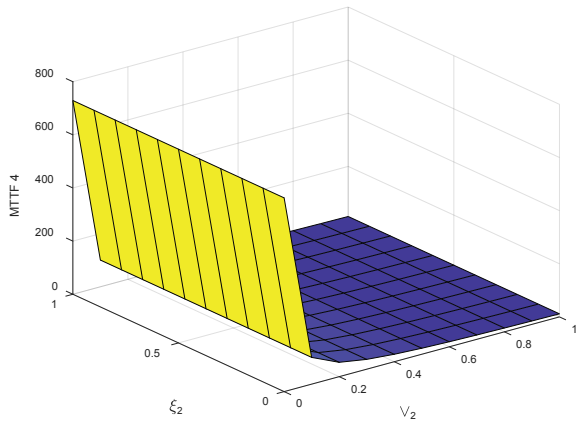


Figure 44. MTTF against scenario 4 for v_2 and x_2.

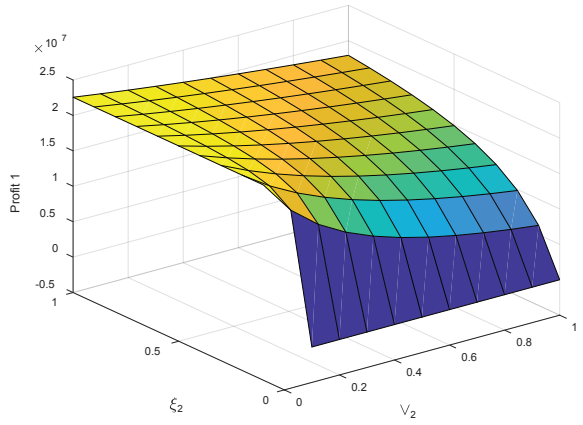


Figure 45. Profit against scenario 1 for v_2 and x_2.

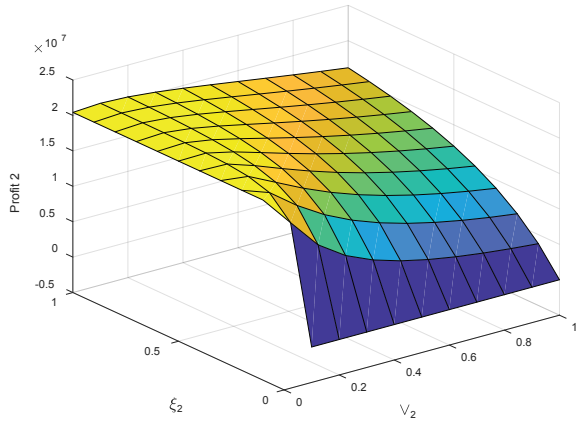


Figure 46. Profit against scenario 2 for v_2 and x_2.

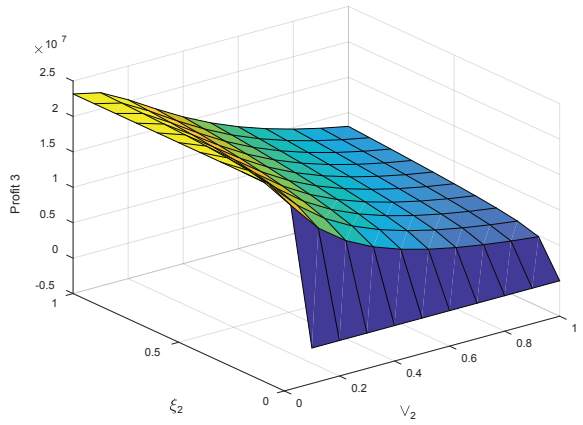


Figure 47. Profit against scenario 3 for v_2 and x_2.

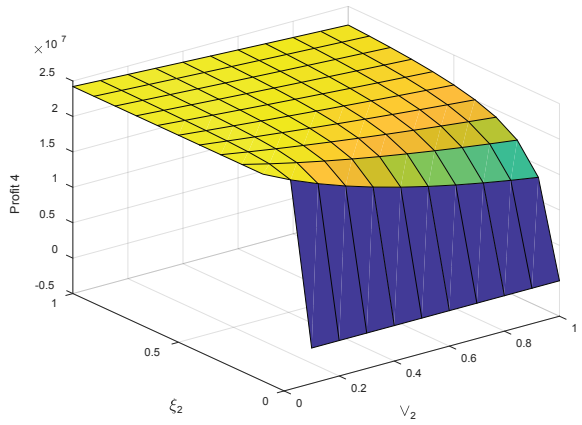


Figure 48. Profit against scenario 4 for v_2 and x_2.

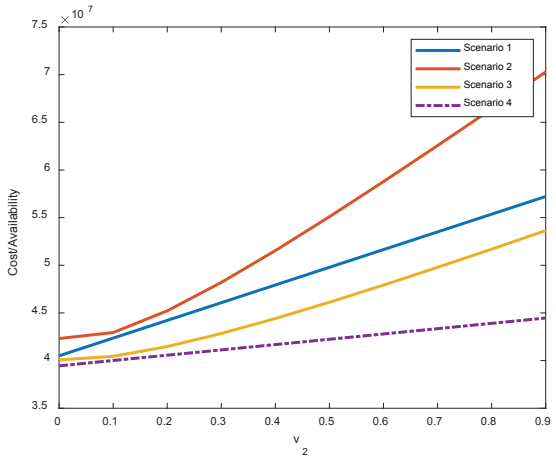


Figure 49. Cost/Availability versus failure rate.

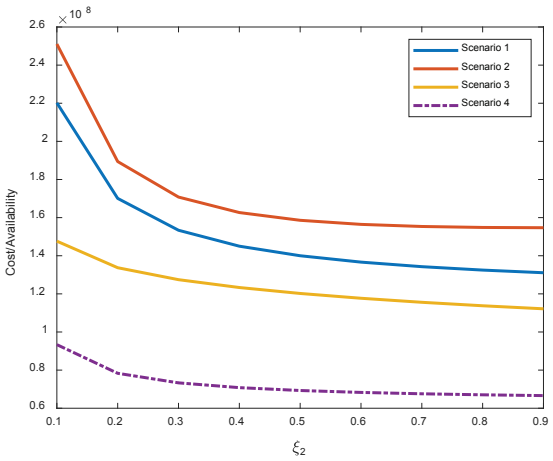


Figure 50. Cost/Availability versus repair rate.

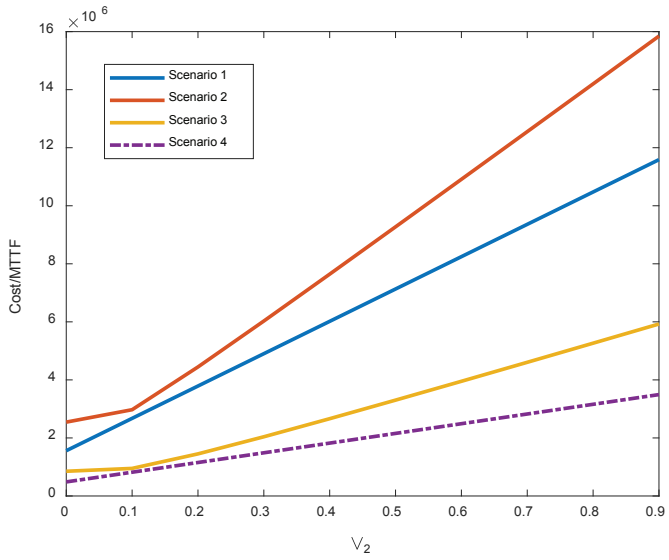


Figure 51. Cost/MTTF versus failure rate.

Table 8. Variation of Availability, MTTF and Profit with respect to failure rate v_3 for the four Scenarios

	Availability				**Mean Time To Failure (MTTF)**				**Profit*10^7**			
	Scenario				**Scenario**				**Scenario**			
v_3	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
0.0	0.950	0.937	0.980	0.984	155.5	54.76	169.4	507.2	2.376	2.344	2.452	2.461
0.1	0.919	0.930	0.979	0.974	28.57	32.92	128.0	94.85	2.298	2.325	2.448	2.435
0.2	0.890	0.900	0.973	0.964	15.73	16.23	73.46	52.31	2.225	2.252	2.433	2.410
0.3	0.862	0.863	0.963	0.954	10.85	9.751	44.98	36.11	2.156	2.157	2.409	2.386
0.4	0.836	0.823	0.951	0.944	8.284	6.705	30.34	27.58	2.092	2.057	2.379	2.361
0.5	0.812	0.783	0.938	0.935	6.698	5.017	22.07	22.30	2.031	1.958	2.345	2.338
0.6	0.789	0.745	0.923	0.926	5.622	3.971	16.97	18.72	1.974	1.864	2.309	2.315
0.7	0.768	0.710	0.908	0.916	4.844	3.267	13.60	16.13	1.920	1.776	2.271	2.292
0.8	0.747	0.677	0.893	0.908	4.255	2.766	11.24	14.17	1.868	1.693	2.233	2.270
0.9	0.728	0.646	0.877	0.899	3.794	2.393	9.517	12.63	1.820	1.617	2.194	2.248

Table 9. Variation of Availability, MTTF and Profit with respect to repair rate x_3 for the four Scenarios

	Availability				**Mean Time To Failure (MTTF)**				**Profit*10^7**			
	Scenario				**Scenario**				**Scenario**			
x_3	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
0.0	0.000	0.000	0.000	0.000	56.00	30.32	107.6	185.1	0.000	0.000	0.000	0.000
0.1	0.919	0.874	0.959	0.974	56.00	20.68	61.34	185.1	2.299	2.186	2.399	2.436
0.2	0.944	0.891	0.962	0.982	56.00	19.31	58.88	185.1	2.361	2.228	2.406	2.456
0.3	0.953	0.893	0.962	0.985	56.00	18.77	58.00	185.1	2.383	2.233	2.407	2.463
0.4	0.957	0.893	0.962	0.986	56.00	18.47	57.55	185.1	2.394	2.233	2.407	2.467
0.5	0.960	0.893	0.962	0.987	56.00	18.29	57.27	185.1	2.400	2.232	2.407	2.469
0.6	0.962	0.892	0.962	0.988	56.00	18.16	57.09	185.1	2.405	2.231	2.406	2.470
0.7	0.963	0.892	0.962	0.988	56.00	18.07	56.95	185.1	2.408	2.230	2.406	2.471
0.8	0.964	0.891	0.962	0.989	56.00	18.00	56.85	185.1	2.410	2.229	2.406	2.472
0.9	0.965	0.891	0.962	0.989	56.00	17.95	56.78	185.1	2.412	2.229	2.406	2.473

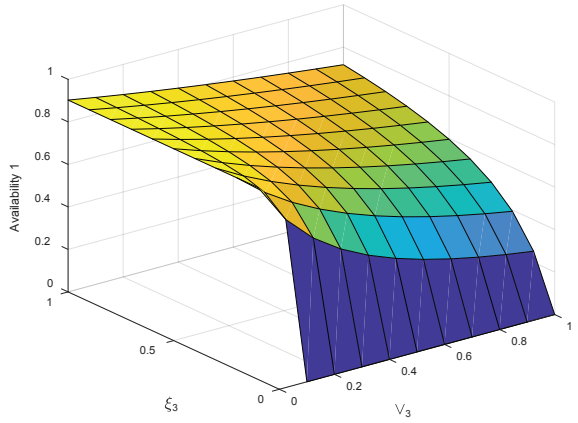


Figure 52. Availability against scenario 1 for v_3 and x_3.

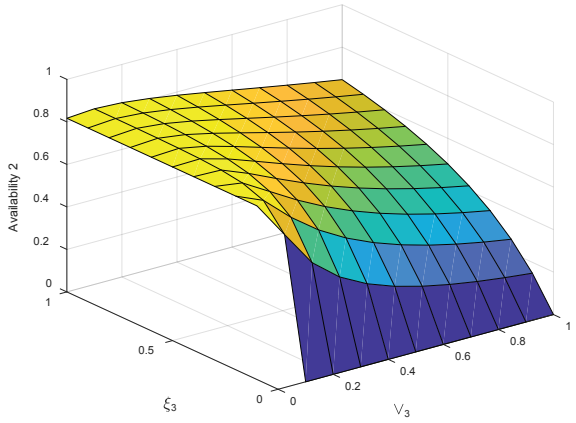


Figure 53. Availability against scenario 2 for v_3 and x_3.

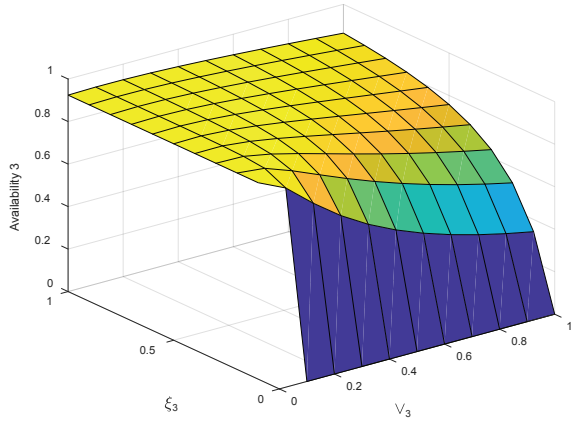


Figure 54. Availability against scenario 3 for v_3 and x_3.

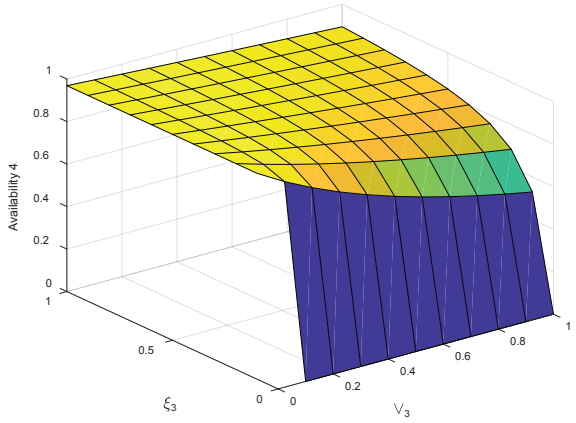


Figure 55. Availability against scenario 4 for v_3 and x_3.

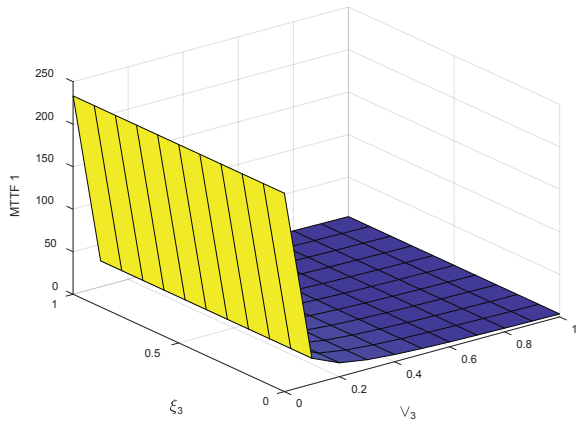


Figure 56. MTTF against scenario 1 for v_3 and x_3

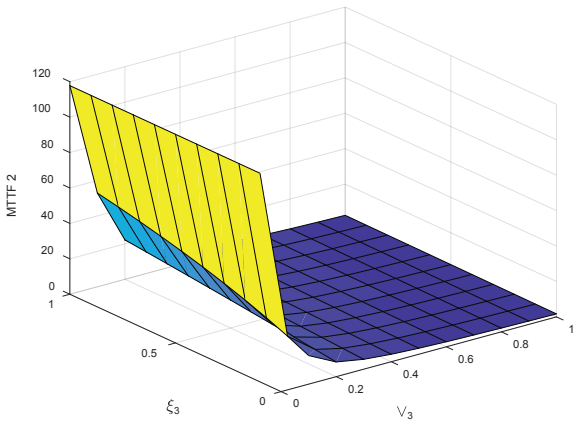


Figure 57. MTTF against scenario 2 for v_3 and x_3.

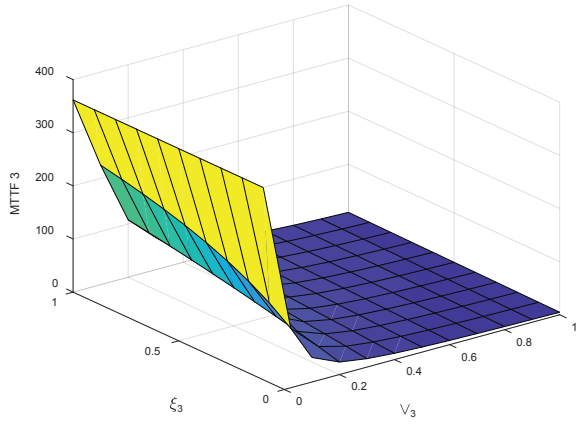


Figure 58. MTTF against scenario 3 for v_3 and x_3.

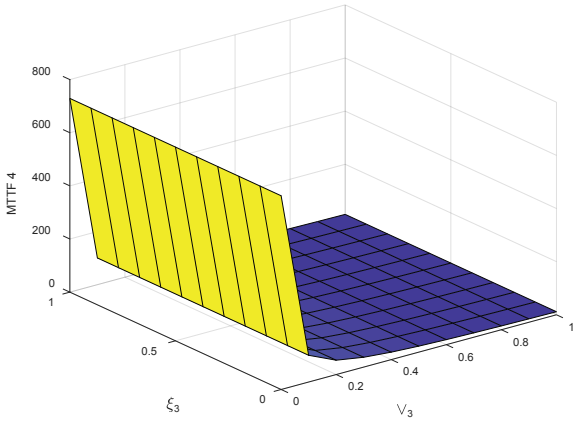


Figure 59. MTTF against scenario 4 for v_3 and x_3.

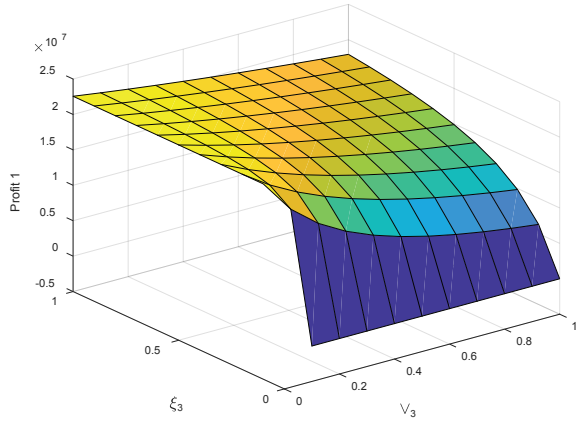


Figure 60. Profit against scenario 1 for v_3 and x_3.

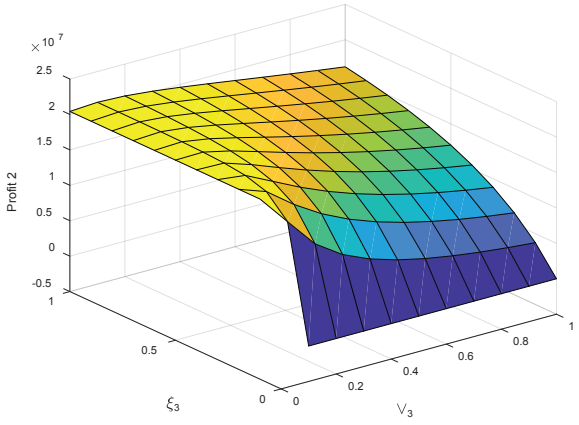


Figure 61. Profit against scenario 2 for v_3 and x_3.

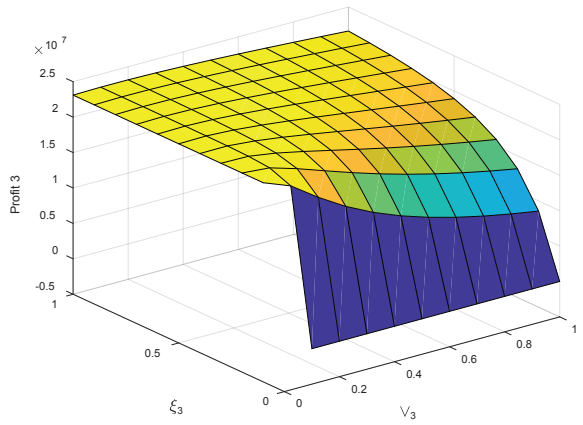


Figure 62. Profit against scenario 3 for v_3 and x_3.

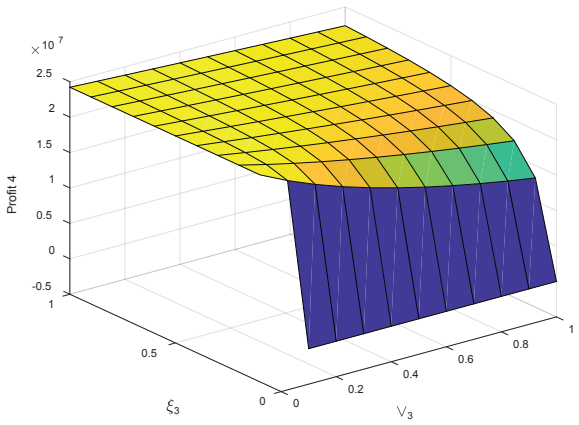


Figure 63. Profit against scenario 4 for v_3 and x_3.

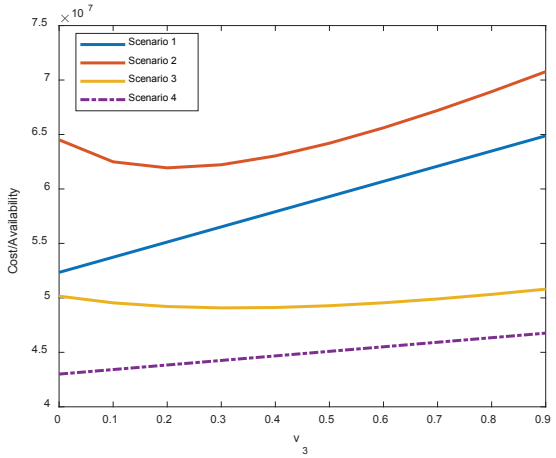


Figure 64. Cost/Availability versus failure rate.

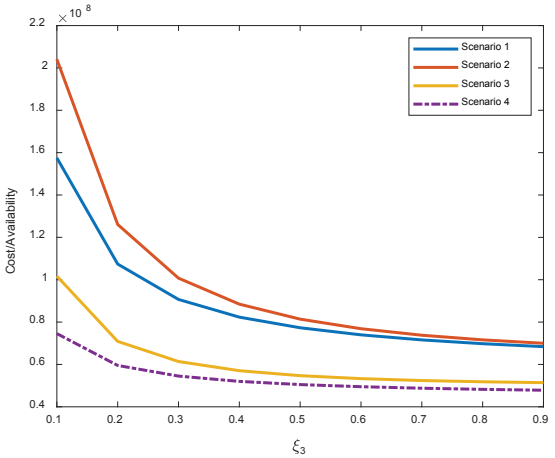


Figure 65. Cost/Availability versus repair rate.

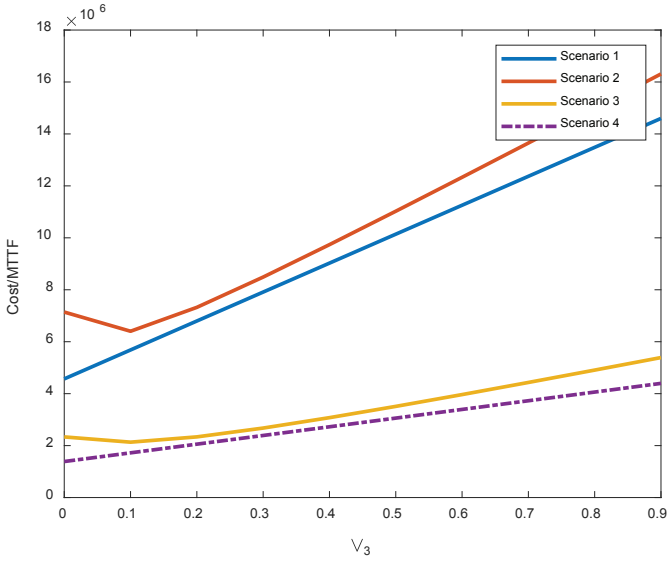


Figure 66. Cost/MTTF versus failure rate.

Table 11. Variation of Availability, MTTF and Profit with respect to repair rate x_4 for the four Scenarios

	Availability				Mean Time To Failure (MTTF)				Profit*10^7			
	Scenario				Scenario				Scenario			
x_4	1	2	3	4	1	2	3	4	1	2	3	4
0.0	0.000	0.000	0.000	0.000	56.00	18.29	57.27	185.1	0.000	0.000	0.000	0.000
0.1	0.901	0.650	0.851	0.968	56.00	18.29	57.27	185.1	2.252	1.625	2.128	2.420
0.2	0.937	0.783	0.917	0.980	56.00	18.29	57.27	185.1	2.342	1.958	2.294	2.450
0.3	0.949	0.840	0.942	0.984	56.00	18.29	57.27	185.1	2.374	2.101	2.355	2.461
0.4	0.956	0.872	0.955	0.986	56.00	18.29	57.27	185.1	2.390	2.181	2.387	2.466
0.5	0.960	0.893	0.962	0.987	56.00	18.29	57.27	185.1	2.400	2.232	2.407	2.469
0.6	0.963	0.907	0.968	0.988	56.00	18.29	57.27	185.1	2.407	2.267	2.420	2.471
0.7	0.964	0.917	0.971	0.989	56.00	18.29	57.27	185.1	2.412	2.267	2.429	2.473
0.8	0.966	0.925	0.974	0.989	56.00	18.29	57.27	185.1	2.415	2.313	2.437	2.474
0.9	0.966	0.931	0.977	0.990	56.00	18.29	57.27	185.1	2.418	2.329	2.442	2.475

Table 10. Variation of Availability, MTTF and Profit with respect to failure rate v_4 for the four Scenarios

	Availability				Mean Time To Failure (MTTF)				Profit*10^7			
	Scenario				Scenario				Scenario			
v_4	1	2	3	4	1	2	3	4	1	2	3	4
0.0	0.949	0.966	0.991	0.984	107.6	307.8	1577	353.5	2.373	2.416	2.478	2.460
0.1	0.924	0.843	0.941	0.976	42.42	12.71	37.10	140.5	2.310	2.109	2.354	2.440
0.2	0.900	0.748	0.897	0.967	26.41	6.492	18.80	87.71	2.251	1.871	2.242	2.419
0.3	0.877	0.672	0.856	0.959	19.17	4.359	12.60	63.75	2.194	1.681	2.140	2.399
0.4	0.856	0.610	0.819	0.952	15.05	3.281	9.400	50.07	2.141	1.526	2.047	2.380
0.5	0.835	0.559	0.785	0.944	12.38	2.630	7.600	41.22	2.089	1.397	1.962	2.361
0.6	0.816	0.515	0.753	0.936	10.52	2.195	6.300	35.03	2.041	1.289	1.884	2.342
0.7	0.797	0.478	0.724	0.929	9.150	1.883	5.400	30.46	1.994	1.196	1.811	2.323
0.8	0.780	0.446	0.697	0.922	8.092	1.649	4.700	26.94	1.950	1.115	1.744	2.305
0.9	0.763	0.418	0.672	0.914	7.253	1.466	4.200	24.15	1.907	1.045	1.682	2.286

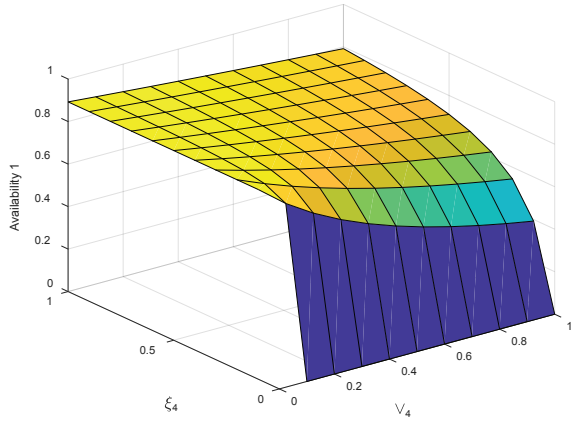


Figure 67. Availability against scenario 1 for v_4 and x_4.

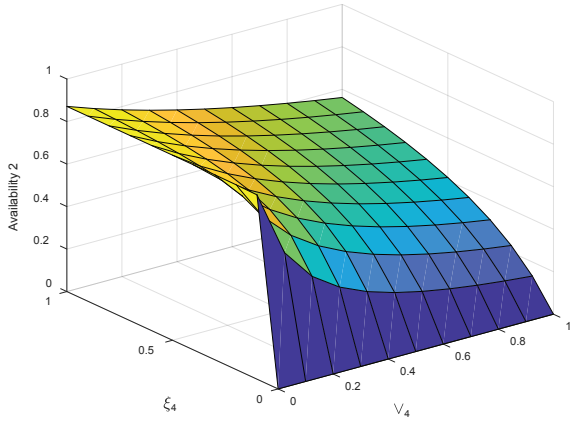


Figure 68. Availability against scenario 2 for v_4 and x_4.

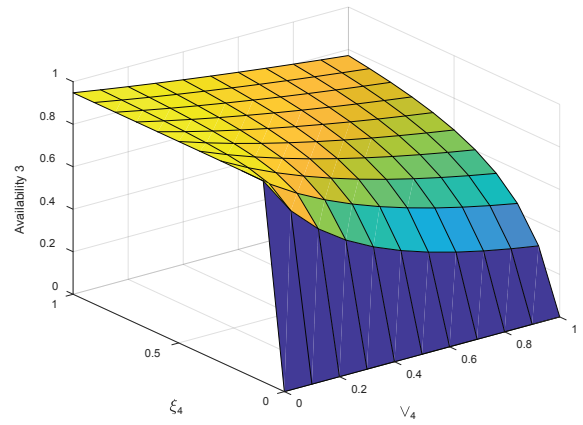


Figure 69. Availability against scenario 3 for v_4 and x_4.

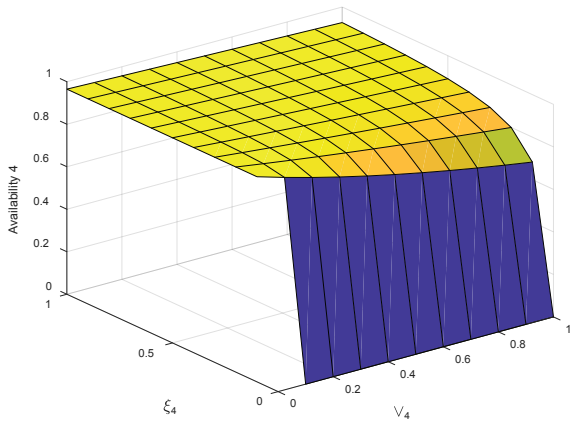


Figure 70. Availability against scenario 4 for v_4 and x_4.

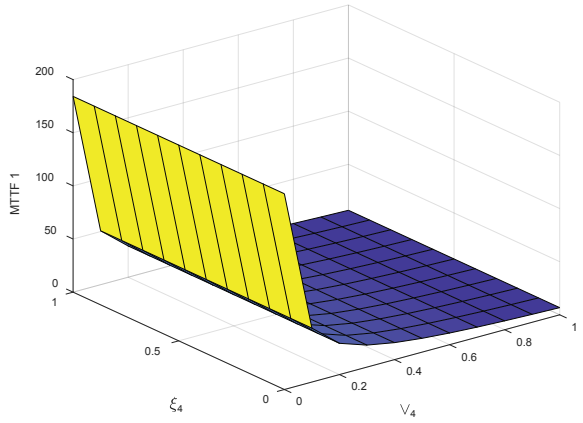


Figure 71. MTTF against scenario 1 for v_4 and x_4.

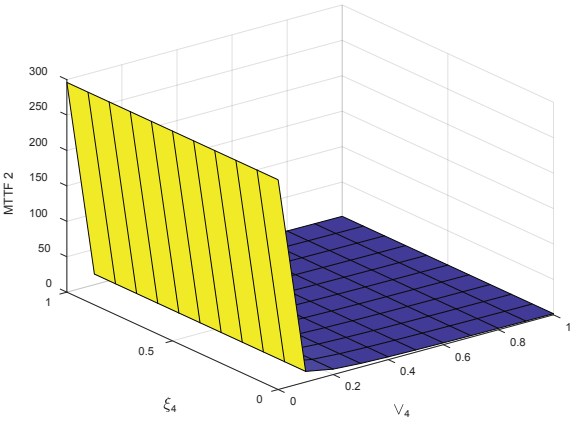


Figure 72. MTTF against scenario 2 for v_4 and x_4.

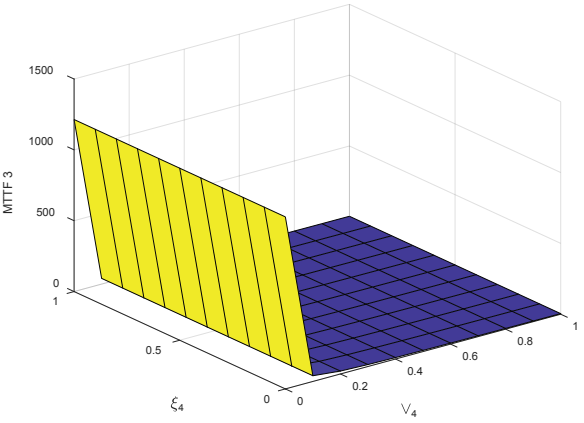


Figure 73. MTTF against scenario 3 for v_4 and x_4.

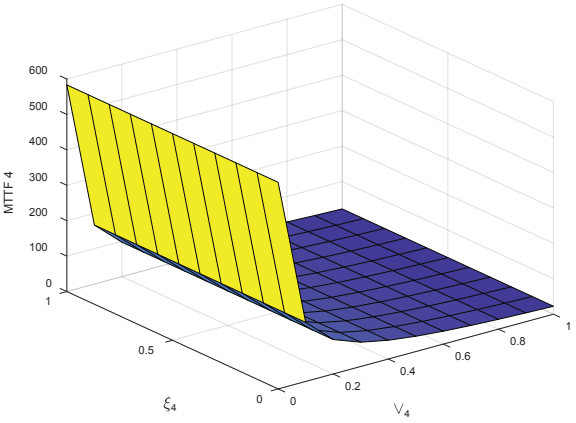


Figure 74. MTTF against scenario 5 for v_4 and x_4.

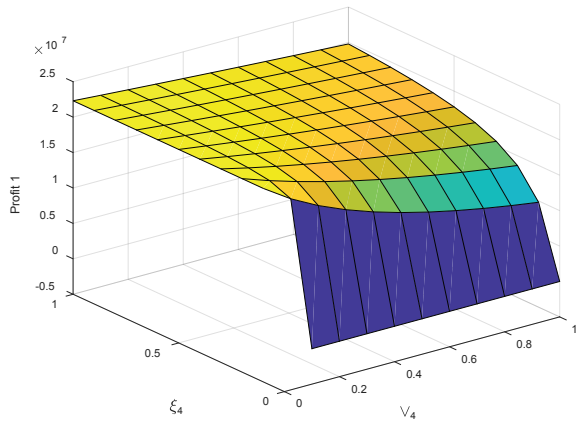


Figure 75. Profit against scenario 1 for v_4 and x_4.

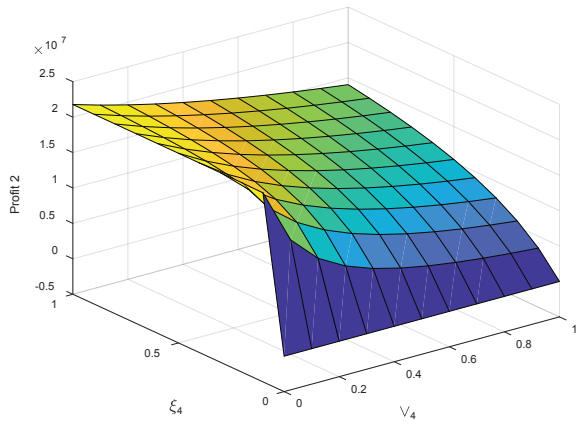


Figure 76. Profit against scenario 2 for v_4 and x_4.

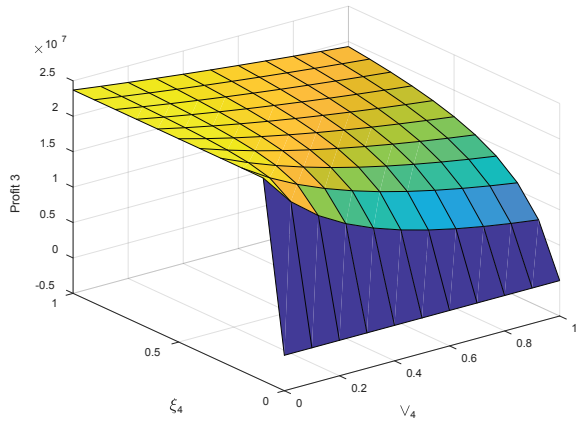


Figure 77. Profit against scenario 3 for v_4 and x_4.

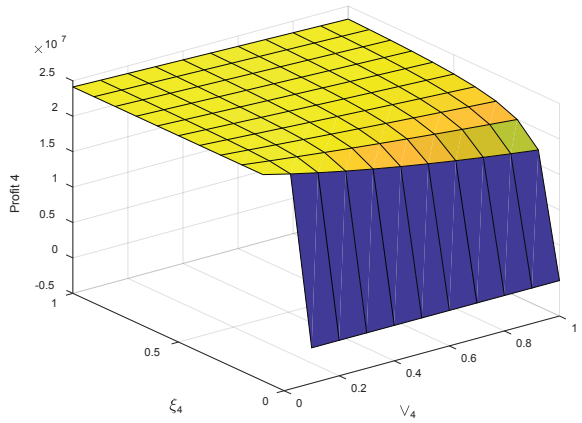


Figure 78. Profit against scenario 4 for v_4 and x_4.

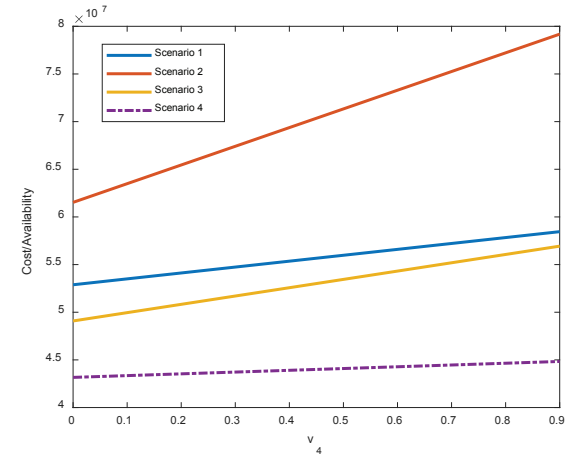


Figure 79. Cost/Availability versus failure rate.

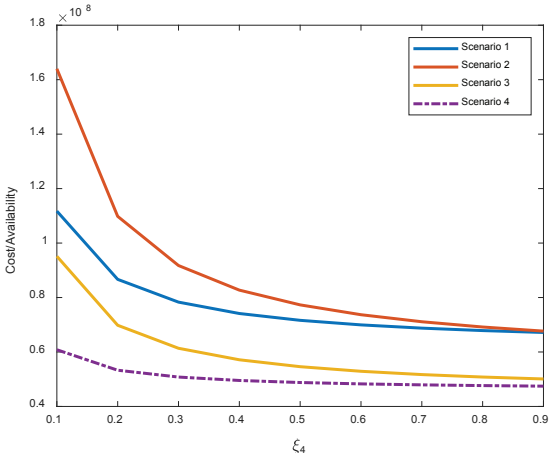


Figure 80. Cost/Availability versus repair rate.

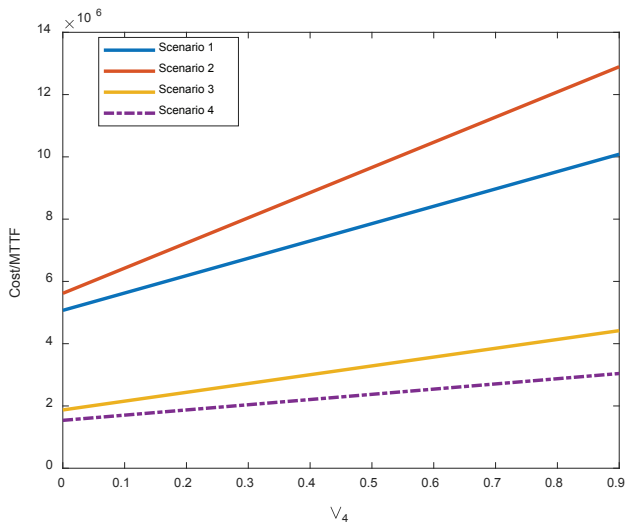


Figure 81. Cost/MTTF versus failure rate.

and at the same time MTTF and Cost analysis increases as the repair rate increases. Lastly, profit analysis was carried out in all the four (4) scenarios and throughout the findings it observed that profit in terms of failure rate decreases as the failure rate increase and also profit increases as the repair rate increases. To this fact, the clients require optimal maintenance action in order to avoid huge downfall and adequate the life span of the network.

CONCLUSION

In this research, computer network system consisting 2-clients, load balancers, 2-fog nodes and 2-cloud server arranged in series-parallel was considered. Thorough differential equations and detailed mathematical expressions of availability, sensitivity, MTTF and profit were derived. It is true from the figures that optimal performance level was obtained. The presented numerical results have shown the implication of repair and failure rates on the network availability. From numerical results, it is enough to show that availability or reliability and MTTF can be increase with adequate maintenance to truncate network downfall through fault tolerant units/subsystem like introducing more load balancers and more number of cloud servers. The presented research work will help plant management to shun away on an erroneous performance assessment caused by poor system design. Failure occurrence and monitoring of condition can be extended and incorporated to allow management in approving the optimal replacement/ maintenance time.

ACKNOWLEDGEMENTS

This paper is supported by the National Social Science Foundation of China (Grant No. 20BTJ044)

AUTHORSHIP CONTRIBUTIONS

Muhammad Salihu Isa initiate the model and do all the writing and mathematical analysis while Jinbiao Wu helps in editing and supervision.

CONFLICT OF INTEREST

Authors have declared that there is no conflict of interest with regard to this research

REFERENCES

[1] Yusuf I, Ismail AL, Singh VV, Ali UA, Sufi NA. Performance analysis of multi-computer system consisting of three subsystems in series configuration using copula repair policy. SN Comput Sci 2020;1:241. [CrossRef]

[2] Yusuf I, Ismail AL, Ali UA. Reliability analysis of communication network system with redundant relay station under partial and complete failure. J Math Comput Sci 2020;10:863–880.

[3] Singh VV, Ismail AL, Chand U, Maiti SS. Performance assessment of complex system under the k-out-of-n: G type configuration with k consecutive degraded states through the copula repair approach. Int J Reliab Qual Saf Eng 2022;29:2150047. [CrossRef]

[4] Liu L. Optimization design of computer network reliability based on genetic algorithms. Chem Eng Trans 2016;51:775–780.

[5] Singh VV, Ismail AL, Yusuf I, Abdullahi AH. Probabilistic assessment of Computer-Based Test (CBT) network system consists of four subsystems in series configuration using copula linguistic approach. J Reliab Stat Stud 2021;13:401–428. [CrossRef]

[6] Isa MS, Yusuf I, Ali UA, Suleiman K, Yusuf B, Ismail AL. Reliability analysis of multi-workstation computer network configured as series-parallel system via Gumbel - Hougaard family copula. Int J Oper Res 2022;19:13–26.

[7] Isa MS, Abubakar MI, Ibrahim KH, Yusuf I, Tukur I. Performance analysis of complex series parallel computer network with transparent bridge using copula distribution. Int J Reliab Risk Saf Theory Appl 2021;4:47–59. [CrossRef]

[8] Ahlawat S, Anand A. An introduction to computer networking. Int J Comput Sci Inform Technol Res 2014;2:373–377.

[9] Garg H. Multi objective non-linear programming problem for reliability optimization in intuitionistic fuzzy environment. Front Inform Syst 2018;2:197–229. [CrossRef]

[10] Garg S, Sejwal S, Solanki J. An approach to resolve heterogeneity using RPC in client server systems. Int J Eng Appl Sci Technol 2019;4:301–305. [CrossRef]

[11] Aydın Er B, Şişman A, Ardalı Y. Applicability of radial-based artificial neural networks (RBNN) on coliform calculation: A case of study. Sigma J Eng Nat Sci 2022;40:724–731. [CrossRef]

[12] Şekerci AZ, Aydın N. A stochastic model for facility locations using the priority of fuzzy AHP. Sigma J Eng Nat Sci 2022;40:649–662. [CrossRef]

[13] Yusuf I, Auta AA. Availability analysis of a distributed system with homogeneity in client and server under four different maintenance options. Life Cycle Reliab Saf Eng 2021;10:355–371. [CrossRef]

[14] Handoko H, Isa SM, Si S, Kom M. High availability analysis with database cluster, load balancer and virtual router redudancy protocol. In proceedings of the 2018 3rd International Conference on Computer and Communication Systems (ICCCS); 2018 Apr 27–30; Negoya, Japan. IEEE; 2018. pp. 482–6. [CrossRef]

[15] Ivanovic M, Vidakovic M, Budimac Z, Dejan M. A scalable distributed architecture for client and server-side software agents. Vietnam J Comput Sci 2017;4:127–137. [CrossRef]

[16] Khouni S, Chemali H. SSEA for PSN: A novel secure technique of communication through IOT devices. Sigma J Eng Nat Sci 2022;40:300–309. [CrossRef]

[17] Kenan E, Mustafa CK, Boru B. Comparison of gesture classification methods with contact and non-contact sensors for human-computer interaction. Sigma J Eng Nat Sci 2021;40:219–226.

[18] Shahen AFMS, Mehmet AK. Optimization of drones communication by using meta-heuristic optimization algorithms. Sigma J Eng Nat Sci 2022;40:108–117. [CrossRef]

[19] Kamal J, Murshed M, Buyya R. Workload-aware incremental repartitioning of shared-nothing distributed databases for scalable OLTP applications. Future Gener Comput Syst 2016;56:421–435. [CrossRef]

[20] Khan Z, Alam M, Haidri RA. Effective load balance scheduling schemes for heterogeneous distributed system. Int J Electr Comput Eng 2017;7:2757–2765. [CrossRef]

[21] Ke JC, Liu TH. Repairable system with imperfect coverage and reboot. Appl Math Comput 2014;246:148–158. [CrossRef]

[22] Isa MS, Yusuf I, Ali UA, Jinbiao W. Series-parallel computer system performance evaluation with human operator using gumbel hougaard family copula. In: Isa MS, Yusuf I, Ali UA, Jinbiao W, editors. Computational Intelligence in Sustainable Reliability Engineering. New Jersey: Wiley; 2023. pp.109–127. [CrossRef]

[23] Kumar P, Jain M, Meena RK. Optimal control of fault tolerant machining system with reboot and recovery in fuzzy environment using harmony search algorithm. ISA Trans 2022;119:52–64. [CrossRef]

[24] Ritu A, Vaishali T, Mangey R. Multi-state system analysis with imperfect fault coverage, human error and standby strategies. Rev Investig Oper 2020;41:214–231.

[25] Tyagi V, Arora R, Mangey R, Ioannis ST. Copula based measures of repairable parallel system with fault coverage. Int J Math Eng Manag Sci 2021;6:322–344. [CrossRef]

[26] Yusuf I, Ismail AL, Sufi NA, Ambursa FU, Sanusi A, Isa MS. Reliability analysis of distributed system for enhancing data replication using gumbel hougaard family copula approach joint probability distribution. J Ind Eng Int 2021;17:59–78.

[27] Abdullahi S, Yusuf I. Reliability assessment and profit analysis of distributed data center network topology. Life Cycle Reliab Saf Eng 2022;11:59–78. [CrossRef]

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1239–1244, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00098

Research Article

Solving difference equations using fourier transform method

Sunnet AVEZOV[1], Ahmad ISSA[1,*], Murat DÜZ[1]

[1]*Department of Mathematics, Faculty of Science, Karabuk University, Karabuk, 78050, Türkiye*

ARTICLE INFO

Article history
Received: 28 January 2023
Revised: 04 April 2023
Accepted: 07 July 2023

Keywords:
Coefficients Matrix; Fourier Transform Method; Non Homogeneous Linear Difference Equations

ABSTRACT

This article mainly focuses on presenting a new accurate technique (Fourier Transform Method) for solving linear of m[th] order Difference Equations with constant coefficients. Also, a new lower triangular matrix was introduced to overcome problems related to finding the Fourier Transform of polynomials by rewriting standard-based polynomials through the fallen power polynomial base. Besides, five examples have been presented to illustrate the validity and accuracy of this method. The results reveal that the Fourier transform method is very effective and attractive in solving the difference equations.

Cite this article as: Avezov S, Issa A, Düz M. Solving difference equations using fourier transform method. Sigma J Eng Nat Sci 2024;42(4):1239–1244.

INTRODUCTION

Difference equations have motivated a variety amounts of research. They arise in dynamical systems, biology, electrical circuit analysis, economics, and models [1-4]. Many researchers have introduced different methods to solve these equations such as Berinde [5] solved the second-order difference equation by converting them into two first-order nonhomogeneous difference equations. In [6] Feldmann implemented the so-called discrete Laplace transformation method to solve systems of difference equations and linear difference equations with constant coefficients. Several methods for solving linear difference equations are Dirichlet series transform method [7], Chebyshev series [8], Differential transform method [9], Taylor polynomial [10], and other methods (see [11-15]).

In this article, we will employ Fourier Transform Method (FTM) to solve non homogeneous linear of m^{th}-order difference equation with constant coefficients, that

$$a_m\Delta^m U_n + a_{m-1}\Delta^{m-1}U_n + \cdots + a_1\Delta U_n + a_0U_n = \zeta(n), \forall m = 1,2,\ldots \quad (1)$$

where ζ(n) is a given function, a_m, a_{m-1},… a_1, a_0 are constant coefficients, $\Delta^i U_n = \Delta^{i-1}\Delta U_n$ for all $i \geq 2$ and $\Delta U_n = U_{n+1} - U_n$. The article is organized as follows: In Section 2, new basic theorems about the Fourier transform method related to difference equations are introduced. In section 3, five examples are solved by using FTM. Finally, the conclusions are presented in section 4.

Preliminaries

Definition 2.1. [16] Let $U(t)$ be piecewise continuous function defined on the entire number line, then its Fourier transform in the angular frequency form is following:

$$\mathcal{F}[U(t)] = \int_{-\infty}^{\infty} U(t)e^{-i\omega t}dt = \widehat{U}(\omega). \quad (2)$$

*Corresponding author.
*E-mail address: ahmad93.issa18@gmail.com

This paper was recommended for publication in revised form by Editor in Chief Ahmet Selim Dalkilic

Definition 2.2. [16] The inverse Fourier transform of $\widehat{U}(\omega)$ is given by:

$$\mathcal{F}^{-1}\left[\widehat{U}(\omega)\right] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{U}(\omega)e^{i\omega t}d\omega = U(t). \quad (3)$$

The difference equations are the problem of finding the given sequence over the relation of previous terms of a defined sequence on natural numbers. To make it possible to take the Fourier transform of the sequence we define a new function based on that sequence as follows:

Let U_n be any given sequence defined on natural numbers we define corresponding function

$$U(t) = \sum_{n=0}^{\infty} U_n S_n(t), \quad (4)$$

where $S_n(t) = \begin{cases} 1, & if\ n \le t < n+1 \\ 0, & otherwise \end{cases}$.

The function $U(t)$ defined in this way with domain $[0,\infty)$ coincides with the sequence U_n in the natural numbers. In this case, the sequence U_n is the solution of the difference equation (1) if and only if the function $U(t)$ satisfies this equation.

Now we give the Fourier transform of this function and the same shifting properties as well.

$$\mathcal{F}[U(t)] = \widehat{U}(\omega) = \int_{-\infty}^{\infty} U(t)e^{-i\omega t}dt = \int_{-\infty}^{\infty} e^{-i\omega t}\sum_{n=0}^{\infty} U_n S_n(t)\,dt$$

$$= \sum_{n=0}^{\infty} U_n \int_{-\infty}^{\infty} S_n(t)e^{-i\omega t}dt = \sum_{n=0}^{\infty} U_n \int_{n}^{n+1} 1\, e^{-i\omega t}dt$$

$$= \sum_{n=0}^{\infty} U_n \frac{e^{-i\omega t}}{-i\omega}\Big|\begin{matrix} t = n+1 \\ t = n \end{matrix} = \frac{1-e^{-i\omega}}{i\omega}\sum_{n=0}^{\infty} U_n e^{-i\omega n},$$

so

$$\mathcal{F}[U(t)] = \frac{1-e^{-i\omega}}{i\omega}\sum_{n=0}^{\infty} U_n e^{-i\omega n}. \quad (5)$$

The Basic Theorems

Theorem 2.1.1. Let $\widehat{U}(\omega)$ the Fourier transform of $U(t), i, e.\ \mathcal{F}[U(t)] = \widehat{U}(\omega)$, then

$$\mathcal{F}[U(t+k)] = e^{i\omega k}\widehat{U}(\omega) - \hat{S}_0\left(e^{i\omega k}U_0 + e^{i\omega(k-1)}U_1 + \cdots + e^{i\omega}U_{k-1}\right).$$

Proof. We assume that $U(t + k)$ is presented as follows

$$U(t+k) = \sum_{n=0}^{\infty} U_n S_n(t+k),$$

where $S_n(t+k) = \begin{cases} 1, & if\ n \le t+k < n+1 \\ 0, & otherwise \end{cases}$

$$= \begin{cases} 1, & if\ n-k \le t < n-k+1 \\ 0, & otherwise \end{cases}$$

Now using Fourier transform on both sides of Eq (6), we obtain

$$\mathcal{F}[U(t+k)] = \mathcal{F}\left[\sum_{n=0}^{\infty} U_n S_n(t+k)\right] = \sum_{n=0}^{\infty} U_n \mathcal{F}[S_n(t+k)]$$
$$= \sum_{n=k}^{\infty} U_n e^{i\omega k}\mathcal{F}[S_n(t)] = e^{i\omega k}\sum_{n=k}^{\infty} U_n e^{i\omega k}\hat{S}_0(\omega)e^{-i\omega n}$$
$$= e^{i\omega k}\hat{S}_0(\omega)\left(\sum_{n=0}^{\infty} U_n e^{-i\omega n} - U_0 - e^{-i\omega}U_1 - \cdots - e^{-i\omega(k-1)}U_{k-1}\right)$$
$$= e^{i\omega k}\widehat{U}(\omega) - \hat{S}_0\left(e^{i\omega k}U_0 + e^{i\omega(k-1)}U_1 + \cdots + e^{i\omega}U_{k-1}\right),$$

where

$$\mathcal{F}[S_n(t)] = \int_n^{n+1} e^{-i\omega t}dt = \frac{1-e^{-i\omega}}{i\omega}e^{-i\omega n}$$
$$= \mathcal{F}[S_0(t)]e^{-i\omega n} = \hat{S}_0(\omega)e^{-i\omega n}.$$

Theorem 2.1.2. If $U(t) = 1$ then $\mathcal{F}[U(t)] = \frac{1}{i\omega}$, for $Im(\omega) < 0$.

Proof. By using Eq (5) we get

$$\mathcal{F}[1] = \frac{1-e^{-i\omega}}{i\omega}\sum_{n=k}^{\infty} e^{-i\omega n} = \frac{1-e^{-i\omega}}{i\omega}\frac{1}{1-e^{-i\omega}} = \frac{1}{i\omega},$$

for

$$\left|e^{-i\omega}\right| < 1 \Leftrightarrow \left|e^{-i(x+y)}\right| < 1 \Leftrightarrow \left|e^{-ix}\right|\left|e^{y}\right| < 1 \Leftrightarrow Im(\omega) < 0.$$

Theorem 2.1.3. If $U(t) = t$, then $\mathcal{F}[U(t)] = \frac{1}{i\omega(e^{i\omega}-1)}$, for $Im(\omega) < 0$.

Proof. By using Eq (5) we obtain

$$\mathcal{F}[t] = \frac{1-e^{-i\omega}}{i\omega}\sum_{n=k}^{\infty} ne^{-i\omega n} = \frac{1-e^{-i\omega}}{i\omega}\frac{e^{-i\omega}}{(1-e^{-i\omega})^2}$$
$$= \frac{1}{i\omega(e^{i\omega}-1)},$$

for $Im(\omega) < 0$.

Theorem 2.1.4. Let $k \in \mathbb{N}$, $t \ge k$ and $U(t) = t^{\underline{k}} = t(t\text{-}1)\ldots(t - k + 1)$ then $\mathcal{F}[U(t)] = \frac{k!}{i\omega(e^{i\omega}-1)^k}$,

for $Im(\omega) < 0$.

Proof. We know that: $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}, |x| < 1$

By differentiating both sides with respect to x, we get:

$$\sum_{n=1}^{\infty} nx^{n-1} = \frac{1}{(1-x)^2}, \sum_{n=2}^{\infty} n(n-1)x^{n-2} = \frac{2}{(1-x)^3},$$

And more generally

$$\sum_{n=k}^{\infty} n(n-1)(n-2)\dots(n-k+1)x^{n-k} = \frac{k!}{(1-x)^{k+1}}. \quad (7)$$

We can rewrite the equation (7) as follows

$$\sum_{n=k}^{\infty} n(n-1)(n-2)\dots(n-k+1)x^{n} = \frac{x^k k!}{(1-x)^{k+1}}, |x|<1. \quad (8)$$

Now by applying Fourier transform of our function and using Eq (5) we obtain

$$\mathcal{F}[t^{\underline{k}}] = \frac{1-e^{-i\omega}}{i\omega}\sum_{n=k}^{\infty} n(n-1)(n-2)\dots(n-k+1)e^{-i\omega n}$$

$$= \frac{1-e^{-i\omega}}{i\omega}\frac{e^{-i\omega k}k!}{(1-e^{-i\omega})^{k+1}} = \frac{k!}{i\omega(e^{i\omega}-1)^k},$$

for $Im(\omega) < 0$.

Theorem 2.1.5. Let $a \in \mathbb{C}$, $a \neq 0$ and $U(t) = a^t$, then $\mathcal{F}[U(t)] = \frac{e^{i\omega}-1}{i\omega(e^{i\omega}-a)}$, for $Im(\omega) < \log\left|\frac{1}{a}\right|$.

Proof. By using Eq (5) we get

$$\mathcal{F}[a^t] = \frac{1-e^{-i\omega}}{i\omega}\sum_{n=0}^{\infty} a^n e^{-i\omega n} = \frac{1-e^{-i\omega}}{i\omega}\sum_{n=0}^{\infty}(ae^{-i\omega})^n$$

$$= \frac{1-e^{-i\omega}}{i\omega}\frac{1}{1-ae^{-i\omega}} = \frac{e^{i\omega}-1}{i\omega(e^{i\omega}-a)},$$

for $|ae^{-i\omega}| < 1 \Leftrightarrow\ Im(\omega) < \log\left|\frac{1}{a}\right|$.

Lemma 2.1.6. Let $a \in \mathbb{C}$, $a \neq 0$ and $U(t) = ta^t$, then $\mathcal{F}[U(t)] = \frac{a(e^{i\omega}-1)}{i\omega(e^{i\omega}-a)^2}$, for $Im(\omega) < \log\left|\frac{1}{a}\right|$.

Corollary 2.1.7. $\mathcal{F}\left[\frac{1}{a(a-1)}(ta^t - \frac{a}{a-1}(a^t-1))\right] = \frac{1}{i\omega(e^{i\omega}-a)^2}$.

Theorem 2.1.8. If $\widehat{U}(\omega) = \frac{1}{i\omega(e^{i\omega}-a)}$, then $\mathcal{F}^{-1}\left[\widehat{U}(\omega)\right] = U(t) = \frac{a^t-1}{a-1}$, for $Im(\omega) < \log\left|\frac{1}{a}\right|, a \neq 0,1$.

Proof. By definition of inverse Fourier transform, we get

$$\mathcal{F}^{-1}\left[\frac{1}{i\omega(e^{i\omega}-a)}\right] = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{i\omega t}}{i\omega(e^{i\omega}-a)}d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{i\omega(t-1)}}{i\omega(1-ae^{-i\omega})}d\omega$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{i\omega(t-1)}}{i\omega}\sum_{n=0}^{\infty}(ae^{-i\omega})^n d\omega = \sum_{n=0}^{\infty}\frac{1}{2\pi}a^n\int_{-\infty}^{\infty}\frac{e^{i\omega(t-1-n)}}{i\omega}d\omega$$

$$= \sum_{n=0}^{\infty} a^n\mathcal{F}^{-1}\left[\frac{1}{i\omega}\right]_{t=t-1-n} = \sum_{n=0}^{\infty} a^n\begin{cases}1, & if\ 0 \le t < \infty\\ 0, & otherwise\end{cases}|t = t-1-n$$

$$= \sum_{n=0}^{\infty}\begin{cases}a^n, & if\ n+1 \le t < \infty\\ 0, & otherwise\end{cases}$$

$$= \begin{cases}1, & if\ 1 \le t < \infty\\ 0, & otherwise\end{cases} + \begin{cases}a, & if\ 2 \le t < \infty\\ 0, & otherwise\end{cases} + \begin{cases}a^2, & if\ 3 \le t < \infty\\ 0, & otherwise\end{cases}$$

$$+ \dots + \begin{cases}a^n, & if\ n+1 \le t < \infty\\ 0, & otherwise\end{cases} + \dots$$

$$= \begin{cases}1, & if\ 1 \le t < 2\\ 1+a, & if\ 2 \le t < 3\\ 1+a+a^2, & if\ 3 \le t < 4\\ \vdots & \\ 1+a+\dots+a^{n-1}, & if\ n \le t < n+1\\ \vdots & \end{cases}$$

$$= \sum_{n=0}^{\infty}\frac{a^n-1}{a-1}S_n(t) = \frac{a^t-1}{a-1},$$

for $|ae^{-i\omega}| < 1 \Leftrightarrow Im(\omega) < \log\left|\frac{1}{a}\right|, a \neq 0,1$.

Theorem 2.1.9. If $a \neq 0$ $\widehat{U}(\omega) = \frac{1}{i\omega(e^{2i\omega}+a^2)}$, then

$$\mathcal{F}^{-1}\left[\widehat{U}(\omega)\right] = U(t) = \frac{1}{1+a^2}\left(1-|a|^t\left(cos\left(\frac{\pi t}{2}\right)+\frac{1}{|a|}sin\left(\frac{\pi t}{2}\right)\right)\right),$$

$$for\ Im(\omega) < \log\left|\frac{1}{a}\right|.$$

Proof. By definition of inverse Fourier transform, we get

$$\mathcal{F}^{-1}\left[\frac{1}{i\omega(e^{2i\omega}+a^2)}\right] = \frac{1}{2ai}\mathcal{F}^{-1}\left[\frac{1}{i\omega(e^{i\omega}-ai)} - \frac{1}{i\omega(e^{i\omega}+ai)}\right]$$

$$= \frac{1}{2ai}\left(\frac{(ai)^t-1}{ai-1}+\frac{(-ai)^t-1}{ai+1}\right) = \frac{|a|^t}{2ai}\left(\frac{(i)^t}{ai-1}+\frac{(-i)^t}{ai+1}\right)+\frac{1}{2ai}\left(\frac{1}{1-ai}+\frac{1}{1+ai}\right),$$

$$= -\frac{|a|^t}{2ai}\frac{\left((1+ai)e^{\frac{i\pi t}{2}}+(ai-1)e^{-\frac{i\pi t}{2}}\right)}{1+a^2}+\frac{1}{1+a^2},$$

$$= \frac{1}{1+a^2}\left(1-|a|^t\left(cos\left(\frac{\pi t}{2}\right)+\frac{1}{|a|}sin\left(\frac{\pi t}{2}\right)\right)\right),$$

for $|ae^{-i\omega}| < 1 \Leftrightarrow Im(\omega) < \log\left|\frac{1}{a}\right|, a \neq 0,1$.

Fourier Transforms of Polynomials

We have seen above Fourier transform of falling power polynomial

$$t^{\underline{k}} = t(t-1)\dots(t-k+1)$$

can be formulated like $\widehat{U}(\omega) = \frac{k!}{i\omega(e^{i\omega}-1)^k}$. Now if we want to evaluate Fourier transform of any polynomial we might face complex expressions, and it is better to avoid that complexity, if we write a polynomial in the form of a polynomial of falling powers, we can find the Fourier transform more easily and without any confusion. For example instead of $U(t) = t^2$ we can rewrite this function as $U(t) = t(t\text{-}1) + t$ and by taking Fourier transform we get:

$$\mathcal{F}[t^2] = \mathcal{F}[t(t-1)+t] = \frac{2!}{i\omega(e^{i\omega}-1)^2}+\frac{1}{i\omega(e^{i\omega}-1)}.$$

In this case, the problem arises of expressing a polynomial in terms of a polynomial of falling powers.

We want to find a_k such that

$$t^k = a_1 t^{\underline{1}} + a_2 t^{\underline{2}} + \cdots + a_k t^{\underline{k}}, \tag{9}$$

here some powers of t in terms falling powers polynomial:

$$t = t$$
$$t^2 = t + t(t-1)$$
$$t^3 = t + 3t(t-1) + t(t-1)(t-2)$$
$$t^4 = t + 7t(t-1) + 6t(t-1)(t-2) + t(t-1)(t-2)(t-3)$$

If we continue like this, we get a triangle of corresponding coefficients:

$$S = \left(S_{i,j}\right)_{i,j\geq 1} = \begin{pmatrix} 1 & & & & & & \\ 1 & 1 & & & & & \\ 1 & 3 & 1 & & & & \\ 1 & 7 & 6 & 1 & & & \\ 1 & 15 & 25 & 10 & 1 & & \\ 1 & 31 & 90 & 65 & 15 & 1 & \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \ddots \end{pmatrix}$$

Coefficient matrix S is lower triangular matrix that k^{th} row elements gives coefficients of falling power polynomials in the expansion of polynomial t^k. We can formulate this matrix by relation

$$S_{ij} = \begin{cases} 1, if\ i \geq 1, j = 1 \\ 1, if\ i = j \\ 0, if\ i < j \\ S_{i-1,j-1} + j \cdot S_{i-1,j}, if\ i \geq 3, j > 1 \end{cases}.$$

Applications of Fourier Transform Method

In this section, some examples are presented to show the efficiency and validity of the proposed method.

Example 3.1. Consider the following linear of 2^{nd} order difference equation

$$\Delta^2 U_n + 6\Delta U_n + 5U_n = 3n + 7, \tag{10}$$

with $U(0) = U(1) = 0$

We can rewrite equation (10) as follows

$$U(t+2) + 4U((t+1) = 3t + 7. \tag{11}$$

By applying Fourier transform of (11), gives

$$e^{2i\omega}\widehat{U} + 4e^{i\omega}\widehat{U} = \frac{3}{i\omega(e^{i\omega}-1)} + \frac{7}{i\omega}.$$

This can be reduced to

$$\widehat{U} = \frac{3}{i\omega(e^{i\omega}-1)(e^{i\omega}+4)e^{i\omega}} + \frac{7}{i\omega(e^{i\omega}+4)e^{i\omega}}.$$

$\widehat{U}$ can be represented as a partial fraction expansion as follows

$$\begin{aligned} \widehat{U} &= \frac{3}{20}\left[\frac{4}{i\omega(e^{i\omega}-1)} + \frac{1}{i\omega(e^{i\omega}+4)} - \frac{5}{i\omega e^{i\omega}}\right] \\ &\quad + \frac{7}{4}\left[\frac{1}{i\omega e^{i\omega}} - \frac{1}{i\omega(e^{i\omega}+4)}\right], \\ &= \frac{12}{20}\frac{1}{i\omega(e^{i\omega}-1)} - \frac{8}{5}\frac{1}{i\omega(e^{i\omega}+4)} + \frac{1}{i\omega e^{i\omega}}. \end{aligned} \tag{12}$$

By applying inverse Fourier transform of (12), gives

$$\begin{aligned} U(t) &= \frac{3}{5}t - \frac{8}{5}\left[\frac{(-4)^t - 1}{-5}\right] + H(t-1) \\ &= \begin{cases} 0, & t = 0 \\ \frac{3}{5}t - \frac{8}{25}(-4)^t + \frac{17}{25}, & 1 \leq t < \infty \end{cases}. \end{aligned}$$

Where $H(t)$ is Heaviside step function.

Example 3.2. Consider the following linear of 1^{st} order difference equation

$$\Delta U_n = n^2 + n - 2, \tag{13}$$

with the condition $U(0) = 0$.

We can rewrite equation (13) as follows

$$U(t+1) - U((t) = t^2 + t - 2, \tag{14}$$

by the formula (9), we can write equation (14) as follows

$$U(t+1) - U((t) = t^2 - t + 2t - 2, \tag{15}$$

By applying Fourier transform of (15), gives

$$e^{i\omega}\widehat{U} - \widehat{U} = \frac{2}{i\omega(e^{i\omega}-1)^2} + \frac{2}{i\omega(e^{i\omega}-1)} - \frac{2}{i\omega}.$$

This can be reduced to

$$\widehat{U} = 2\left[\frac{1}{i\omega(e^{i\omega}-1)^3} + \frac{1}{i\omega(e^{i\omega}-1)^2} - \frac{1}{i\omega(e^{i\omega}-1)}\right]. \tag{16}$$

By applying inverse Fourier transform of (16), gives

$$U(t) = \frac{2}{3!}\big(t(t-1)(t-2)\big) + \frac{2}{2!}\big(t(t-1)\big) - \frac{2}{1!}(t) = \frac{1}{3}(t^3 - 7t).$$

Example 3.3. Consider the following linear of 3^{rd} order difference equation

$$\Delta^3 U_n + 2\Delta^2 U_n + 5\Delta U_n = 2^n, \tag{17}$$

with the conditions $U(0) = 0$, $U(1) = 1$, $U(2) = 1$.

We can rewrite equation (17) as follows

$$U(t+3)-U(t+2)+4U(t+1)-4U(t)=2^t. \quad (18)$$

By applying Fourier transform of (18), gives

$$e^{3i\omega}\widehat{U}-\hat{S}_0(e^{2i\omega}+e^{i\omega})-\left(e^{2i\omega}\widehat{U}-\hat{S}_0e^{i\omega}\right)+4e^{i\omega}\widehat{U}-4\widehat{U}=\frac{1}{i\omega(e^{i\omega}-2)},$$

So

$$\widehat{U}=\frac{1}{i\omega(e^{i\omega}-2)(e^{3i\omega}-e^{2i\omega}+4e^{i\omega}-4)}+\frac{i\omega(e^{i\omega}-1)}{i\omega(e^{3i\omega}-e^{2i\omega}+4e^{i\omega}-4)}.$$

$\widehat{U}$ can be represented as a partial fraction expansion as follows

$$\widehat{U}=\frac{1/8}{i\omega(e^{i\omega}-2)}-\frac{1/5}{i\omega(e^{i\omega}-1)}+\frac{(43/40)e^{i\omega}}{i\omega(e^{2i\omega}+4)}-\frac{1/20}{i\omega(e^{i\omega}+4)}.$$

By applying inverse Fourier transform of (19), gives

$$U(t)=\frac{1}{25}-\frac{t}{5}+2^t\left(\frac{1}{8}+\frac{91}{200}sin\left(\frac{\pi t}{2}\right)-\frac{33}{200}cos\left(\frac{\pi t}{2}\right)\right).$$

Example 3.4. Consider the following linear of 4^{th} order difference equation

$$\Delta^4U_n-5\Delta^3U_n+8\Delta^2U_n-4\Delta U_n=n^4, \quad (20)$$

with the conditions $U(0) = 1$, $U(1) = -1$, $U(2) = 2$, $U(3) = 0$.
We can rewrite equation (20) as follows

$$U(t+4)-9U(t+3)+29U(t+2)-39U(t+1)+18U(t)=t^4. \quad (21)$$

By the formula (9), we can write equation (21) as follows

$$U(t+4)-9U(t+3)+29U(t+2)-39U(t+1)+18U(t)=t+7t(t-1)+6t(t-1)(t-2)+t(t-1)(t-2)(t-3) \quad (22)$$

By applying Fourier transform of (22), gives

$$e^{4i\omega}\widehat{U}-\hat{S}_0(e^{4i\omega}-e^{3i\omega}+2e^{2i\omega})-9e^{3i\omega}\widehat{U}+9\hat{S}_0(e^{3i\omega}-e^{2i\omega}+2e^{i\omega})+29e^{2i\omega}\widehat{U}-29\hat{S}_0(e^{2i\omega}-e^{i\omega})-39e^{i\omega}\widehat{U}+39\hat{S}_0(e^{i\omega})+18\widehat{U}=\frac{1}{i\omega(e^{i\omega}-1)}+7\frac{2!}{i\omega(e^{i\omega}-1)^2}+6\frac{3!}{i\omega(e^{i\omega}-1)^3}+\frac{4!}{i\omega(e^{i\omega}-1)^4},$$

So

$$\widehat{U}=\frac{1}{i\omega(e^{i\omega}-1)\,(e^{i\omega}-2)(e^{i\omega}-3)^2}\left[\frac{1}{i\omega(e^{i\omega}-1)}+\frac{14}{i\omega(e^{i\omega}-1)^2}+\frac{36}{i\omega(e^{i\omega}-1)^3}+\frac{24}{i\omega(e^{i\omega}-1)^4}+\frac{1-e^{-i\omega}}{i\omega}(e^{4i\omega}-10e^{3i\omega}+40e^{2i\omega}-86e^{i\omega})\right],$$

and this is just rational function in terms of $e^{i\omega}$, except $\frac{1}{i\omega}$ factor. so by partial fractions we get:

$$\widehat{U}=\frac{1}{i\omega}\left(\frac{75}{e^{i\omega}-2}-\frac{243}{4(e^{i\omega}-1)}-\frac{103}{2(e^{i\omega}-1)^2}-\frac{38}{(e^{i\omega}-1)^3}-\frac{21}{(e^{i\omega}-1)^4}-\frac{6}{(e^{i\omega}-1)^5}-\frac{57}{4(e^{i\omega}-3)}+\frac{5}{(e^{i\omega}-3)^2}+(e^{i\omega}-1)\left[\frac{-38}{e^{i\omega}-2}+\frac{55}{4(e^{i\omega}-1)}+\frac{101}{4(e^{i\omega}-3)}-\frac{29}{2(e^{i\omega}-3)^2}\right]\right). \quad (23)$$

By applying inverse Fourier transform of (23), gives

$$U(t)=\frac{-1269}{4!}-\frac{243}{4}t-\frac{103}{4}t^{\underline{2}}-\frac{38}{6}t^{\underline{3}}-\frac{21}{24}t^{\underline{4}}-\frac{1}{20}t^{\underline{5}}+37(2^t)+\frac{405}{4!}(3^t)-\frac{41}{3}t(3^t).$$

Example 3.5. Consider the Fibonacci sequence defined by difference equation

$$\Delta^2U_n+\Delta U_n-U_n=0, U(0)=U(1)=1. \quad (24)$$

We can rewrite equation (24) as follows

$$U(t+2)-U(t+1)-U(t)=0 \quad (25)$$

By applying Fourier transform of (25), gives

$$e^{2i\omega}\widehat{U}-\hat{S}_0(e^{2i\omega}-e^{i\omega})-e^{i\omega}\widehat{U}+\hat{S}_0e^{i\omega}-\widehat{U}=0,$$

so that

$$\widehat{U}(e^{2i\omega}-e^{i\omega}-1)=\frac{e^{2i\omega}-e^{i\omega}}{i\omega}.$$

This can be reduced to

$$\widehat{U}=\frac{e^{i\omega}(e^{i\omega}-1)}{i\omega(e^{2i\omega}-e^{i\omega}-1)}.$$

$\widehat{U}$ can be represented as a partial fraction expansion as follows

$$\widehat{U} = \frac{(e^{i\omega}-1)}{i\omega}\left(\frac{\frac{\alpha}{\alpha-\beta}}{e^{i\omega}-\alpha} - \frac{\frac{\beta}{\alpha-\beta}}{e^{i\omega}-\beta}\right), \qquad (26)$$

where $\alpha = \frac{1+\sqrt{5}}{2}$ is the golden ratio, and $\beta = \frac{1-\sqrt{5}}{2}$ is the silver ratio.

By applying inverse Fourier transform of (26), gives

$$U(t) = \frac{1}{\alpha-\beta}(\alpha\cdot\alpha^t - \beta\cdot\beta^t) = \frac{1}{\sqrt{5}}\frac{(1+\sqrt{5})^{t+1}-(1-\sqrt{5})^{t+1}}{2^{t+1}}.$$

CONCLUSION

In this study, a new accurate technique (Fourier Transform Method) was applied to solve the linear of m^{th}-order difference equation with constant coefficients, also successfully overcoming the finding of the Fourier transform of a polynomial function. In addition, we gave five examples. One of them is the difference equations from the Fibonacci sequence.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Kelley W. G, Peterson A. C. Difference equations: an introduction with applications. Cambridge, Massachusetts: Academic Press; 2001.

[2] Gupta R. C. On particular solutions of linear difference equations with constant coefficients. SIAM Rev 1998;40. [CrossRef]

[3] Agarwal R. P. Difference equations and inequalities: theory, methods, and applications. Boca Raton, Florida: CRC Press; 2000. [CrossRef]

[4] Avezov S, Düz M, Issa A. Solutions to differential-difference equations with variable coefficients by using fourier transform method. Süleyman Demirel Univ Fac Arts Sci J Sci 2023;18:259-267. [CrossRef]

[5] Berinde V. A Method for Solving Second Order Difference Equations. In Cheng S, (editor). New Developments in Difference Equations and Application. London: Routledge; 2017. p. 41-48. [CrossRef]

[6] Feldmann L. On linear difference equations with constant coefficients. Period Polytech Electr Eng 1959;3:247-257.

[7] Fort T. Linear difference equations and the Dirichlet series transform. Amer Math Monthly 1955;62:641-645. [CrossRef]

[8] Olver F. W. Numerical solution of second-order linear difference equations. J Res Nat Bureau Stand 1967;71:111-129. [CrossRef]

[9] Arikoglu A, Ozkol I. Solution of difference equations by using differential transform method. Appl Math Comput 2006;174:1216-1228. [CrossRef]

[10] Hatipoglu V. F. Taylor polynomial solution of difference equation with constant coefficients via time scales calculus. New Trends Math Sci 2015;3:129.

[11] Gencev M, Salounova D. First-and second-order linear difference equations with constant coefficients: suggestions for making the theory more accessible. Int J Math Educ Sci Technol 2022;54:1349-1372. [CrossRef]

[12] Gupta RC. On linear difference equations with constant coefficients: An alternative to the method of undetermined coefficients. Math Mag 1994;67:131-135. [CrossRef]

[13] Rivera-Figueroa A, Rivera-Rebolledo J. M. A new method to solve the second-order linear difference equations with constant coefficients. Int J Math Educ Sci Technol 2016;47:636-649. [CrossRef]

[14] Rivera-Figueroa A, Rivera-Rebolledo JM. A response to Tisdell on 'Critical perspectives of the "new" difference equation solution method'of Rivera-Figueroa and Rivera-Rebolledo. Int J Math Educ Sci Technol 2020;51:150-151. [CrossRef]

[15] Tisdell C.C. Critical perspectives of the 'new'difference equation solution method of Rivera-Figueroa and Rivera-Rebolledo. Int J Math Educ Sci Technol 2019;50:160-163. [CrossRef]

[16] Düz M, Issa A, Avezov S. A new computational technique for Fourier transforms by using the differential transformation method. Bull Inter Math Virtual Inst 2022;12:287-295.

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00099

Research Article

Evaluation of energy consumption and noise reduction change of a strengthened building: An educational building case

Seda YÜKSEL DİCLE[1,*], Fatma ZOROĞLU[2], Ahmet Bircan ATMACA[2], Neşe YÜĞRÜK AKDAĞ[2], Gülay ZORER GEDİK[2]

[1]*Institute for Environmental Design and Engineering, University College London, 14 Upper Woburn place, London, WC1H0NN, UK*
[2]*Building Physics Research Group, Department of Architecture, Yildiz Technical University, Istanbul 34349, Türkiye*

ARTICLE INFO

Article history
Received: 17 November 2022
Revised: 17 December 2022
Accepted: 13 May 2023

Keywords:
Classroom; Educational Building; Energy Efficient Design; Noise Reduction Level; Strengthened Buildings

ABSTRACT

The buildings in various geographical conditions are strengthened due to different reasons, such as earthquakes, fire, or deterioration. One of the important issues is that comfort levels are ensured, and energy consumption is kept at a minimum in educational buildings where renovation and strengthening work is being carried out. In this study, in terms of energy consumption and acoustic comfort, a method is proposed to examine a building damaged in an earthquake and then strengthened. The method has also been created and implemented to provide control and improvement suggestions in terms of thermal and acoustic conditions in strengthening projects of buildings. As a case study, the pre- and post-reinforcement situation of an educational building in a temperate humid climate has been evaluated. In addition, different glazing scenarios that reduce energy consumption while increasing facade noise reduction have been developed and analyzed. The results showed that while sufficient noise reduction is achieved in the building envelope, there is a 15.9% reduction in the total energy consumption of the building. The optimum scenario decreased total energy consumption by 19.4 % in B113 and 26 % in B114 and increased the facade noise reduction levels by 14.2 dB in B113 and 15 dB in B114. The proposed method and the findings will contribute to the design process of newly designed, renovated, and strengthened buildings in terms of energy efficiency and indoor acoustical comfort.

Cite this article as: Yüksel Dicle S, Zoroğlu F, Atmaca AB, Yüğrük Akdağ N, Zorer Gedik G. Evaluation of energy consumption and noise reduction change of a strengthened building: An educational building case. Sigma J Eng Nat Sci 2024;42(4):1245–1260.

INTRODUCTION

Buildings protect their occupants from the negative effects of the external environment. In developed countries, people spend 90% of their lives in buildings. For this reason, indoor comfort conditions (heat, sound, and light) which significantly affect the productivity and health of

***Corresponding author.**
*E-mail address: seda.dicle.22@ucl.ac.uk

This paper was recommended for publication in revised form by Editor in-Chief Ahmet Selim Dalkilic

occupants in the building, are important in terms of the efficiency of the activity performed [1]. In addition, in recent years, there has been an increase in people's awareness of the effects of indoor comfort conditions on human health and well-being [2].

Indoor comfort conditions in educational buildings affect students' physical and mental health and learning efficiency. For educational activities, the students spend about 30% of their daily lives indoors, therefore it is important to provide appropriate conditions in terms of light, sound, and air in the buildings [3,4]. The educational buildings consist of units such as classrooms, laboratories, conference rooms, and office rooms. The basic units of educational buildings are classrooms. The efficient and healthy progress of the teaching-learning activities depends on ensuring the required environmental conditions in the classroom [5,6]. Ensuring acoustic comfort conditions in educational buildings and classrooms constitute an important design criteria in order to continue educational activities without interruption and under appropriate conditions [7,8]. In addition, the priority of educational buildings is to create quality learning conditions, while minimizing unnecessary energy consumption [9].

Nowadays, energy usage should be kept at a minimum while providing comfort conditions in buildings because of the rapid consumption of energy resources. In many developed countries, buildings account for around 35-40% of total energy consumption [10] 65% of the energy consumed in buildings is used in heating-cooling, ventilation, and hot water systems [11].

Comfort conditions and energy consumption have lost their importance during renovations. The strengthening activities in buildings can be defined as building development, upgrades, and modifications according to user requirements and legal expectations [12]. The renovations of existing buildings are generally related to the age of the building, strengthening of the structural system, changing its function, its adaptation to the current architecture, etc. The strengthening of the structural system is often used in Turkey, which is located in the Alpine-Himalayan active earthquake zone. The educational buildings are one of the building typologies damaged in earthquakes. According to the 2020 data of the Ministry of National Education, Republic of Turkey, the number of schools used by about 21% of the country's population is 68.589 [13]. About 50% of the buildings are over 25 years and were built according to the architectural features, construction standards, and old earthquake regulations [14]. For this reason, many educational buildings may be affected by the earthquake, and it becomes necessary to strengthen them in order to continue education after the earthquake. While strengthening the educational buildings, attention needs to be dedicated to the effects of the units on the physical environmental comfort conditions. Renovation activities due to strengthening works can cause significant changes on building facades that directly and indirectly affect indoor comfort conditions. Indoor environmental quality in buildings is evaluated in terms of thermal comfort, indoor air quality, noise level, and illumination level for users [15].

The design of the building facade, which is one of the important design decisions in buildings, directly affects both energy consumption and acoustic comfort conditions. In a previous study, it was determined that the strengthening carried out in the facade of educational buildings affects the visual comfort of the indoor environments [16]. When the properties of transparent areas change, acoustic and thermal comfort issues can arise in addition to daylighting issues. Therefore, besides daylight, factors related to heat and noise should also be taken into account in transparent space design decisions[14]. In order to provide acoustic comfort conditions in the buildings, some measures are taken against noise in the building envelope. These measures can be exemplified as the use of double walls, using materials with high noise reduction in transparent and opaque areas, and determining transparency ratios [17].

Although design guidelines and technical standards focus separately on a single environmental factor, the changes made to the facade of the building affect the thermal and acoustic comfort conditions with the energy consumption together [9,18]. In the literature, there are many studies on the effect of changing the transparent area properties of buildings' comfort conditions. The windows are considered to be the weakest element of the facade in terms of thermal and acoustical aspects. The effect of transparent areas on both energy consumption and noise control is greater than that of opaque areas [19]. Therefore, the design strategies that focus on improvements in transparent area properties are considered important in terms of thermal and acoustical properties [20,21].

However, few studies have been conducted to investigate energy consumption while providing acceptable thermal and acoustic comfort in relation to the transparent area properties of educational buildings. In studies, the effect of various glass types and different Window to Wall Ratios (WWR) on building energy performance has been investigated [22–24]. The studies investigating the effect of changes in the WWR on noise control in educational buildings are very limited. They are generally aimed at measuring the current situation or determining the acoustical comfort of the occupants [25–27]. There are few research methods or studies on the changes in energy consumption and noise control after strengthening educational buildings. The changing environmental conditions, energy consumption and acoustic comfort conditions should be taken into account in the implementation of strengthening projects of buildings.

Suitable indoor conditions for educational buildings should be provided to improve productivity, reach a healthy environment, etc., while providing energy efficiency. For this reason, even after strengthening, it is important that they protect and/or improve the performance of the buildings in terms of noise control and energy consumption.

This study has investigated the effect of the change in the Window to Wall Ratio (WWR) on noise control and energy consumption after the strengthening in classrooms of a sample building. In addition, the effects of transparent area change have been evaluated by developing suggestions aiming to provide comfort conditions with minimum energy consumption.

METHODOLOGY

In this study, an educational building, that is strengthened, has been determined as a sample building. It is determined by measurements in the previous studies that the current state of the educational building is not at an acceptable level of acoustical, visual, and thermal comfort according to the relevant regulations [16,28]. An educational building was evaluated in terms of facade noise reduction and energy consumption for before and after strengthened situations including different transparent area scenarios. The facade noise reduction evaluations were made for selected classrooms, while the changes of energy consumption amounts were made for both classrooms and the whole building.

The method steps of the study are shown in Figure 1. The proposed method for the study consists of six parts. The first stage is the determination of the strengthened building and its properties. The information was obtained about the environmental, physical, and indoor properties of the building. In order to determine the noise reduction of the building envelope, the opaque and transparent area properties of the facade were defined. Also, the thermal properties of the building envelope were defined to determine total energy consumption. Then, with the information obtained during the determination and identification stages, the analysis was done to determine the effect of the situation before and after the strengthening of the building facade. After this stage, building facade scenarios were suggested for situations where acceptable limit values for the comfort conditions cannot be achieved. The suggested scenarios were developed for transparent areas that have weaker properties in providing comfort conditions on facades.

The scenarios that provided acceptable values in terms of thermal and acoustic performance and had different transparent area properties were evaluated and the most appropriate solution was selected by comparing the energy consumption ratios of the scenarios. The evaluations of

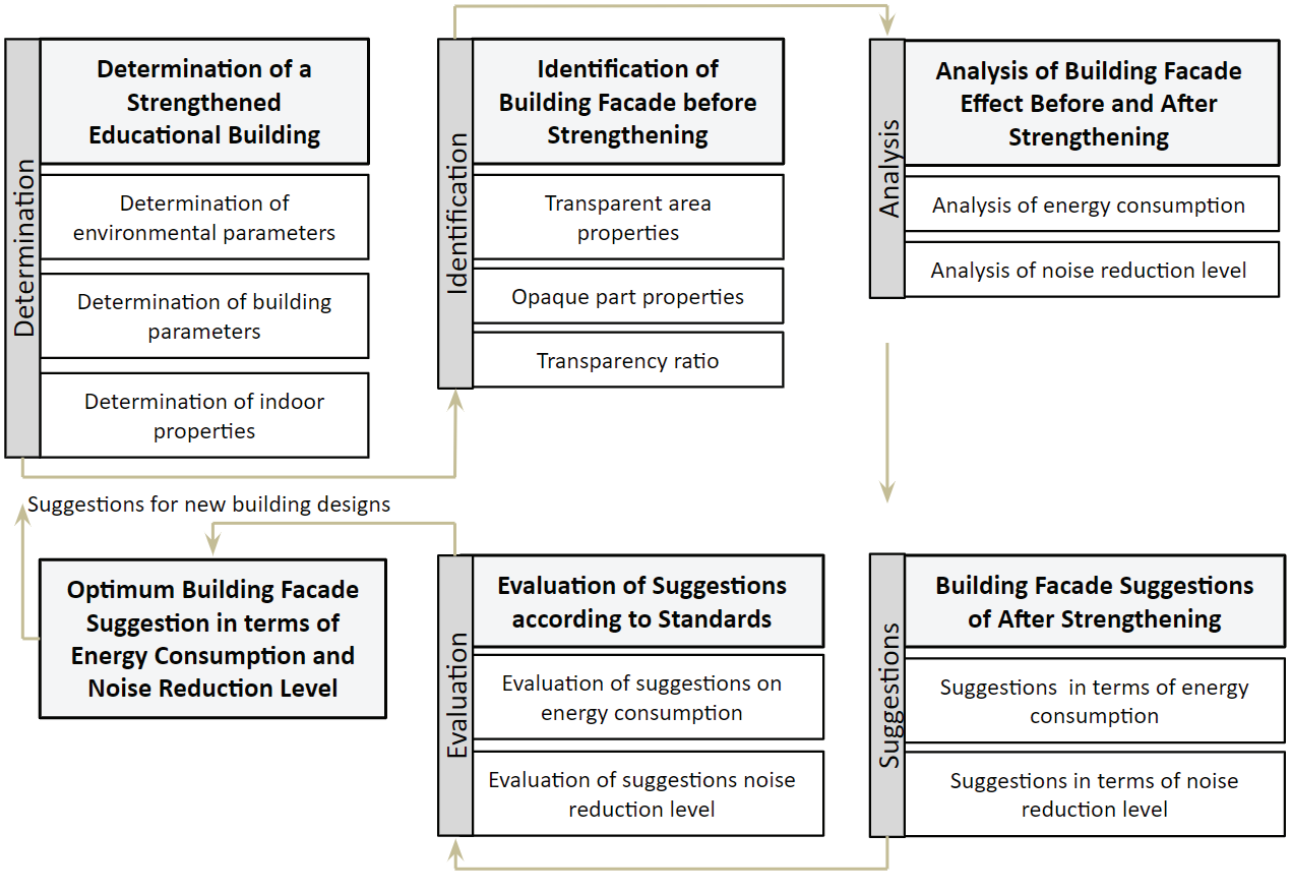


Figure 1. Evaluation stages of the strengthened educational building in terms of energy consumption and noise reduction level.

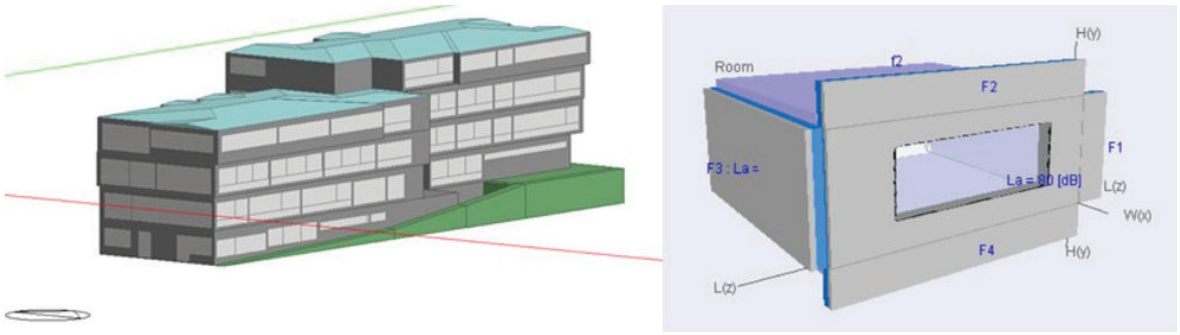


Figure 2. The sample building and classroom visuals designed with Design Builder and Kalksandstein Schallschutzrechner (KS)

acoustical and thermal comfort conditions were made according to international standards, and the results of noise reduction and energy consumption were compared for scenarios that met the conditions. Finally, the optimum facade designs were recommended, providing minimum energy consumption and acceptable noise reduction from the suggested scenarios for transparent areas. The stages of the method mentioned can also be implemented for future studies and different building types.

In the sample building, which was found to have insufficient comfort conditions in previous studies, evaluations were made for the situation before and after strengthening, and scenarios were created to provide appropriate thermal comfort and acoustic comfort. The scenarios for the energy consumption and facade noise reduction were determined through the simulation programs. The energy consumption analyses of the educational building and selected classrooms were made in the Design Builder program. Design Builder is a simulation program that uses the Energy Plus simulation engine to determine thermal and visual comfort conditions and to calculate energy loads [29]. The facade noise reduction analyses of the classrooms were made in the Kalksandstein Schallschutzrechner (KS) program [30]. The KS program calculates in accordance with the 12354-3 ''Building acoustics - Estimation of acoustic performance of buildings from the performance of elements" standard [31]. Figure 2 shows the three-dimensional model of the sample building and classroom.

The Properties of Educational Building and Sample Classrooms

The educational building included in the study is on Barbaros Avenue in Beşiktaş, Istanbul, which is exposed to high-level of highway noise. The building is located in an area with temperate humid climate conditions. Figure 3 shows the site plan and view of the educational building. The educational building has five floors including a basement. The building has 26 classrooms, six laboratories, one conference hall, one music workshop, one painting workshop, one dining hall, one archive, and five WC areas. 534 students study between 08:15-15:20 on weekdays [13]. The classrooms chosen for this study are on the east side of the educational building and on the 1st floor. Figure 4 shows the location of the sample classrooms on the floor plan.

The classrooms have capacity of 24 students. They are 7.0 m wide, 8.2 m long and 3.44 m high. The floors in the classrooms are covered with resin-based linoleum. Plastic paint is used on the ceiling and walls. Transparent areas have 6 mm Low-e double glass with a 16 mm air gap, and the parapet height of the windows is 84 cm. An aluminum

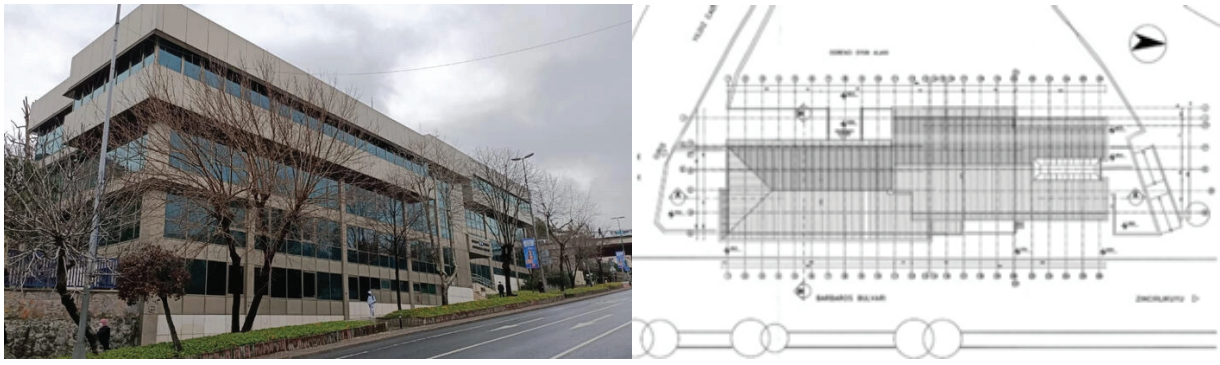


Figure 3. The view of the educational building from Barbaros avenue and site plan.

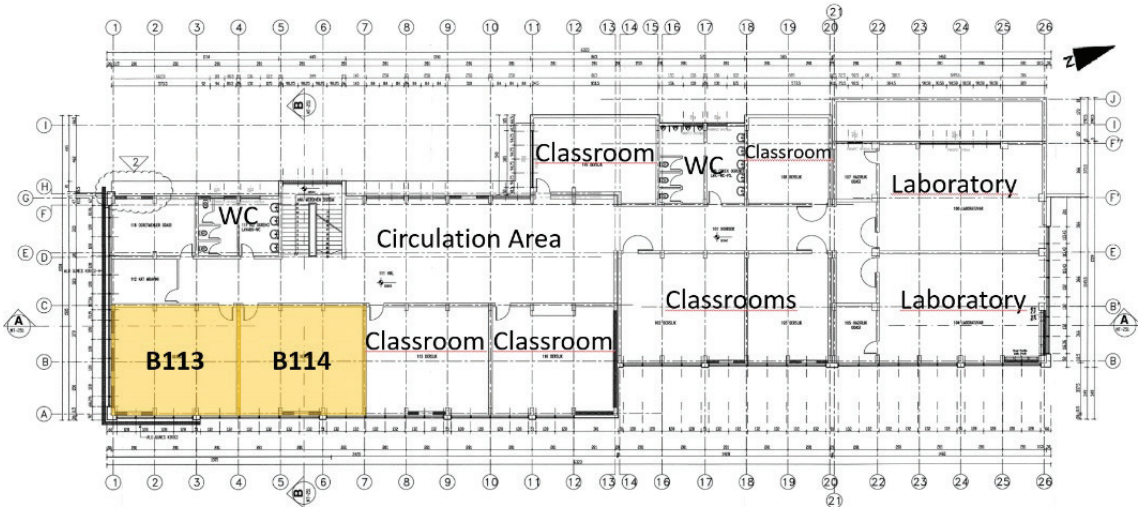


Figure 4. Floor plan and classrooms B113, and B114.

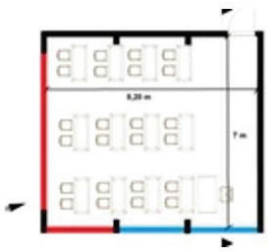
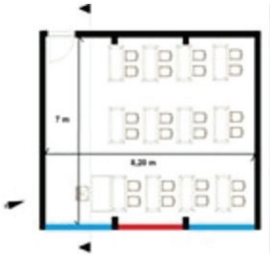
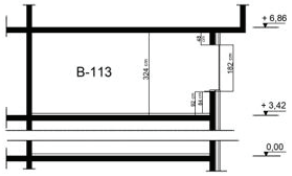
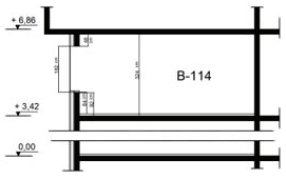
Table 1. The current facade layers properties of the educational building

Building Components	Layers	Thickness (m)	Thermal Conductivity (W/mK)	Overall Heat Transfer Coefficient-OHTC (W/m^2K)	Density (kg/m^3)
Wall	Aluminum	0.02	160.0	0.423	2800
	Air Gap	0.06	0.30		-
	Thermal Insulation	0.06	0.034		35
	Cement Plaster	0.03	0.72		1760
	R. Concrete	0.2	1.13		2000
	Gypsum Plaster	0.25	0.80		1300
Floor	Linoleum	0.002	0.30	1.684	1000
	Tile	0.03	1.10		2100
	Mortar	0.015	1.40		2100
	Leveling concrete	0.035	0.88		2800
	R. Concrete Floor	0.2	1.13		2000
	Gypsum Plaster	0.025	0.40		1000
Ceiling	Linoleum	0.002	0.30	1.684	1000
	Tile	0.03	1.10		2100
	Mortar	0.015	1.40		2100
	Leveling Concrete	0.035	0.88		2800
	R. Concrete Floor	0.2	1.13		2000
	Gypsum Plaster	0.025	0.40		1000
Indoor wall	Gypsum Plaster	0.02	0.40	1.603	1000
	Brick	0.19	0.72		1920
	Gypsum Plaster	0.02	0.40		1000
Window	Sunguard Low-e Glass	0.006	0.94	OHTC 1.842	-
	Air Gap	0.016	-	SHGC 0.164	
	Green Tinted Low-e Glass	0.006	1.00	LT 0.121	

cladding wall is used in the opaque areas of the building facade. Opaque areas have 5 cm thermal insulation. Table 1 shows layer details of the selected classrooms.

The educational building, which was built in 1992, was strengthened in 2008 and reached its current situation. Table 2 shows the plans and sections of the classrooms in 1992 and 2008. During the strengthening made in 2008, the entire transparent area on the south facade of classroom B113 was covered with a concrete curtain wall. On the eastern facade, the transparent area decreased by triple. There

Table 2. Plan and sections of sample classrooms (The red lines of the transparent areas show the surfaces that were closed during strengthening, and the blue lines of the transparent areas show the current transparent areas - not closed)

		Classroom B113	Classroom B114	Classroom Section
1992-2008 WWR Change				
WWR	1992	South-East: 46.90% South-West: 60.25%	South-East: 46.90%	
	2008	South-East: 31.6%	South-East: 31.6%	

is a similar situation in classroom B114. The transparent area on the eastern facade of classroom B114 was covered by a concrete curtain wall by ⅓. This situation also exists on other floors and classrooms. The effect of the closed transparent areas of the classrooms on energy consumption and the facade noise reduction was calculated with simulation programs.

Definitions Regarding Analysis Variable

The consumed heating, cooling, and lighting loads and the total energy load were analyzed in order to provide thermal comfort conditions in the educational building and the selected classrooms. In addition, the noise reduction value was analyzed in order to provide the required acoustic comfort conditions in classrooms. While the energy consumption analyses were made for the whole building and selected classrooms, the facade noise reduction evaluations were made for the classrooms only.

In the study, outdoor climate data for Istanbul were taken from the DesignBuilder - Energyplus program [32]. During the education period, the air conditioning systems in the educational building are used between 08:15 and 16:00 on weekdays. They are not used during the summer period. The classrooms are heated by radiators with a boiler heating water system during the heating period. They are cooled with split air conditioners during the cooling period. The indoor temperature was determined as 21 °C during the heating period and 26 °C during the cooling period in the simulation program. The illumination level was set as 300 lux which is specified in the standard for the lighting of the classrooms. The default lighting system recommended for educational buildings was chosen in the program [33].

Outdoor noise levels were taken from the current noise maps created by Istanbul Metropolitan Municipality Environmental Protection Directorate [34]. According to the noise map, the noise level on the classroom facades was accepted as 80 dBA. It is recommended in the Regulation on Protection of Buildings against Noise that existing buildings should meet the minimum D acoustic performance class limit value. The classrooms in educational buildings are determined as 1st-degree sensitive volumes in the regulation. The required limit value is given as "Lden-26" according to the degree of sensitivity and the D acoustic performance class in the regulation [35]. Lden states the day-evening-night noise level. It is a descriptor of noise level based on energy equivalent noise level (Leq) over a whole day. According to the current regulation and the outdoor noise determined as 80 dBA, the minimum noise reduction level that the facade should provide is DnT,A,tr (weighted standardized level difference) 54 dB (80-26 = 54 dB).

RESULTS

Analysis of Energy Consumption and Noise Reduction of the Facade Before and After Strengthening of the Building

The energy consumption and the facade noise reduction were evaluated according to the changes before strengthening the educational building. The building's opaque and transparent component layers and properties of the facade layers remained unchanged during the structural system's strengthening, but WWR of the facades was decreased in the south and west directions. After the strengthening, the transparent area was completely closed on the south facade of classroom B113, and the WWR was reduced by 15.3% on the eastern facade of B113 and B114 classrooms. Figure 5 shows the heating, cooling, lighting, and total energy

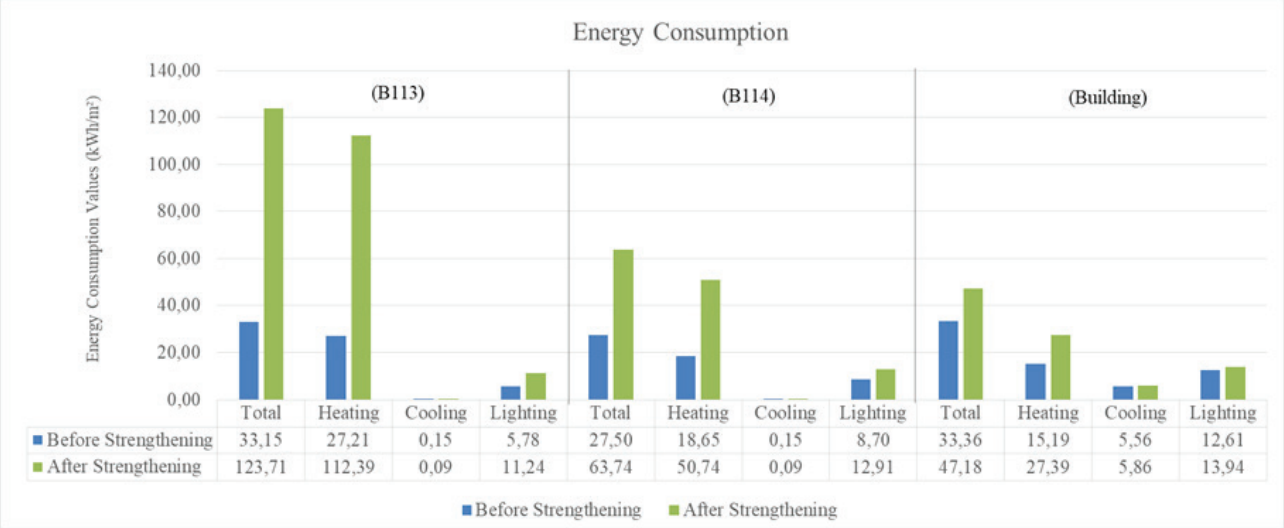


Figure 5. Energy consumption of before and after strengthening of the educational building.

consumption. Figure 6 shows the facade noise reduction levels of B113 and B114 classrooms, and also the whole building before and after strengthening.

As a result of completely closing the transparent areas on the south facade and reducing them on the east facade of the classrooms, solar heat energy gains decreased greatly. For this reason, the heating energy consumption of the classrooms increased by 313% in B113 and by 172% in B114. In addition, the total energy consumption increased by 273% in the B113 and 131% in the B114. Although the WWR change on the east facade was the same in both classrooms, the entire transparent area on the south facade was also covered in B113. Therefore, the proportional increase in heating, lighting, and total energy consumption in B113 was higher than in B114. However, no significant difference was observed in cooling energy consumption.

It was determined that there was an 80% increase in heating energy consumption, a 10% increase in lighting energy consumption, and a 5% increase in cooling energy consumption in the whole building. There was an increase of 41% in the total energy consumption of the building used for heating, cooling and lighting. Accordingly, it was observed that the change in the transparency ratio after strengthening greatly affected the energy consumption ratio and that energy consumption was saved by utilizing solar heat energy before strengthening.

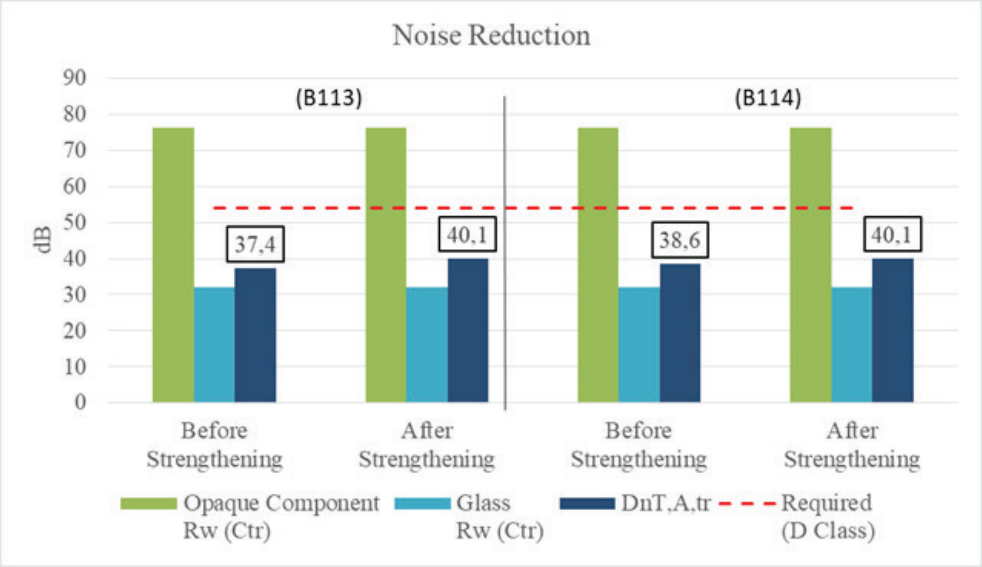


Figure 6. Facade noise reduction levels of classrooms B113 and B114.

As a result of reducing the transparent areas on the facades of the classrooms, the facade noise reduction levels increased in both classrooms with different amounts. There was an increase of 2.7 dB for classroom B113 and 1.5 dB for classroom B114 in facade noise reduction level due to the decrease in WWR, which varies at different ratios depending on the location of the classrooms. After strengthening, facade noise reduction levels increased to 40.1 dB in the B113 and B114 classrooms due to the fact that the transparent area ratios after strengthening were the same. The decrease in the transparency ratio caused the facade noise reduction level to increase. However, this increased ratio could not exceed 1.5-3 dB levels due to the insufficient insulation properties of the transparent area.

The transparency ratio, which has a significant effect on the noise reduction level of the facade, was less effective when the insulation properties of the weak facade element were insufficient. It was observed that the increase in the sound reduction level after the decrease in the transparency ratio will only provide acceptable conditions with the transparent area improvements.

Improvement Scenarios of Transparent Area in Educational Building

In this study, it was determined that the indoor noise level after strengthening did not meet the requirements of the Regulation on Protection of Buildings Against Noise, which is in force in Turkey and determined according to COST Action TU 0901 'Integrating and Harmonizing Sound Insulation Aspects in Sustainable Urban Housing Constructions' [35,36]. The transparent areas have a greater impact on the facade noise reduction and the thermal load than the opaque areas. Different transparent area scenarios were developed for the current situation, since the noise reduction level of the transparent areas did not meet the requirements given in the regulation and the energy consumption of the building was high after strengthening. The scenarios were created according to the fact that the transparent area sections have different layer properties and layer numbers are selected from the most commonly used glass in the construction market.

In order to provide the required noise reduction level on the building facades consisting of opaque and transparent components, noise control is required through the composite wall. The noise reduction level of the composite wall varies according to the sound insulation of the opaque and transparent components and the transparency ratio. Although the sound insulation level of the opaque component is high and the transparency ratio is low in the educational building, the required noise reduction level cannot be achieved due to the low sound insulation level of the transparent component. Since the noise reaching the facade was far above the acceptable value for a building to be used for educational purposes, scenarios were determined based on the options that meet the acoustic requirement. To limit multiple scenario options, energy consumption data was evaluated for sections where sufficient noise reduction value can be achieved.

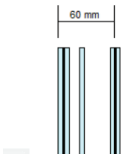
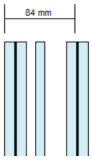
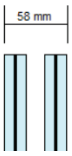
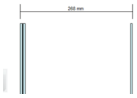
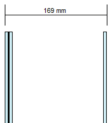
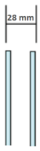
The sample educational building is located in a temperate humid climate type. The energy consumption of the building after strengthening is considerably higher than before strengthening. After strengthening, the WWR of the building facade decreased. This situation reduced the solar energy gain from transparent areas. It also caused an increase in the energy consumed for heating. In addition, the reduction of transparent areas also reduced daylight gain, resulting in an increase in the energy consumed for lighting. When developing the scenarios for energy consumption, the effect of reducing transparent areas, which enable solar energy gain, on heating, lighting, and total load was attached to importance. The scenarios developed for the after strengthening were created by considering these criteria of noise reduction and energy consumption.

The glasses with different heat conduction coefficient properties were chosen among the 2 or 3-layered glasses with high sound insulation properties, which are widely used in the market. Table 3 shows the parameters which are the layering details, weighted sound reduction index (Rw), overall heat transfer coefficients (OHTC), solar heat gain (SHGC), and light transmittance (LT) in the transparent area scenarios. The five scenarios were evaluated in terms of facade noise reduction and energy consumption. In addition, in Scenario 6, only the low-e characteristic of the currently used glass type was changed. Since this situation has no effect on the noise reduction level, only energy consumption assessments were made for Scenario 6.

Evaluation of Improvement Scenarios in terms of Energy Consumption

The total energy consumption reduction was achieved with the proposed building facade sections developed for the current situation (after strengthening) of the educational building. Figure 7 shows the energy consumption values of the whole building for the before and after strengthening, and different glass usage scenarios. Compared to the after strengthening situation, the highest decrease in the heating energy consumption was 20.6% in Scenario 1 (three-layered glass), 19.89% in Scenario 3 (two-layered glass) while in Scenario 6 there was an increase of 8.9% (two-layered glass) only. While the highest increase in cooling energy consumption was in Scenario 3 with 57.4%, there was a decrease in Scenario 6 with 8.27%. The highest decrease in lighting energy consumption was 48.3% in Scenario3. According to the current situation, lighting energy consumption decreased by 45-49 % in Scenario 1, 2, 3, 4 and 5. The maximum decrease in the total energy consumption was 20% in Scenario 1 and 19% in Scenario 3. When only the energy consumption of the building is evaluated, it is seen that the use of Scenario 1 and Scenario 3 is more appropriate than the others.

Table 3. The suggested sections and glass properties for the building envelope

	Glass Section	Glass Layer	Glass Rw (Ctr) dB	OHTC (W/m2K) U value	SHGC	LT
Scenario 1		12.5 mm Lamine Acoustic Glass, 10 mm Air Gap, 6mm Float Glass, 20 mm Air Gap, 10.5 mm Lamine Acoustic Glass	52(-1,-5)	1.34	0.47	0.59
Scenario 2		26.28 mm Laminated Glass, 10 mm Air Gap, 12mm Float Glass, 10 mm Air Gap, 25.14 mm Laminated Glass	49(0,-3)	1.66	0.45	0.66
Scenario 3		20.5 mm Lamine Acoustic, Glass 16 mm Air Gap, 20.5 mm Lamine Acoustic Glass	54(-3,-8)	1.78	0.55	0.65
Scenario 4		12.5 mm Laminated Glass, 250 mm Air Gap, 6mm Float Glass	54(-1,-5)	2.65	0.571	0.65
Scenario 5		12.5 mm Laminated Glass, 150 mm Air Gap, 6mm Float Glass	53(-2,-7)	2.65	0.571	0.65
Scenario 6		6 mm Low-e Glass, 16 mm Air Gap, 5.6 mm Green Glass	37(-2,-5)	2.15	0.175	0.174

Note: Rw: weighed sound reduction index, OHTC: overall heat transfer coefficient, SHGC: Solar Heat Gain Coefficient, LT: Light Transmittance

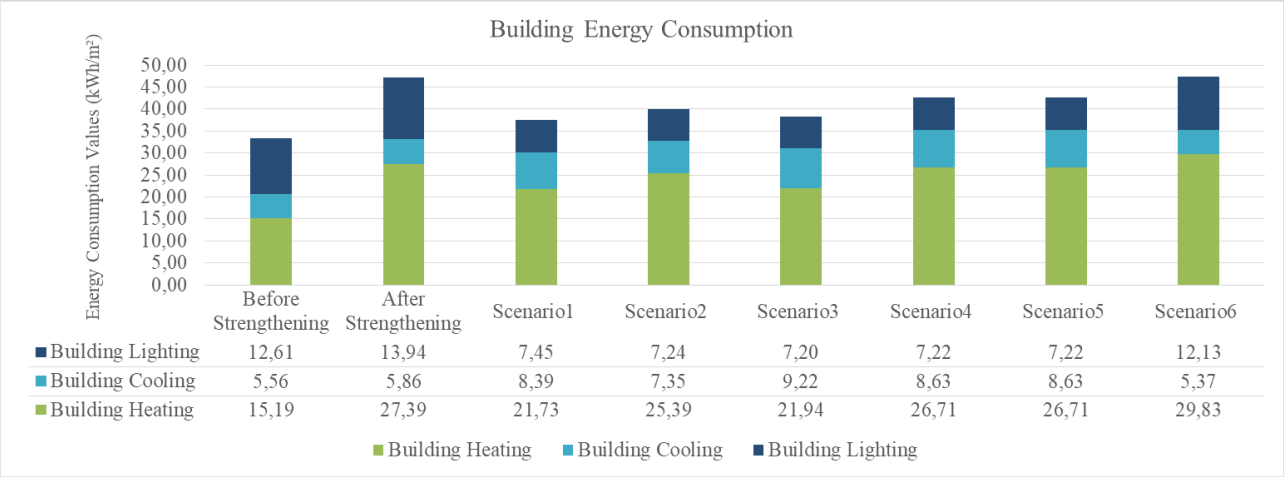


Figure 7. Energy consumption of the whole building for before and after strengthening, and scenarios.

Figures 8 and 9 show the energy consumption amounts of B113 and B114 classrooms for the scenarios developed in the study. There was an increase in energy consumption after strengthening compared to before strengthening in the classrooms. In the B113 classroom, a decrease of 15.1% was achieved in Scenario 1 and 15.9% in Scenario 3, compared to the current situation. On the other hand, cooling energy consumption increased by 56% in Scenario 1 and by 68.5% in Scenario 3, compared to the current situation. In terms of total energy consumption in classroom B 113, it was observed that there was a 19.4% decrease in Scenario 1 and 20.2% in Scenario 3. The decrease in energy consumption is quite similar in Scenario 1 and Scenario 3 according to the current situation. In the B114 classroom, it has achieved a decrease of 16% in Scenario 1 and 3 in terms of heating energy consumption compared to the current situation (Fig 9). In terms of cooling energy consumption, it has been observed that there is an increase of 29% in Scenario 1 and 38% in Scenario 3. In terms of total energy consumption in classroom B114, it has been observed that there is a decrease of 26% in Scenario 1 and 3 compared to the current situation.

Energy consumption evaluations were made for 6 different scenarios that provide acoustic comfort conditions in regions with high noise levels. According to the evaluation results, total energy consumption decreased in all scenarios except Scenario 6. Although it provides a reduction in energy consumption after strengthening, the Scenario 4 and Scenario 5 which have a U value that is not suitable for TS 825, were also not evaluated. Scenario 1 and Scenario 3 were determined as the most appropriate glass selections when the energy consumptions for total, heating, cooling, and lighting loads were examined on all buildings and sample classrooms. The window in Scenario 1 has three layers, and the window in Scenario 3 has two layers and their U values (OHTC) are different from each other and comply

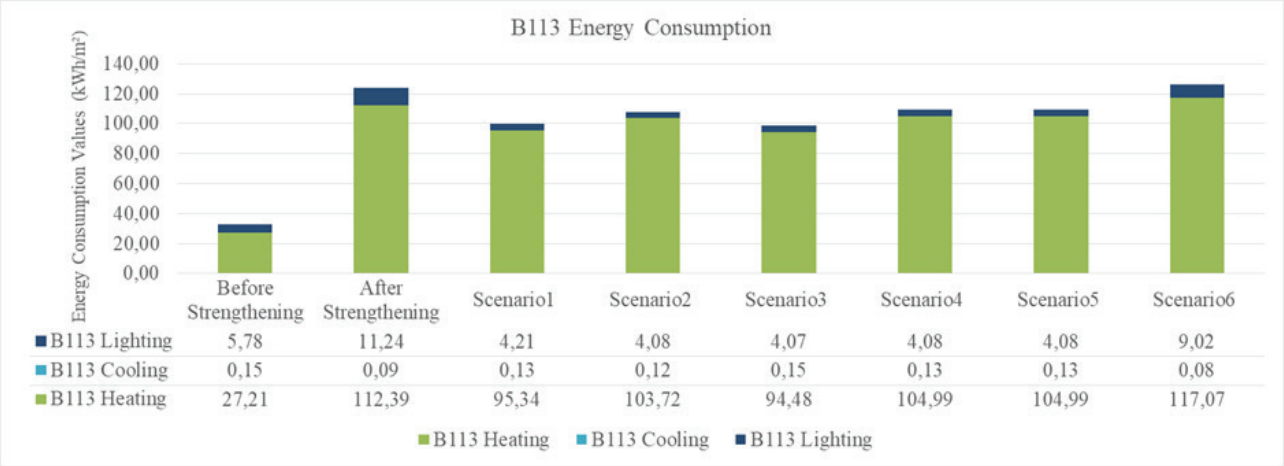


Figure 8. Classroom B 113 energy consumption values for before and after strengthening and scenarios.

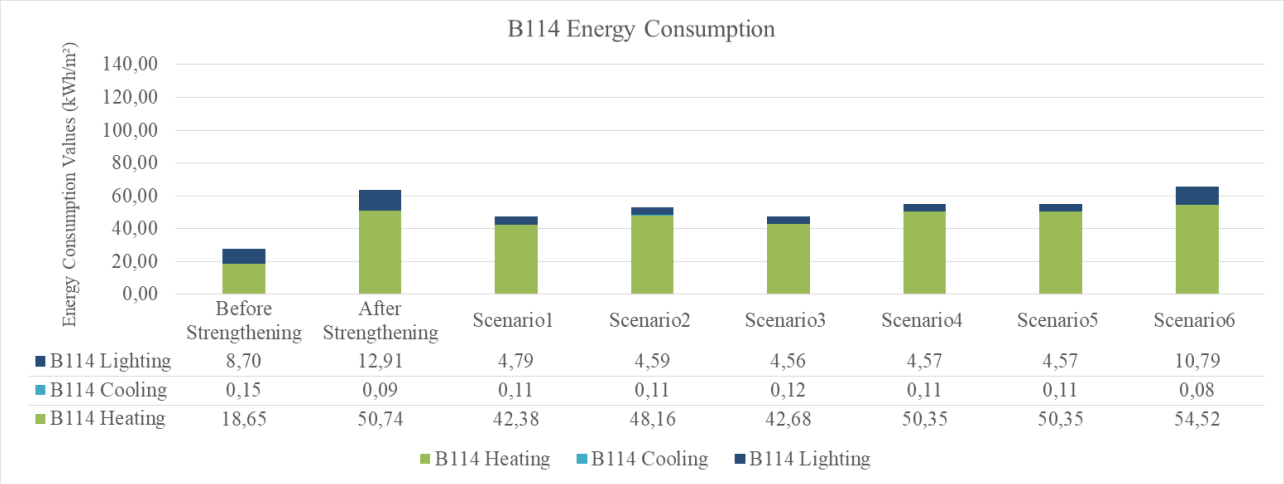


Figure 9. Classroom B114 energy consumption values for before and after strengthening and scenarios.

with TS825. However, the results are similar in terms of total energy consumption due to the difference in layer properties, thicknesses, and air gap ratios. In addition, there is a balance and similarity in total energy consumption because the amount of energy consumed for lighting in Scenario 3 is lower than in Scenario 1.

Evaluation of Improvement Scenarios in terms of Facade Noise Reduction

An increase in facade noise reduction was achieved with the building facade scenarios developed for the educational building. Figure 10 for B113 and Figure 11 for B114 show the noise reduction value of the classrooms for the before and after strengthening, and scenarios. In all scenarios, transparent area alternatives were selected that provide sufficient noise reduction value as required by the regulation.

In the classroom B113, for all scenarios, it was observed that the facade noise reduction levels increased by 14.2 dB, 13.2 dB, 13.2 dB, 16.2 dB, and 13.2 dB, respectively, compared to the current situation. In the classroom B114, for all scenarios, it was observed that the facade noise reduction levels increased by 15 dB, 14 dB, 14 dB, 17 dB, and 14 dB, respectively, compared to the current situation. The noise reduction levels ranged from 1.5 dB to 2.7 dB, depending on how the transparency ratio changed after strengthening. However, it was observed that this level increased to 17 dB with the improvement of the transparent area properties in the scenarios.

Scenario 1 and Scenario 4 have the highest facade noise reduction levels among the scenarios. DnT,A,tr values are provided for Scenario 1 to 54.3 dB and Scenario 4 to 56.3 dB. In Scenario 4 of the double-layered glass section, which is the most suitable option, the sound reduction level increases due to the increase in the air gap between panes of glass. In Scenario 1, it is seen that high noise reduction value is achieved by using 3-layer and laminated sound-reduction glasses. When the scenarios were examined, it was determined that the noise reduction level of the laminated sound-reduction glasses was higher than the laminated glass scenarios for 3-layered glasses (Scenario 1 and Scenario 2). When the results of Scenario 4 and Scenario 5 were compared, it was observed that the level of facade noise reduction improved due to the increase in the air gap between panes of glass.

It was observed that increasing the air gap between panes of glass or using glasses with noise-reducing properties in noise-sensitive volumes such as classrooms will increase the level of facade noise reduction. Therefore, comfort conditions in accordance with the regulations can be provided with appropriate glass selections in spaces with the same transparency ratio and facade properties in areas with high outdoor noise.

Optimum Building Envelope Scenarios in terms of Energy Consumption and Noise Reduction

The total energy consumption and facade noise reduction level have been examined according to the analyses made for the before and after strengthening and the glass scenarios for classrooms B113 and B114. When the situation before strengthening is compared with the current situation, it has been determined that the total energy consumption and facade noise reduction in both classrooms have increased in the current situation due to the decrease in the WWR. When the transparent area scenarios suggested

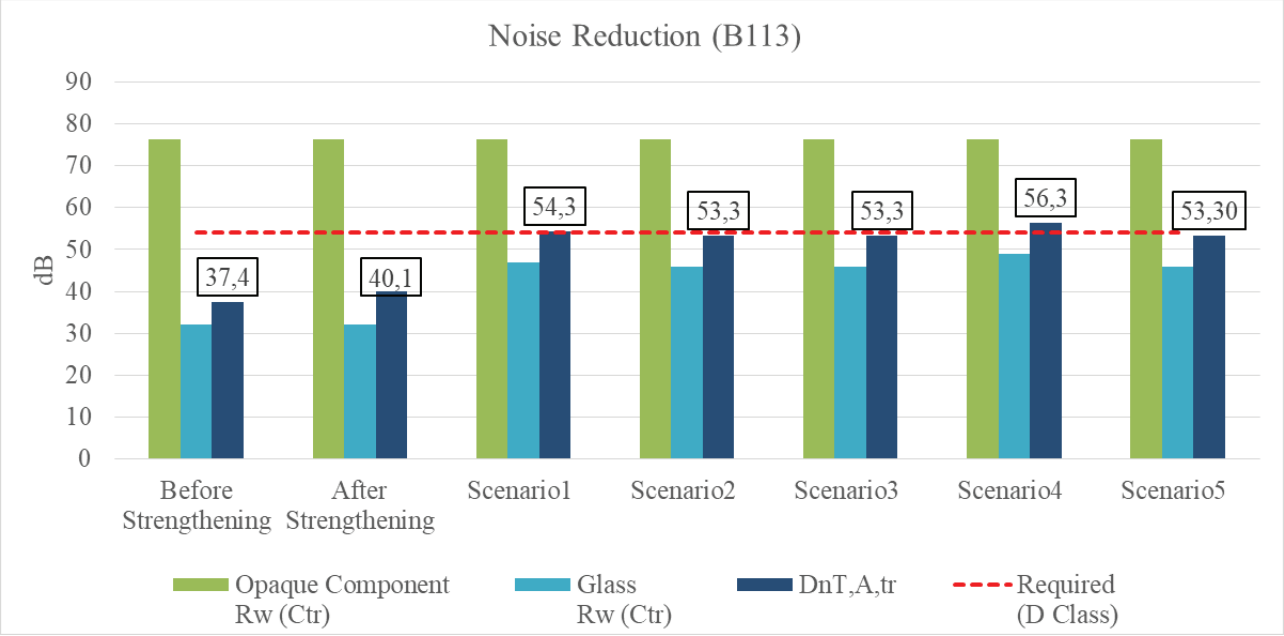


Figure 10. Classroom B 113 noise reduction levels for before strengthening, current situation and scenarios.

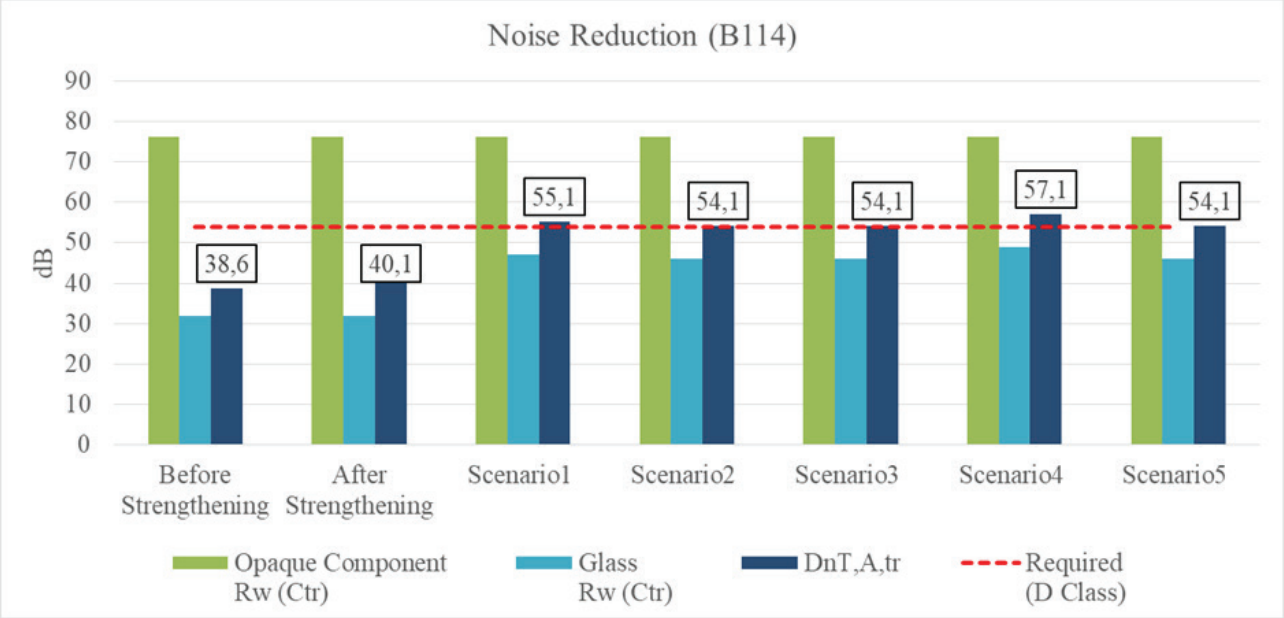


Figure 11. Classroom B 114 noise reduction levels for before strengthening, current situation and scenarios.

for improvement are compared with the current situation, it has been seen that the amount of energy consumption decreases and the sound insulation level increases. Figure 12 shows the total energy consumption and facade noise reduction levels for classrooms B113 and B114 before and after strengthening for 5 scenarios.

The lowest energy consumption was achieved with Scenario 3 in classroom B113 and Scenario 1 in classroom B114. The heat transfer coefficient and solar heat gain coefficient are lower in Scenario 1 than Scenario 3. While this situation reduces the ratio of the heating energy transfer in Scenario 1, it also decreases the solar heat gain from the outdoor environment. Therefore, similar energy consumption was obtained in Scenarios 1 and 3. In addition, the low energy consumption for lighting in Scenario 3 led to results similar to Scenario 1. Since the U value of Scenario 1 is lower, it causes a

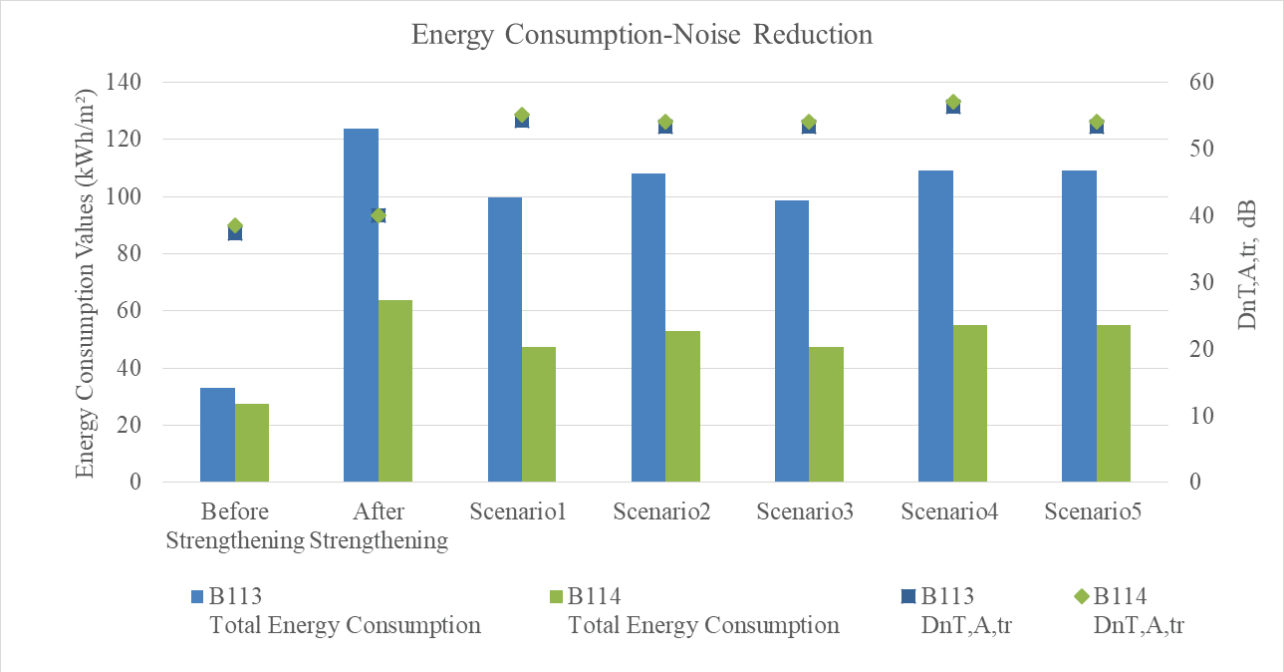


Figure 12. Classrooms B113 and B114 total energy consumption and noise reduction levels.

proportionally smaller increase in cooling loads compared to Scenario 3. Similar results were obtained in heating, lighting, and total energy consumption due to layer properties.

The highest sound insulation level was achieved with Scenario 1 and Scenario 4 in classrooms B113 and B114. The acoustic laminate glass was used in Scenario 1 and laminated glass was used in Scenario 4. In addition to the properties of the glasses, the number of layers and the gaps between the layers are also different. Although the glass thicknesses are less in Scenario 4, the sound insulation level is close to Scenario 1 due to the air gap between the panes of glass. Changes in layer thicknesses, glass properties, and distances between layers have been effective in providing the highest sound insulation level in different scenarios.

In the education building, which is in a location where the outdoor noise is high, transparent area proposals have been created that provide high facade noise reduction value in accordance with the Noise Regulation. When the energy consumption values for these scenarios are evaluated, it is seen that the energy consumption values for all scenarios decrease. In scenarios that provide the required noise reduction value and reduce energy consumption, Scenario 4, Scenario 5, and Scenario 6 , which were scenarios with U-values that do not comply with TS 825, were excluded from the evaluation in terms of noise reduction. According to the noise reduction evaluations, the scenarios that provide the highest reduction are Scenario 1 and Scenario 4. However, since Scenario 4 could not meet the standard requirements of TS 825, it is appropriate to choose Scenario 1 as the optimum scenario.

DISCUSSION

In this study, different scenarios for transparent areas were evaluated in order to provide a high noise reduction level and low energy consumption in a strengthened building located on an area with high outdoor noise. In order to optimize energy consumption and noise reduction, transparent area scenarios that provide heat conservation during the heating period and minimize the heating effect of solar radiation during the cooling period, and provide high sound transmission loss were determined. Among the suggestions, evaluations were made for six different glass scenarios with different glass thicknesses, number of layers and air gaps that meet the minimum values of the regulations.

As a result of the strengthening of the building, the energy consumed for heating, cooling, and lighting increased significantly. This situation shows that the windows should be arranged by considering thermal properties such as heat transfer coefficient and solar heat gain coefficient on facades with reduced transparency. In addition, it was observed that the level of facade noise reduction increased, but that the acoustic performance class requirements specified in the regulation could not be met due to the low insulation properties of the transparent area. This situation shows that even if the noise reduction value increases on the facades with reduced transparency, the insulation properties of the glasses should also be regulated in order to provide an adequate level.

After strengthening, the decrease in the WWR has reduced the benefit from the heating effect of the solar radiation. The observed reduction is evident in the energy consumption data for heating, cooling, and lighting. In the modeled scenarios for the existing conditions, even with identical WWR in the classrooms, the decrease in energy consumption is comparatively smaller in B113, which initially had a south-facing window, than in B114. Furthermore, the scenarios reveal a relatively higher increase in cooling energy consumption for B113 in comparison to B114.

Since the WWR decreased in the current situation compared to the before strengthening, the need for heating and lighting energy in classrooms B113 and B114 increased with the effect of the direction factor, and this situation also increased the total energy consumption. The increase in total energy consumption in classroom B113 was higher than in B114. Yıldız et al., investigated the effects of glass types and WWR on energy performance of a school building located in Izmir. They found that when WWR is increased heating loads decrease and cooling load increase [22]. Although, the objectives and methods of the studies are different, the results show that when WWR is decreased, energy consumption increases. It should be remembered that there are more parameters that affect the results, like climate type. Therefore, this study suggested a usable method for all buildings, which have different characteristics and environmental conditions.

Although the sound insulation of the opaque component is high, a sufficient noise reduction value could not be achieved on the facade due to the low sound insulation of the transparent area. In the current situation, since the WWR decreased after the strengthening, it was observed that the facade noise reduction level increased in the classrooms, but due to the low insulation level of the transparent area, adequate insulation was not provided.

Alonso et al., investigated acoustic retrofit strategies of windows depending on the regulations of certain European countries and other countries [20]. They stated that upgrading strategies focused on windows can be sufficient to comply with national building regulations in European countries and various countries globally. However, Alonso et al. did not consider Turkey in their evaluation of these regulations. Our research expands upon their work by examining an educational facility in compliance with Turkish standards, thereby providing a valuable contribution to their study. We concluded that scenarios 1 and 4 met the standard requirements for Turkey. In addition to proposing a method to evaluate strengthening designs, this study contributes to the literature on this aspect.

Scenario 1 (three-layered glass) is the proposal that provides the most reduction in the total energy consumption of

the building compared to the current situation considering the calculation results of the improvement scenarios. When the total energy consumption and noise reduction levels for classrooms B113 and B114 were analyzed, Scenario 1 was determined to be the most suitable alternative. It is seen that when Scenario 1 is applied to the whole building, there is a 15.9 % decrease in total energy consumption, and the facade noise reduction level in the classrooms increases and meets the requirements of the noise regulation valid in Turkey.

The sound insulation levels of the opaque and transparent areas of the facade did not change before and after the strengthening, but the WWR decreased. Although the facade noise reduction increased due to the decrease in the WWR, the limit values given in the regulation could not be achieved. With the scenarios created, the sound insulation values of transparent areas have been increased and all scenarios provided the required facade noise reduction limit values specified in the regulation. Although the scenarios with different layer properties provide similar noise reduction, the highest facade noise reduction level for the classrooms was obtained in Scenarios 1 and 4. Due to its physiological structure, the human ear cannot perceive low sound level differences below a certain value. For this reason, the energy consumption amounts of the scenarios gain importance in the selection of the optimum transparent area. In this study, when the scenarios for transparent areas are considered together with energy consumption and noise reduction, Scenario 1 (three-layered glass), which provides both the high noise reduction level and low energy consumption, is determined as the most suitable option for the transparent area.

CONCLUSION

The buildings need to be renewed or strengthened due to wear during their lifetime and/or damage as a result of natural disasters. After the renewal and strengthening of the structure system, providing indoor comfort conditions is important for health, efficiency and comfort. While providing comfort conditions in the indoor environment, it is necessary to keep energy consumption at a minimum and facade noise reduction at maximum level. The sectional properties of the building envelope should be improved in order to ensure that the indoor environment is minimally affected by external climatic conditions and noise sources.

A method proposal was created to examine the changes made in the envelope of an educational building, whose building's column system was renewed and strengthened against earthquakes. The optimum building envelope scenarios have been developed in terms of energy consumption and noise control. The building envelope was examined for the before and after strengthening conditions, and also six different transparent area scenarios were analyzed. The analyses were evaluated within the scope of the relevant standards and regulations and by comparing them with each other.

The changes in the transparency ratio after strengthening process, directly affect the indoor comfort conditions depending on the direction and the location. In strengthened buildings, the necessary changes should be made in the selection of glass type and the cross-section of the transparent areas, especially considering the change in energy consumption caused by the reduction in the size of the glass areas and depending on the direction of the facade. The reduction in WWR resulted in lower solar heat energy gains. So, the heating energy consumption of the classrooms increased. However, no significant difference was observed in cooling energy consumption. This study found that the total energy consumption increased by 273% in the classroom B113 and 131% in the classroom B114 after strengthening while the facade noise reduction levels (DnT,A,tr) increased by 2.7 dB in B113 and 1.5 dB in B114.

This study revealed the effect of choosing the appropriate transparent area on energy consumption and facade noise reduction of strengthened buildings. The results showed that the optimum scenario (Scenario 1) decreased the total energy consumption by 19.4% in B113 and 26% in B 114 and increased the facade noise reduction levels by 14.2 dB in B113 and 15dB in B 114.

It is expected that other alternative scenarios and the optimum building envelope glass section will contribute to future renovation and/or strengthening studies. Even if the scenarios in the study were able to decrease energy consumption in the current situation, none of the scenarios have not the building physics performance of before strengthening. The designers should consider the performance of after strengthening conditions to be able to build energy efficient buildings. Considering only the issues related to strengthening, it is determined that after the strengthening application, the building's energy consumption is negatively affected and the building does not provide sufficient conditions in terms of sound and light. In strengthening projects, the control is also required in terms of heat, sound and light during the project phase. Therefore, the method proposed in this study is important for future studies as it will lead to the control and improvement of thermal and acoustic conditions in strengthening projects.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Zomorodian ZS, Tahsildoost M. Assessment of window performance in classrooms by long term spatial comfort metrics. Energy Build 2017;134:80-93. [CrossRef]

[2] Huang L, Zhu Y, Ouyang Q, Cao B. A study on the effects of thermal, luminous, and acoustic environments on indoor environmental comfort in offices. Build Environ 2012;49:304–309. [CrossRef]

[3] Paschoalin Filho JA, Guerner Dias AJ, Storopoli JH, Ghermandi A, de Carvalho HC. Relationship between environmental indoor conditions of a classroom and the performance of undergraduate students. Int J Archit Res 2022;16:359-377. [CrossRef]

[4] Shrestha M, Rijal HB, Kayo G, Shukuya M. A field investigation on adaptive thermal comfort in school buildings in the temperate climatic region of Nepal. Build Environ 2021;190:107523. [CrossRef]

[5] Subhashini S, Thirumaran K. A passive design solution to enhance thermal comfort in an educational building in the warm humid climatic zone of Madurai. J Build Eng 2018;18:395-407. [CrossRef]

[6] Lakhdari K, Sriti L, Painter B. Parametric optimization of daylight, thermal and energy performance of middle school classrooms, case of hot and dry regions. Build Environ 2021;204:108173. [CrossRef]

[7] Ou Y, Zhang X, Zhang L, Yu S. A review on the current condition and control of campus noise pollution in central region of China. E3S Web Conf 2021;269:1010. [CrossRef]

[8] Yılmaz Karaman O, Berber Üçkaya N. Acoustic comfort in lecture halls: The Dokuz Eylül University Faculty of Architecture. Megaron 2015;10:503–521. [Turkish] [CrossRef]

[9] Catalina T, Iordache V. IEQ assessment on schools in the design stage. Build Environ 2012;49:129-140. [CrossRef]

[10] Turkish Building Sector. Energy efficiency technology atlas - Executive summary and roadmap. Available at: https://www.giz.de/de/downloads/giz2019-en-turkish-building-sector.pdf. Accessed on Jul 1, 2024.

[11] Zoroğlu F, Gedik G. An evaluation of thermal comfort conditions in shopping malls in İstanbul. In proceedings of the International Research Conference on Sustainable Energy, Engineering, Materials and Environment; 2017 Jul 26–28; Newcastle, United Kingdom. 2017.

[12] Ali H, Hashlamun R. Envelope retrofitting strategies for public school buildings in Jordan. J Build Eng 2019;25:100819. [CrossRef]

[13] Republic of Turkey, Ministry of National Education. Istanbul/Beşiktaş Sakıp Sabancı Anatolian High School Available at: http://sabancilisesi.meb.k12.tr/tema/index.php. Accessed on Nov 1, 2021.

[14] Unver R, Öztürk L, Adıgüzel Ş, Çelik Ö. Effect of the facade alternatives on the daylight illuminance in offices. Energy Build 2003;35:737-746. [CrossRef]

[15] Lai ACK, Mui KW, Wong LT, Law LY. An evaluation model for indoor environmental quality (IEQ) acceptance in residential buildings. Energy Build 2009;41:930-936. [CrossRef]

[16] Atmaca AB, Zoroğlu Çağlar F, Rengin Ünver F, Zorer Gedik G. A method for determining and improving the visual comfort change in strengthened buildings: educational building-the example of classroom. J Fac Eng Archit Gazi Univ 2022;37:1915-1930. [Turkish] [CrossRef]

[17] Akdağ NY. A simple method to determinate required Rtr values of building envelope components against road traffic noise. Build Environ 2004;39:1327-1332. [CrossRef]

[18] Guan H, Hu S, Lu M, He M, Mao Z, Liu G. People's subjective and physiological responses to the combined thermal-acoustic environments. Build Environ 2020;172:106709. [CrossRef]

[19] Yun GY, McEvoy M, Steemers K. Design and overall energy performance of a ventilated photovoltaic façade. Sol Energy 2007;81:383-394. [CrossRef]

[20] Alonso A, Suárez R, Patricio J, Escandón R, Sendra JJ. Acoustic retrofit strategies of windows in facades of residential buildings: Requirements and recommendations to reduce exposure to environmental noise. J Build Eng 2021;41:102773. [CrossRef]

[21] Balasbaneh AT, Yeoh D, Abidin ARZ. Life cycle sustainability assessment of window renovations in schools against noise pollution in tropical climates. J Build Eng 2020;32:101784. [CrossRef]

[22] Yıldız Y, Göksal Özbalta T, Durmuş Arsan Z. Impact of window-to-wall surface area for different window glass types and wall orientations on building energy performance: A case study for a school building located in Izmir, Turkey. Megaron 2011;6:30-38. [Turkish]

[23] Alwetaishi M. Impact of glazing to wall ratio in various climatic regions: A case study. J King Saud Univ Eng Sci 2019;31:6-18. [CrossRef]

[24] Alwetaishi M, Taki A. Investigation into energy performance of a school building in a hot climate: Optimum of window-to-wall ratio. Indoor Built Environ 2020;29:24-39. [CrossRef]

[25] Avsar Y, Gönüllü M. A map preparation for outdoor noises of educational buildings in fatih district of İstanbul. In proceedings of the International Symposium on Noise Control and Acoustics for Educational Buildings; 2000 May 24, İstanbul, Türkiye. 2000.

[26] Yang HS, Cho HM, Kim MJ. On-site measurements for noise reduction through open windows of classrooms with different building dispositions. Appl Acoust 2018;139:165-173. [CrossRef]

[27] Shield BM, Dockrell JE. The effects of noise on children at school: A review. Build Acoust 2003;10:97-116. [CrossRef]

[28] Atmaca AB, Akdağ NY, Ünver R, Gedik GZ. Evaluation of building envelope in terms of acoustical, visual and thermal comfort in educational buildings: A secondary school example. In proceedings of the 3rd National Building Physics and Environmental Control Congress; 2018 May 10-11; İstanbul, Türkiye. 2018.

[29] DesignBuilder Software n.d. Available at: https://designbuilder.co.uk/ Accessed on Oct 19, 2021.

[30] KS-Schallschutzrechner Version 8.00. Available at: https://www.kalksandstein.de/service-und-aktuelles/news/ks-schallschutzrechner-version-800-veroeffentlicht Accessed on Oct 19, 2021.

[31] EN ISO 12354-3:2017 Building acoustics - Estimation of acoustic performance of buildings from the performance of elements. Available at: https://www.iso.org/standard/70244.html. Accessed on Jul 1, 2024.

[32] EnegyPlus. Weather data file - İstanbul. Available at: https://energyplus.net/weather-search/istanbul. Accessed on Jul 17, 2021.

[33] EN ISO 12464-1:2011 Light and lighting-Lighting of work places Part 1 : Indoor work places. Available at: https://standards.iteh.ai/catalog/standards/cen/75239d59-3e2c-4c3a-b262-e1a80fe62a6e/en-12464-1-2011. Accessed on Jul 1, 2024.

[34] İBB Çevre Koruma ve Kontrol Dairesi Başkanlığı. Noise maps - İstanbul. Available at: https://cevrekoruma.ibb.istanbul/gurultu-haritalari. Accessed on Aug 18, 2021.

[35] Ministry of Environment Urbanization and Climate Change. Binaların gürültüye karşı koruması hakkında yönetmelik - Birinci bölüm. Available at: https://www.resmigazete.gov.tr/eskiler/2017/05/20170531-7.htm. Accessed on Jul 1, 2024.

[36] European Cooperation In Science & Technology. Building acoustics throughout Europe Volume 1: Towards a common framework in building acoustics throughout Europe. Available at: https://www.cost.eu/publication/building-acoustics-throughout-europe-volume-1-towards-a-common-framework-in-building-acoustics-throughout-europe/. Accessed on Jul 1, 2024.

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1261–1291, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00096

Review Article

Unlocking biogas production potential: Evaluating the environmental impact and biodegradability of pharmaceutical and medical wastes

Abdulhalim Musa ABUBAKAR[1,3,*], Nasir Musa HARUNA[2], Zidani Danladi AHMED[3], Aminullah Zakariyyah ABDUL[3,4], Muhammad Abbagoni ABUBAKAR[3], Amina Mohamed ALI[5], Muhammad Jamil Umar SABO[6], Balasubramanian THIAGARAJAN[7]

[1]*Department of Chemical Engineering, Modibbo Adama University (MAU), P.M.B 2076, Yola, Adamawa State, 640261, Nigeria*
[2]*Department of Animal Science and Range Management, Modibbo Adama University (MAU), P.M.B 2076, Yola, Adamawa State, 640261, Nigeria*
[3]*Department of Chemical Engineering, University of Maiduguri (UNIMAID), P.M.B 1069, Maiduguri, Borno State, 600004, Nigeria*
[4]*Department of Chemical Engineering, Federal University Wukari (FUW), P.M.B 1020, Wukari, Taraba State, 670101, Nigeria*
[5]*Medicinal Research Institute, Centre d'Etudes et de Recherche de Djibouti, IRM-CERD, 77101, République de Djibouti*
[6]*Department of Surgery, University of Maiduguri Teaching Hospital (UMTH), PMB 1414, Bama Road, Maiduguri, Borno State, 600230, Nigeria*
[7]*Sri Lalithambigai Medical College Chennai, The Tamilnadu Dr MGR Medical University, 600032, India*

ARTICLE INFO

Article history
Received: 05 January 2023
Revised: 12 July 20233
Accepted: 18 August 2023

Keywords:
Anaerobic Reactors; Antibiotics; Biogas, Pharmaceutical Wastewater; Medical Waste; Waste Management, Wastewater Treatment

ABSTRACT

Addressing the pollution problems caused by improper disposal of effluents and wastes from pharmaceutical companies and medical facilities are important for the safety of humans and animals. This work highlights the challenges and developments in medical and pharmaceutical waste management practices across the world as well as their potential for bio-energy production. It involves the study of these waste properties, their impacts on the ecosystem and treatment or recycling methods. Various studies have shown that successes have been recorded in converting some antibiotic contaminated wastewater to biogas in advance anaerobic digesters. Moreover, not all medical wastes are degradable, the use of placentas, hospital cotton waste, human urine, waste blood and surgery waste has been used in biogas plants built at close proximity to hospitals, in some cases. However, such plants are few and are only located in Tanzania, India and Philippines, among others to generate biogas to power hospitals, boil hot water needed by patients and for cooking. This is because the level of awareness as regards the dangers associated with indiscriminate disposal of medical and pharmaceutical waste is low and hence the development of waste disposal policy by countries is often overlooked. The implication of this is the spread of diseases in affected areas which can results in epidemics. It is therefore necessary to formulate policies that allow the harnessing of these wastes to biogas/bioenergy or the creation of better waste management practices that is environmentally safe.

Cite this article as: Abubakar AM, Haruna NM, Ahmed ZD, Abdul AZ, Abubakar MA, Ali AM, Sabo MJU, Thiagarajan B. Unlocking biogas production potential: Evaluating the environmental impact and biodegradability of pharmaceutical and medical wastes. Sigma J Eng Nat Sci 2024;42(4):1261–1291.

***Corresponding author.**
*E-mail address: abdulhalim@mau.edu.ng

This paper was recommended for publication in revised form by Editor in Chief Ahmet Selim Dalkilic

INTRODUCTION

Wastes in general, can be categorized into easily degradable, hardly degradable or non-biodegradable materials. Pharmaceutical waste are unused, spilt, or expired drugs or vaccines that is itself an inhibitor in anaerobic digestion systems or in simple terms, 'hardly degradable' [1]. Pharmaceutical residues are harmful to natural water sources and are just 3% of medical waste; hence referred to as micropollutants [2–7]. They are found in water bodies due to common usage and manufacturing procedures caused by increased number of emerging pharmaceutical industries [8, 9]. In that sense, nearly, 10,000 liters of pharmaceutical wastewater is generated in India daily [10]. Their characteristics short-life, high organic load and high chemical oxygen demand (COD) makes them difficult to treat using neither physical, chemical or biological means [11–13]. Medical/clinical waste are of human and veterinary origins and includes body fluids, placenta, blood, sharp objects, culture dishes and pathological waste [7]. Anaerobic digestion, as one typical biological treatment process is favored by proteineous medical and pharmaceutical waste in different anaerobic reactors [14–16]. Presently, the effect of only few out of 3000 various medical and pharmaceutical residues in wastewater or pure form (e.g., penicillin bacterial residue) for biogas production has been analyzed [17, 18].

Biofertilizer, biohydrogen and biogas are three resulting by-products of pharmaceutical effluents and medical waste via AD, with near zero carbon emission [19–21]. The technology involves 3-4 intermediate biological decomposition steps and is carried out in fixed dome, balloon or tubular digester and floating drum bioreactors via wet or dry procedures, influenced by several factors [22–27]. Most importantly, pharmaceutical waste/effluents AD is enhanced using various types of microorganisms (e.g., bacteria, protozoa, helminths, microalgae and fungi) and is efficient when proper mixing technology is involved [28–30]. For instance, Clostridium plays an important role in the degradation of compounds containing cellulose and starch, Bacillus helps in the decomposition of proteins and fats and rhizosphere microbiota only degrade plant waste and organisms [23, 27, 31, 32]. Microalgae and blue-green algae (cyanobacteria) can be used in many ways. Namely, as food by humans, a potential pharmaceutical contaminant remover, heavy metal remover, in biofertilizer production by fixing nitrogen, bioplastic production and as biogas/biofuel/biodiesel/biohydrogen production ingredient [33–35]. Contaminants elimination using microalgae follows 3 mechanisms comprising of bioadsorption, biodegradation and accumulation [36]. Microalgal AD is suitable for biogas production as earlier affirmed by scientist since the 1950s [33]. Notably, the same author selectively degrades 9 antibiotics and 1 antidepressant (10 PhACs) at laboratory scale. In 1980s, potentials of fungi species numbering over 14,000 to digest pharmaceutical compounds in effluents, was proposed [37, 38].

An overview of biogas development from the past up to this present time shows that there has been rapid growth in the technology since 1920 in terms of its utilization for street lighting, cooking fuel and transportation fuel [20, 24]. A plant to digest agricultural, municipal and industrial waste, including medical and pharmaceutical wastes had been developed with set target for future multiplication in Skåne, Poland, Pakistan, Austria, Denmark, Turkey, Argentina, Sweden and other European nations [24, 27, 39–45]. Thus, precise aim of building these plants is to reduce organic waste pollution, deforestation, emission reduction and sustainable energy development. But Gittelson et al. [46] lists several reasons why biogas is not sustainable, arguing that the gas is highly toxic, flammable, potentially explosive and contains greenhouse gases which are released via combustion processes and diffusive emissions comparable to manures. The authors also maintained that, transporting biogas would require pipelines with exact problems faced with natural gas (such as leaks and emissions) which did not 'as believed' solve the emission issue it was designed to address. However, the new assertion hasn't been analyzed by scientist to challenge the continuous development of biogas plants across the globe. Moreover, apart from this technology, some wastes, especially of medical and pharmaceutical facilities would lie unattended to, thereby posing serious health issues to humans and animals. Hence, the general objective of this work is to review the pharmaceutical industry and related products and the treatment of wastewaters emerging from them. This work tries to study the dangers associated with medical and pharmaceutical waste in the ecosystem; their possible decomposability, by x-raying their previous exploitation for biogas generation; and the need for future research on these two wastes sources, as specific objectives. This was carried out by studying the pharmaceutical industry products and effluents and their pollution effects and the roles of different bioreactors in enhancing their conversion into useful products.

LITERATURE REVIEW

Pharmaceutical Industry

End-users of products from pharmaceutical companies including raw materials, veterinary medicine, antibiotics and cosmetics are humans and animals [47–49]. In pharmaceutical industries, large varieties of these products emanates from five manufacturing processes, namely, fermentation, extraction, chemical synthesis, formulation and packaging [50–52]. Any of these processes has the potential of generating wastewaters containing higher organic loads of harmful constituents to plant, human, animal and aquatic life [48, 50]. The chemical synthesis and fermentation categories of bulk pharmaceutical production process, produces higher volume of wastewater containing salts, recalcitrant organics, spent solvents and pharmaceutical residue (PR) [50, 52]. Currently, the world is witnessing

an increased growth in the pharmaceutical industry as core components of the chemical industry [53]. The pharmaceutical supply chain normally has four components, namely, the primary manufacturer, secondary manufacturer, distribution centers, and retail outlets (including hospitals and pharmacies) [54]. Majority of the globe's largest producers of pharmaceuticals are actively based in Ireland, where considerable amounts are also exported [55]. There are 250 multinational pharmaceutical companies all over India which is regarded as one of the largest and fastest growing industries [56–58]. Andersson & Karlsson [59] reported that, Sweden uses more than 1000 different active substances in about 7600 pharmaceuticals. First pharmaceutical industry in Nigeria was the May & Baker Nigeria Plc founded in 1944 – adding up to over 115 registered manufacturers in Nigeria which started in early 1960s [60]. However, due to high population figures in China, they are understandably, the highest manufacturer and consumers of pharmaceuticals globally [61]. Presently, the largest pharmaceutical company in South Africa, 'Aspen Pharmacare', was established in 1850; while 'Adcock Ingram', the second largest pharmaceutical firm in the country, trace its roots to 1890 [62]. Pharmaceutical products from Kenya often end up in Uganda, Burundi, DRC, Tanzania, Rwanda, Malawi, the Comoros and Ethiopia, among other destinations [63]. List of pharmaceutical companies in Kenya, Nigeria, Egypt and South Africa can be found in the literature [60, 63–65]. Pharmaceutical companies in Nigeria are located in Lagos, Ogun, Kano, Imo and Anambra states. However, the activities of some unregistered pharmacies in cities like Lagos, makes the garnering of accurate data of its usage in that part of the country difficult [66]. In 2021, revenues from pharmaceuticals was forecasted by WHO [67] to reach $1400 billion. North America accounts for 49% of the revenue, Africa, 0.7%, Europe, 21.8%, Asia-Pacific, 21.7%, Latin America, 4.6%, and Turkey, Middle-East-Eurasia 2.2%. In Africa, most of the pharmaceutical industries (representing 80%) are concentrated in just 8 countries (Egypt, South Africa, Algeria, Tunisia, Nigeria, Morocco, Kenya, and Ghana), according to a 2020 report which puts their sum at approximately 600 [68]. Experts predicts a further growth (from 2020 onwards) in pharmaceutical manufacturing in Africa from a value of $40-$65 billion – a growth trend that was maintained compared to 2013's worth of $20.8 billion [69]. Asplund [13] predicts that, the world's medicine market may go back to pre-pandemic levels in 2022, approaching almost $1.8 trillion by 2026.

Pharmaceuticals

Preferably, the main role of pharmaceuticals is to influence the function of human biological elements, especially their cells and also methanogens responsible for anaerobic digestion [70]. Pharmaceuticals are generally classified into antibiotics, tranquilizers, antiepileptic, diuretics, azole (fungicide), anti-inflammatory, human drugs, analgesics, illicit drugs, blood lipid regulators, beta-blockers, anticancer drugs, antirheumatic, personal care products (PCPs), psychotropic, cholesterol medicines, steroids and related hormones [13, 38, 71–74]. Some examples of pharmaceuticals of veterinary and human prescription, and hormones are paracetamol, rifaximin, sparfloxacin, sulfamerazine, ceftazidime, cilastatin, triclosan, β-estradiol, roxithromycin, minocycline, sulfisoxazole, androstenedione, megestrol acetate, ciprofloxacin, caffeine; chloramphenicol, clindamycin, danofloxacin, chlortetracycline, diclofenac, enrofloxacin, doxycycline, estriol, ethisterone, estrone, fenbendazole, sulfone, fenbendazole, ketoprofen, cloprop, clofibric acid, carbamazepine, ibuprofen, flunixin, florfenicol, ivermectin, marbofloxacin, lincomycin, metronidazole, ofloxacin, norfloxacin, oxytetracycline, sarafloxacin, sulfacloropyridazine, progesterone, fluoxetine, sertraline, sulfadiazine; sulfadoxine; sulfadimethoxine, clotrimazole, simvastatin, tamoxifen, sulfamethoxazole, sulfamazine, sulfathiazole, sulfaquinoxalin, tetracycline, testosterone, tilmicosin, and trimethoprim, among others [75–78]. Lincomycin ($C_{18}H_{34}N_2O_6S$) (LCM) is used to control gram-positive bacterial infection of human and animal and ranked second in 2012 in China in terms of consumption [61, 79]. LCM industries generates solid waste byproduct called lincomycin fermentation residues (LFR) containing residual LCM (whose concentration is around 2000 mg/kg), mycelial cells and different organic matters which can be degraded by black soldier fly larvae (BSFL) and associated microbiota [79]. Vultures and fish are reported to be gravely affected by some pharmaceuticals. In Pakistan, a study points at diclofenac residues as responsible for reduction of vulture population in the country, while mixtures of fluoxetine, ciprofloxacin and ibuprofen in g/L concentration range results in deaths of fishes [57].

Antibiotics Production and Uses

Being one of the most widely used pharmaceuticals, antibiotics occur in form of a compound or various array of molecules often channeled for the treatment and prevention of microbial infections in veterinary medicine and humans [47, 80]. In human and veterinary applications, there are about 250 different types of antibiotics [81]. Annually, China produces around 210,000 and consumes about 90,000 metric tons of antibiotics [61]. But according to Zhong et al. [17], in 2007, China produces 1.21 million tons of antibiotics, being the largest exporter and manufacturer in the world. Spain and Germany are the highest end-users of antibiotics within the European Union (EU) [82]. Hazardous waste such as antibiotic mycelial residues (AMRs) or antibiotic bioferment residues are produced in excess of 1.3 million tons in China, while 80,000 tons of veterinary antibiotics were used in 2013 in the same country [82, 83]. Consumption of veterinary antibiotics in 2013 in the United States was 11,000 tons, but global consumption of the generality of known antibiotics is still in thousands of tons [82, 84]. Mitchell et al. (2013)[85] stated that, of the 16,000 metric tons of antibiotics traded annually in the

country, approximately 80% are used in animal husbandry. Antibiotics are present in mixture form in the environment and are regarded as pollutants since they kill or inhibit the growth and development of microorganisms and aquatic animals [47, 86, 87]. Basically, they find applications in farming, human medicine, aquaculture and veterinary medicine, which end up in sewage or residues/excreta of poultry farms and livestock, causing heavy loses of microorganisms used in aerobic and anaerobic active sludge for effluent treatment [73, 84, 87]. Ofloxacin is an example of synthetic antibiotics widely used in aquaculture [70]. For instance, the antibiotics used as feed additives, therapeutic animal treatments, growth promoters and those used to prevent infections in pig farms can be excreted unchanged to the ecosystem [88]. This is the reason some veterinary pharmaceuticals are found in manures of some animals which contaminates the soil as well as ground and surface waters as a result of runoff from fields [89, 90]. Therefore, the digestion of some categories of antibiotic-contaminated cattle faeces or dungs is affected, as active methanogens are inhibited [91].

Few examples of antibiotics are amoxicillin, nitrofurantoin, gentamicin, rifampicin, doxycycline, tylosin, neomycin, erythromycin, sulfamethoxazole (SMX), tetracycline, roxarsone, carbadox, monensin, ofloxacin, streptomycin, flumequine, ciprofloxacin, bambermycin, apramycin, bacitracin, spectinomycin, oleandomycin, tiamulin, efrotomycin, trimethoprim, LCM and penicillin [23, 73, 92]. Erythromycin ($C_{37}H_{67}NO_{13}$), a macrolide antibiotic that emanates from *Saccharopolyspora erythraea* (or actinomycete), sulfamethoxazole (SMX), tetracycline, fluoroquinolones, ciprofloxacin (CIP) and enrofloxacin (ENR) are the most widely used human and veterinary antibiotics pausing as pollution threats to the environment [31, 47, 93–95]. An original form of erythromycin derivative called clarithromycin can be discharged unavoidably to the environment through human and animal excretion as it is not absorbed or utilized fully by the patient animal or human [96]. Clarithromycin utilization doubled to 15 tons per annum from 2002 to 2009 in Germany [96]. Erythromycin and tetracycline both inhibit protein synthesis while tetracycline forms complexes with ions present or bind to soil particles, where in 2018, 25% was used in veterinary medicine accounting for 33% of antibiotic consumption in 2012 in Europe [95, 97, 98]. SMX is basically a sulfonamide (SAs) bacteriostatic antimicrobial antibiotic used to arrest urinary tract infections, as it inhibits bacterial proliferation; as well as to prevent or treat mastitis in cattle and respiratory infections when combined with trimethoprim [86, 99–101]. But SAs, tetracyclines, β-Lactams, macrolides, fluoroquinolones and nitroimidazole derivatives are the most prescribed antibiotics in animal husbandry veterinary pharmaceuticals with a share use of 90% in the UK, 77% in the EU and above 50% in Denmark and Korea [91, 97, 102]. β-Lactam residues are hardly spotted in dairy manure compared to tetracycline residues that is persistent and common [103]. SAs, lincosamide, and macrolide residues, according to the same author is typically below 1 ppb in US dairy manure samples measured. SAs utilization in Germany is up to 9% and 2-11% in Europe, among which high amount of sulfadiazine (an SAs group) was detected in chicken and pig manure [97]. Visca et al. [93] reported that, SMX (which is quite mobile) at initial concentrations of 4-20 mg/kg in the soil, would half within 4-13 days. Tetracycline kills microorganisms by terminating protein synthesis after binding to the cell ribosomes, while other sub-classes such as oxytetracycline (OTC) and chlortetracycline (CTC) helps in improving health and growth efficiency when used in animal feeds, as well as for therapeutic purposes [23, 104, 105]. Tetracycline, CTC, and OTC (first isolated in 1940s) are cheap and measured at concentrations of ≈ 5, 10, and 250 μg/kg respectively in dairy manure [23, 103]. OTC, CTC and ENR are persistent veterinary antibiotics in the environment that is used in livestock farming [23, 106]. ENR is usually administered orally to cattle or by subcutaneous injection to treat alimentary tract and respiratory infections [93]. Both fluoroquinolones and CIP has high affinity for soil and persist for several months in the environment [93].

Amoxicillin belongs to the penicillin class and can be used to treat gastro-intestinal veterinary infections [91, 107]. An oral drug known as cephalexin in conjunction with cephaloglycin is used in treating diseases caused by gram-negative bacteria [108]. The outer membrane of gram-negative bacteria makes them more resistant to antibiotics, because it thwarts their penetration into the cells compared to gram-positive bacteria [109]. However, both cubical gram positive and circular gram negative forms of bacteria are killed by amoxicillin along with anaerobic bacteria as stated by Nuengjamnong et al. [110]. Hitherto, more than 700,000 deaths per year is said to be a result of antimicrobial resistance globally, and further estimates put this figure at 10 million by 2050 [31, 77]. Practical occurrence is at Puri (Bay of Bengal), India, in which previous investigations indicates the presence of 38 multi-resistant bacteria [105]. Based on experience and as stated by Oliver et al. [103], antibiotic-resistant bacteria (ARB) are present in dairy manure.

Pharmaceutical Effluent

The poor biodegradability of pharmaceutical plants' wastewater or effluent is due to large compositions of inorganic and organic toxic pollutants (such as intermediate products, catalysts, spent solvents, additives and reactants), high salt concentration and its characteristic dark color, despite having higher chemical oxygen demand (COD) and a low biochemical oxygen demand (BOD) [63, 111–116]. BOD and COD in wastewater can be determined using procedural steps illustrated by Njuguna et al. [63]. The pH of pharmaceutical effluents ranges from 4-11 while its salt content (e.g., chlorides, bromates and sulfates) could be above 1 g/L [117, 118]. Therefore, pharmaceutical industry wastewater comes from chemical synthetic plants

whose constituents are complex and pose serious risk to the ecosystem by inhibiting active biomass even at negligible amounts [53, 111, 119]. The high COD level in the toxic effluents is normally reduced using a chemical treatment technique called wet oxidation: which is defined as the transfer of one or more electrons from a reductant (an electron donor) to an oxidant (an electron acceptor) having higher affinity for electrons [11, 53, 120]. Otherwise, the wastewater discharged from pharmaceutical industries are difficult to treat because of the variable character the constituent toxic compounds possess [11, 121–123]. Above all, treatment of chemical synthesis-based pharmaceutical effluents involves a lot of complex operations due to formation of chemical reactions in the process [122]. Notwithstanding its characteristic high COD level, anaerobic treatment is generally favored [111]. A starting specimen for formulating commercially vital amoxicillin and ampicillin or 6-Aminopenicillanic acid (6-APA) are pollutants commonly found in the chemical synthesis and fermentation class of pharmaceutical wastewaters [107]. Others, as investigated in Kocaeli Province (Turkey) chemical synthesis pharmaceutical industry effluents by Gulmez et al. [122], are solvents such as isopropanol, methylene chloride, tetrahydrofuran, ethyl acetate, pyridine, methyl isobutyl ketone and methyl ethyl ketone, used at different production stages. Examples of wastewaters of pharmaceutical industry sources from manufacturing processes are azalide antibiotic raw water, azalide antibiotic pre-treated water, broad spectrum tetracycline antibiotic raw water, diuretic production raw water, disinfectant chlorhexidine-dihydrochloride raw water and molasses slops raw water [119]. It is however obvious that pesticides and some micropollutants are not scrutinized for potential environmental effect in the literature compared to drugs like analgesics, antituberculosis, antibiotics and antihistamines despite their utilization on the same scale [89, 124].

Environmental Contamination

Pharmaceutically active compounds (PhACs), about 2300 active pharmaceutical ingredients (API) and pharmaceutical wastewaters are anthropogenic contaminants that affects human health and the ecosystems when released into the surrounding environment above certain concentrations; because they contain high COD due to the presence of organic and inorganic constituents [105, 125–128]. In such plants, wastes or effluent disposal is subject to further processing in pharmaceutical wastewater treatments plants (WWTPs), which is not often the case because half of the pharmaceutical wastewater generated globally are discharged without prior treatment as reported in the literature [127, 128]. Zhan et al. [129] describe pharmaceutical process residues (PPR) as herbal and antibiotic fermentation residues. Numerous pharmaceutical residues are practically unaffected or barely removed even after passing through WWTPs [126, 130]. In addition, non-industry pharmaceutical wastes released to the environment are as a result of prescriptions (veterinary medicinal products) taken by humans and animals which are often excreted with faeces and urine in its original form (or parent substances) [73, 131–133]. Typical instances are the detection of LCM (7820 mg/L) in swine manures and 3 swine farms aqueous wastewaters (166 mg/L) in China due to intense application [61].

Researchers across the globe have detected several PhACs and PCPs in soils, ground water, drinking water supplies, sea water, surface water (lakes, rivers, and streams), municipal wastewater, wastewater treatment plants effluents and slaughterhouse wastewater, with detrimental effect to livestock, aquatic life and human health even at low concentrations (sludges) [50, 57, 73, 81-82, 130-131, 134-136, 137, 138,]. However, Alenzi et al. [92] attributed this to huge demands for medicines by users on daily basis. Concentrations of these micropollutants (pharmaceutical compounds) vary from ng/L to μg/L in natural water bodies and domestic effluents to up to mg/L levels in streams [77, 81, 86, 139–141]. Notable drugs or PhACs that have been detected in dirty water, soil and garbage are pivalic acid, 2-ethylhexanoic acid (2-EHA), antiepileptic drug carbamazepine (CBZ), vitamins, antidepressants, sleep aids and narcotic pain relievers to mention a few [121, 140, 142, 143]. Moreover, 80 persistent pharmaceuticals and medical substances (such as cytostatic drugs, clofibrate, antibiotics and analgesics) have been spotted in drinking water, sewage, sewage treatment plant wastewaters, surface and ground waters of about 10 countries, since 1980s as described by Nguluka et al. [57] and Fountoulakis et al.– while 200 different others are found globally in river waters according to Alenzi et al. 92]. The manufacturing, trade, storage and use of 12 persistent organic micro-pollutants (POPs) had earlier been restricted at a global treaty Stockholm Convention to safeguard the ecosystem [144]. The POPs covered are chlordane, aldrin, dieldrin, DDT, endrin, hexachlorobenzene, heptachlor, mirex, polychlorinated biphenyls, polychlorinated dibenzofurans, polychlorinated dibenzo-pdioxins, toxaphene, beta hexachlorocyclohexane, alpha hexachlorocyclohexane, chlordecone, hexabromobiphenyl, commercial octabromodiphenyl ether (hexabromodiphenyl ether and heptabromodiphenyl ether), pentachlorobenzene, lindane, perfluorooctane sulfonic acid (PFOS), its salts and perfluorooctane sulfonyl fluoride (PFOS-F).

Notable PhACs Pollution

Canada, USA, Brazil and many European nations found over 80 drugs in their aquatic ecosystems – UK alone detects 70 [92, 145]. This includes up to 83 ng dm^{-3} of paracetamol (in groundwaters) in the vicinity of Gdańsk [98]; a 1994 discovery of clofibric acid in Berlin (Germany) surface waters [146], around 0.3-19 ngL^{-1} concentration of clofibric acid in North Sea, and approximately 100 ngL^{-1} in estuaries, both in UK [105], and as well as in the 1970s in WWTPs samples in the USA [145]; 1.7g/l concentrations of antibiotics in surface waters according to United States

Geological Survey (USGS) [90] and; 28-31 mg/L of ciprofloxacin in India's treated effluents and natural water bodies as well as UK, USA and Germany in the range of ng/L to g/L [57], to mention a few. Again, 3800 µg/L of metaxalone (a muscle relaxant), according to Colella [147], are found in New York's pharmaceutical manufacturing wastewaters. Clotrimazole in concentrations ranging from 3-54 ng/L are present in Germany and UK ecosystems, as reported by Pavithra et al. [105]. The same authors also position the presence of diclofenac concentrations at around 2.3 µg/L in UK pharmaceutical effluents. Propranolol, carbamazepine and ibuprofen had contaminated Doñana Park in Spain, whereas in Warsaw, Poland, clindamycin in concentrations of 134 ng/L have been found in influents pharmaceutical treatment plants [74, 78].

According to literature, analysis of dams, hospital effluent and sachet waters in Nigeria shows the presence of some antibiotics in alarming proportions. Usuma dam [148], which delivers treated water to Nigerian capital territory for instance, contains metronidazole, trimethoprim, ciprofloxacin, ibuprofen and amoxicillin whose risk levels are classified into low, medium and high to aquatic species, according to Ilechukwu [149]. Previous reports show high proportion of amoxicillin (8066 µg/L) in Sango Ota, Ogun state in Nigeria which has few pharmaceutical companies and whose population largely depends on the antibiotic [150]. Similar studies reveal the presence of 34.31 µg/L of acetaminophen in Ahmadu Bello University Teaching Hospital (ABUTH) samples of the hospital plant's wastewater treatment effluent in Zaria, Kaduna state and 2.57 µg/L of the same pharmaceutical product in irrigation wells in Sango Ota, a pharmaceutical industrial area in Ogun state in the country [150]. Nigerians presumed sachet water as 'pure' and so is taken in every state in the country. However, detection of some pharmaceuticals in Ogun and Lagos sachet water can be attributed to several functional industries in those areas and so should be a health concern for food safety bodies and regulators in the country [149]. Analysis of wastewaters in Kenya points to high concentrations of ibuprofen (26.54 mg/L), ciprofloxacin (14.98 mg/L), norfloxacin, sulfamethoxazole (62.83 mg/L), metronidazole (29.92 mg/L), paracetamol, trimethoprim (208.30 mg/L) and zidovudine [63]. In Machakos, 49,300 ng/L of sulfamethoxazole and 2800 ng/L of norfloxacin in Kangemi cities in Kenya has been detected in influent waters as well as many other pharmaceuticals in Nairobi River Basin in the country [78, 150]. According to literature, 8430 ng/L of trimethoprim is concentrated in influents wastewaters in Kampala, Uganda, 1193 ng/L of erythromycin in Choutrana, Tunisia, and 88,012 ng/L of ciprofloxacin, 20,656 ng/L of metronidazole and 5742 ng/L of ofloxacin are found in Durban in South Africa [78].

Frascaroli et al. [78] collectively reports concentrations of pharmaceuticals such as azithromycin (115,413 ng/L), clarithromycin (6917 ng/L), oxytetracycline (1531 ng/L), roxithromycin (19,135 ng/L) in Sanya – as well as tetracycline (374 ng/L) and sulfadiazine (574 ng/L) in Xinjiang cities in China high-strength wastewaters. LCM of concentrations between 3000-9000 mg/L in fermentation broth are said to be discharged from manufacturing facilities in the same country [61]. Studies had shown the occurrence of ibuprofen, propranolol and carbamazepine in Yangtze Estuary in China [74]. Collectively, Zhu et al. (2021)[61] stated that, roughly, 2200 metric tons of pharmaceutical effluents are released from production centers in the country.

Pharmaceutical Wastewater Treatment

There are three processes applied by wastewater treatment experts to disinfect it for public use and environmental protection, and they include physical, chemical and biological methods, which is further divided into preliminary, primary, secondary, tertiary and advanced treatment processes [13, 151–153]. In the pharmaceutical industry, two biological treatment techniques is used to treat an effluent of pharmaceuticals manufacturing origin, which is aerobic and anaerobic [47, 154-155]. Typical dual combinations of the biological processes is at the Brazilian wastewater treatment plants employing a hybrid units of anoxic, aerobic and anaerobic techniques [77, 156]; the one investigated by Inanc et al. [157] using aerobic-anaerobic techniques to disinfect a chemical synthesis-based pharmaceutical plant byproduct and; the successful toxic and organic matter removal from pharmaceutical effluent that is composed of nitroaromatic compounds using sequencing batch biofilter, which are known for combining the two technologies [158]. Alternatively, the sole use of anaerobic treatment method have been characterized with less energy consumption and high efficiency in treating high-strength pharmaceutical effluents, making it convenient for sludge digestion [153, 159-160]. Though it has the disadvantage of longer start-up time owing to low methanogens growth rate, it is still the most promising technique compared to the aerobic process which gained popularity in the 1960s [16, 121], and under which several innovations has been done to advance the technology. Currently, power is generated by anaerobic decomposition of sewage sludge in 146 UK facilities, amounting to 66% of all municipal sewage sludge in the country [80]. In the past, Schlott et al. [161] produced a design, erection and start-up of an anaerobic treatment scheme for pharmaceutical waste, whereas Murugesan et al. [162] happily shares the development in utilizing the process in Netherlands at the moment. Almost all anaerobic digestion plants are operated at mesophilic temperature regime – so even when typical influent temperatures received are less than 18°C, 30% of energy is expended to heat it prior to treatment, according to Enright et al. [55]. Therefore, a cost-effective option is psychrophilic anaerobic digestion (PAD) occurring at temperature less than 20°C. Activities of methanogens are significant in this process, and so the measure of the specific methanogenic activity (SMA) highlights the activity of the organism

in pharmaceutical wastewater being digested in bioreactors [55, 70, 90, 128]. Again, the digestion of some pharmaceutical wastewater containing high level of sulphate is reduced to sulphide by sulfate-reducing bacteria, also responsible for hydrogen sulphide generation in anaerobic bioreactors [16, 23, 111]. But since 1970, better significance is attached to a treatment option called anaerobic biofilter because of its advantage over the two conventional oxygen-related methods for toxic elimination from pharmaceutical effluents [121].

Pharmaceutical sludge has complex compositions, bad odor, and is highly corrosive due to the presence of salts, suspended solids, pathogenic microorganisms and refractory antibiotics (e.g., aureomycin, benzylpenicillin, berberine hydrochloric, colistin sulphate, ofloxacin, etc), with imminent danger to the environment if not properly treated before discharge [95, 115, 163–165]. Even though pharmaceuticals exist in low amounts in surface waters and domestic raw water, it rivals hospital and pharmaceutical dirty water which has higher concentrations (≅ 100–500 mg/L), in terms of their persistence and toxicity [30, 47, 74, 94, 166-167]. Consequently, high antibiotic levels in wastewaters is difficult to purify or remove and as such, genetically modified strains of microorganism is endorsed to treat such antibiotics contaminated effluents [87]. The term 'pharmaceutical removal' as used here, implies the loss of the parent compound via physical, chemical, bio-decomposition and sorption to solid organic matter [73, 168]. Classically, to achieve this objective, 4 kinds of sewage sludge can be recognized and treated in pharmaceutical wastewater treatment plants, namely, primary and secondary (called mixed sludge), aerobic/anaerobically fermented and dehydrated sludge, using landfill, incineration and biodegradation technique [127, 165, 169-171]. Both aerobic and anaerobic process and landfill is proven as ineffective technique in pharmaceutical sludge treatment, and as such, a burden in Europe that produces 15 million tons of sludge in 2021 alone [127, 165]. Practical basis is drawn from non-removal of carbamazepine, ciprofloxacin and triclosan containing PhACs in sewage sludge using aerobic and anerobic means according to Mejías et al. [127]. The same author refer to PhAC removal from sewage sludge as time-dependent, giving that it took doxycycline and tetracycline 77 days to be slightly removed. Affirmed that some bacterial strain can be used to degrade ciprofloxacin from pharmaceutical wastewater.

Advance chemical oxidation process (AOP) is a method known to increase treatment efficiency of industrial wastewater and promote its biodegradability [11, 172]. The process involves oxidation with oxygen at 200–300^{o}C and the use of energy oxidants (e.g., peroxone, hydrogen peroxide, sulfate radical, photons, and/or ozone) [11, 73, 173-174]. Ozone can be used to break down pharmaceuticals in water, but is however energy intensive and costly [11]. Ozonation is also a COD and color remover from wastewater [38, 112, 175]. Biodegradable nanoparticles can also be used to detect contaminants in water for its purification [176]. In addition, Zhang et al. (2022)[177] draws the attention of researchers to biological nitrogen removal from pharmaceutical dirty water, being an inhibitor and a bio-toxic. Alternatively, pollutants in sludge can be eliminated using biochar/activated biochar; especially caffeine which is a common pharmaceutical wastewater contaminant, as well as the use of biogas residue biochar to adsorb tetracycline [178–181]. Other progress in this direction includes the utilization of *Washingtonia robusta* to remove improved pollutants and heavy metals from pharmaceutical wastewater by Al-Samrraie et al. (2022)[182], tea-based materials and coffee utilization to remove pharmaceuticals in contaminated water, reviewed by Madikizela & Pakade (2023) [183], the use of hybrid constructed wetlands [184], and the use of air and pure oxygen to treat pharmaceutical industry wastewater, researched by Gnanavel & Muthusamy (2018).

Roles of Anaerobic Treatment Reactor Types

Several anaerobic reactors of unique configurations were previously studied for pharmaceutical wastewater treatment in the literature [16, 24, 47-48, 50–52, 77, 86, 94, 99, 107, 111, 115, 121-122, 128, 157-158, 185–194]. They include, anaerobic contact reactor (ACR), anaerobic baffled reactor (ABR), anaerobic batch reactor, anaerobic filters, anaerobic fluidized bed reactor (AFBR), anaerobic fixed film reactor (AFFR), anaerobic fluidized membrane bioreactor (AFMBR), anaerobic bio-entrapped membrane reactor (ABEMR), anaerobic membrane bioreactor, anaerobic mesophilic fixed film reactor (AMFFR), anaerobic plug-flow reactors (APFRs), anaerobic structured bed biofilm reactor (ASBBR), aerobic/anaerobic sequencing batch reactor (ASBR), anaerobic suspended film contact reactor (ASCR), anaerobic thermophilic fixed film reactor (ATFFR), anaerobic up-flow packed bed reactor, biofilm airlift suspension reactor (BASR), continuous stirred tank acidogenic reactor, electrochemical membrane bioreactors (EMBR), expended granular sludge blanket (EGSB), fluidized bed reactor (FBR), hybrid up-flow anaerobic sludge blanket (HUASB) reactor, modified internal circulation (MIC) anaerobic reactor, up-flow anaerobic bio-filter process (UABP), up-flow anaerobic filters (UAF), up-flow anaerobic hybrid reactors (UAHR), up-flow anaerobic sludge blanket (UASB), up-flow anaerobic stage reactor (UASR), and up-flow bio-electrochemical system (UBES). The treatment occur via some processes including stripper columns, coagulation, microfiltration, photocatalysis, sonolysis, Fenton/photo-Fenton process and reverse osmosis [8, 172, 195-196].

Commonly, WWTPs are not efficiently designed to remove micropollutants many of which are adsorbed to sludges, so that pharmaceutical drugs elimination is rarely complete [145]. But reactor technologies like ASBR, UASB, AMBR and MBBR are considered a new development even if they still have their challenges to effectively get rid of high resistant antibiotics for high biogas recovery [197].

Anaerobic fixed-film fixed-bed reactor is more efficient in removing organics and is best suited for pretreating herbal pharmaceutical effluent having high concentrations of organics [198]. ASBR (as shown in Figure 1) demonstrates a better treatment of pharmaceutical wastewater containing ≅ 40 mg/L of SMX; where above this limit, the reactor is disrupted leading to its failure, substrates utilization is inhibited and biogas production is affected [52]. Practically, Aydin et al. (2015)[95] studied the treatment of five synthetic pharmaceutical wastewater containing each of sulfamethoxazole–erythromycin–tetracycline (ETS), sulfamethoxazole–tetracycline (ST), erythromycin–sulfamethoxazole (ES) and erythromycin–tetracycline (ET) for 360 days inside an ASBR. UASB is basically a methane producing digester or an advanced type of clarigester that is widely applied owing to their low cost, ability to resist pH and temperature fluctuations, microbial diversity, and large biomass retention [162, 187]. Sreekanth et al. (2009)[128] studied the HUASB reactor using a terbinafine hydrochloride pharmaceutical effluent from Hyderabad with pH of 7-7.5 and BOD:COD ratio of 0.45-0.6. Likewise, the UAHR configuration combines the merits associated with both UAF and UASB, while limiting their disadvantage, making it effective in treating medium-to-high strength wastewater at high loading rates and short retention time [162]. UASR design separates the acidogenesis and methanogenesis stages, allows rapid recovery from hydraulic and organic shock loads, reduces biomass washout and needs no gas or sludge separation equipment, thereby demonstrating identity with ABR (consisting of UASB bioreactors connected in series) [52].

AMBR process amalgamates membrane separation and anaerobic biological treatment [111, 134, 160, 199]. Its benefits are low energy demand, less sludge production, enhanced biogas recovery, longer solid retention time, shorter hydraulic retention time, rapid start-up of microbial processes and a novel approach for treating effluents with variable flow, high COD and salinity content, suspended solids and inhibitory compounds (e.g., oil, fat and grease) [31, 73, 159]. But occasionally, different behaviors are noticed. For instance, even at high sludge retention period, reports presented by Cha et al. [142] shows that CBZ removal using the MBR technology was ineffective. A retention time of 435 days in an AMBR have also been reported – typically, pharmaceutical wastewater containing β-lactams antibiotics (BLAs) had been operated for 253 days in a 180 litres reactor [30, 159]. In another study, slow-growing anammox microorganisms was retained using an AMBR which boosted its activity 19 times [193]. Xiao et al. [200,201] investigate the removal of 5 PhACs in a bench-scale AMBR used to treat synthetic sludge. Pharmaceuticals like carbamazepine [13] and hydrochlorothiazide shows inefficient removal in activated sludge process (ASP) and MBR system. On the other hand, the advanced technology of the EMBRs are more efficient compared to ASPs and MBRs due to their low energy requirement [73]. Again, the novel ABEMR have superior merits over AMBR in terms of optimum organic removal and methane yield during pharmaceutical wastewater treatment [189]. However, problems with advanced technologies are limited applications (only at bench and pilot scales), membrane fouling, costly membrane materials and high energy needs [73].

EGSB is the modified but similar version of the UASB and UABP are used for treating high-strength pharmaceutical wastewater despite the problems associated with its use such as poor biomass retention, longer start-up caused by slow-growing microorganisms and inability to output solid-free effluent [162, 189]. Hu et al. 197] studied the performance of a novel UBES for the removal of β-lactams pharmaceutical effluents under different hydraulic retention time. The use of PAM as flocculant during wastewater removal from pharmaceutical factory workshop and after adding 200 mg/L of calcium chloride coagulant significantly neutralizes the effluent [143]. Though[105] affirms that electrocoagulation is 20 times more efficient compared to chemical coagulation. Previously, Mestre et al. shows that the novel powdered activated carbon and pine nut shell can be used to adsorb pharmaceutical compounds while Baaloudj et al. [202] proposes a technique for using sillenite ($Bi_{12}TiO_{20}$-BTO) catalyst to clean pharmaceutical wastewater composed of hazardous cephalosporin. Zero channeling of flow, compact bioreactor volume, recycle flow influent concentration dilution, uniform mixing resulting in high conversion rate, long biomass retention and high mass transfer rates are the advantages fluidized bed bioreactors offer during fungal treatment of wastewater [3, 38, 203].

A review on the use of conventional and nonconventional reactor systems to treat pharmaceutical effluents and produce biogas is shown in Table 1.

Medical Waste

Medical waste is basically all wastes generated from health facilities like dispensaries, medical centers and general hospitals, as a result of medical treatment, scientific research, diagnosis, immunization of humans or animals and/or therapeutic procedures such as injections, biopsy, resection of gangrenous organs, surgery, para-clinical exams, dialysis, delivery and autopsy [205–209]. The hospital could be tagged as an important source of pharmaceutical to the ecosystem which represents a small fraction of residue generated from typical municipality [147, 207, 210-211]. Medical wastes could be classified into infectious waste, hazardous waste, sharp waste and toxic waste which falls under two categories of risk and non-risk waste [7]. WHO report has it that 85% of hospital wastes are non-hazardous and the remaining 15% which is toxic is broken down into 10% hazardous (or infectious) waste, and about 5% non-infectious but hazardous (e.g., radioactive, chemical and pharmaceutical) [207, 212, 213]. But, based on literature report, other countries generate 10-25% of hazardous medical waste above the WHO report and 75-90% of such waste are non-hazardous [214, 215]. Longe & Williams

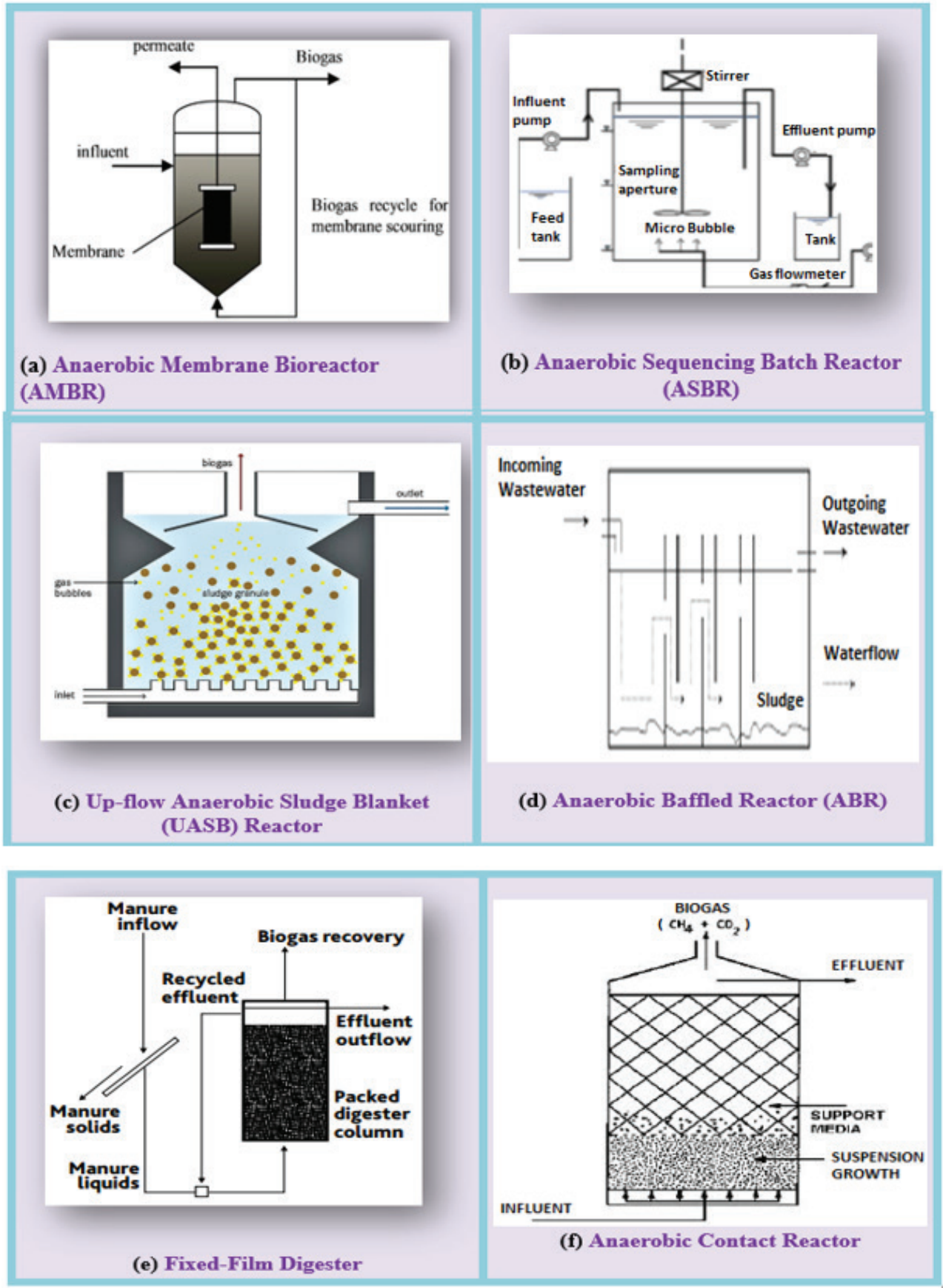


Figure 1. Schematic diagram of notable reactor types [From Chen & Neibling [22], with permission from Stanford.edu]

Table 1. Previous studies on different method of biogas production and treatment of pharmaceutical wastewater

Type of waste treated	Method	Biogas/ methane produced	Material & amount removed	Drawback	References
Pharmaceutical wastewater	Anaerobic bio-entrapped membrane reactor (AnBEMR)	Methane: 142.2±34.4 mL CH_4/ gCOD	5–10% higher TCOD removal efficiency than	Particular waste removed is not stated	[115]
Antibiotic	Thermal hydrolysis pretreatment and Anaerobic Digestion	Biogas: 289.90 mL gVS^{-1} untreated & 27.6% more after pretreatment	Spectinomycin mycelial residues of 265 mg L^{-1}	Serum bottles and not standard reactor types was used	[83]
Pharmaceutical waste fermentation broth	Mesophilic reactor	Methane: 22.5 L d^{-1} in 10 days only	Spent mycelia + Corn-grass silage + Pig slurry	Production of gas forced-stopped by an inhibitor	[49]
Pharmaceutical effluent	Anaerobic fluidized bed reactor	Biogas: 5.62 L/day	91.2 % COD removal	Particular effluent removed is not known	[204]
Antibiotic wastewater	Anaerobic multiphase hybrid reactor (AMHR)	Biogas: 8780 mL/ day	Sewage sludge + cow dung + Wastewater	Wastewater % removed is not mentioned	[191]
Pharmaceutical industrial effluent	Anaerobic Suspended Film Contact Reactor	Methane: 0.29-0.33 m^3 CH_4/ kgCOD	-	Type of effluent is not stated	[186]
Pharmaceutical effluent	Anaerobic batch reactor (bottles)	45-60 mL	Cow dung + Effluent	No effluent type or constituent and not a membrane system used	[58]

(2006)[216] enumerates the type of medical waste generated in Nigeria where they list waste human blood and products of blood, pathological wastes, body parts, and contaminated animal carcasses from autopsy or surgery as the most prevalent.

Medical waste generation in China was foretold to increase far above 50% by 2030 from a current yearly production of around 650,000 tons looking at a growth rate of 19-25%, largely because of huge population [217–220]. During the COVID-19 period, about 2,600 tons of biomedical waste was produced from 24 temporary quarantine facilities and households, 8 residential centers and 91 COVID-19 hospitals in South Korea [221]. New Delhi, the capital of the second most populous nation in the world, India, generates 65 tons of biomedical waste – where elsewhere in the country (Tamil Nadu), the massive nonmedical waste (food leftover, plastics and paper) generated are subject for research in the medical institution (Christian Medical College) in that part of the country, which is generating approximately 11,000 kg/day [212, 222]. In neighboring state within Pakistan, Peshawar, private hospitals produce 79 kg/day of medical waste, government teaching hospitals generates 900 kg/day, and government non-teaching hospitals produce 167 kg/day as reported by Giakoumakis et al. [223]. In Dhaka, Bangladesh, around 200 metric tons of medical waste (about 6% of overall waste) are produced daily – whereas on April 2020, the COVID-19 epidemic results in 14,500 tons of biomedical waste in the whole country [221, 224]. Annually, Malaysia is said to generate up to 33,000 tons of clinical waste [225]. In North America, Brazil, as estimated, 76% of the township dispose municipal and medical waste together in landfills; non-disposable medical devices (e.g., endoscopes) are reused in 41% of Canadian hospitals; and 2.5-3 million tons of biomedical waste is generated annually in the US [221, 226–228]. According to Chisholm et al. [229], Tunisia generates approximately 18,000 tonnes/year presently of medical waste, in addition to 8000 tonnes of hazardous waste. Morocco, Nigeria and majority of other African countries has no such information at the moment [226, 229].

Based on Chisholm et al. [229]'s assertion, approximately, 0.5 kg/bed/day of hazardous waste are produced in high-revenue countries, higher than 0.2 kg/bed/day rate of production in low-revenue countries. And according to Burik [213], a hospital bed has the capacity of generating 0.5 kg of waste per day, far below the rate in various countries as illustrated in Table 2.

Infectious and Sharp Waste

Infectious waste originates from infected patients in quarantine wards and are essentially cultures, contaminated blood, tissues, excreta and transmittable agents or

Table 2. Average daily weight of medical waste generated in hospitals across the world per bed

Country	Total medical waste (kg/bed/day)	Reference	Country	Total medical waste (kg/bed/day)	Reference
Algeria	1-1.72	[229]	Latvia	1.18	[223]
Bangladesh	0.57	[205]	Lebanon	2.45	[223]
Belgium	1.4	[223]	Libya	1.3	[223]
Brazil	4.4	[223]	Malaysia	1.9	[225]
Bolivia	0.5	[223]	Mexico	2-2.2	[221]
Bulgaria	2	[223]	Netherland	1.7	[223]
Canada	8.2	[223]	Nigeria	2.5	[223]
China	0.5-0.8	[230]	Norway	3.9	[223]
Ecuador	0.4	[223]	Pakistan	2-2.2	[223]
Egypt	0.7-1.7	[229]	Palestine	0.33-1.57	[205]
Ethiopia	1.1-6.03	[229]	Portugal	1.5	[223]
France	3.3	[223]	Serbia	1.9	[223]
Germany	3.6	[223]	Spain	4.4	[223]
Greece	1.4.-1.9	[230]	Sudan	0.38-0.87	[229]
India	0.5	[223]	Taiwan	1.9	[223]
Indonesia	2.23	[221]	Tanzania	0.84-5.8	[206]
Iran	3.5-4.42	[206]	Thailand	2.85	[221]
Ireland	5.03	[231]	Turkey	0.63	[230]
Italy	1-2.08	[232]	UK	3.3	[223]
Japan	2.3	[231]	USA	8.4	[223]
Jordan	3.49	[206]	Vietnam	1.57	[223]
Kazakhstan	5.34	[223]	Yemen	2.41	[223]
Kuwait	3.8	[223]			

equipment from laboratory tasks [207, 224, 231-233]. They results in diseases such as tetanus, tuberculosis, diarrhea, whooping cough and pneumonia that are spread by pathogens (e.g., streptococcal infections) due to inappropriate handling and waste management practices [209, 212, 234-235]. To arrest the dangers caused by open defecation, the Nigerian government in 2019, signed an executive order in that respect, as excreta could host harmful pathogens [149]. Most notable outbreak is the COVID-19 virus which affects approximately 75 million people worldwide as well as the Ebola virus [233, 236]. More than 50% of deaths is linked to health-care waste related activities and diseases according to Odumosu [7], surpassing the 1996 WHO number equivalent to almost 50,000 deaths globally [212, 235]. The largest deaths of infected humans were recorded in 2014, where Ebola took the life of 8 people in Nigeria and hundreds more in Africa – preceded by the COVID-19 virus that kills 1.66 million across the globe in 2020 [233, 236]. Due to rapid industrialization, 3.4 million people die annually across the globe from waterborne diseases [48]. To avoid future occurrence of the menace caused by infectious waste, the presence of incinerators, just like the one at Al Shifa hospital (in Gaza Strip-Palestine) is a way of ensuring effective disposal of infectious toxic waste [205].

However, literature documentation of infectious waste generated is scanty, but there has been a report by Hassan & Shareefdeen [233] which puts the amount of infectious clinical waste produced in underdeveloped nations at 0.2 kg/day. This quantity falls short of amount from Quezon City in four specialty hospitals (namely, Philippine Children's Medical Center-PCMC, National Kidney and Transplant Institute-NKTI, Philippine Heart Center-PHC and Lung Center of the Philippines-LCP) with total capacities of 1009 beds and mean infectious waste per day of 579 kg [237]. Sharps are syringes, broken ampoules, surgical aids/instruments, hypodermic needles, needles, scalpels, knives, broken glass infusion sets, blades and glassware [1, 7, 224, 233]. As for sharps, 91 healthcare services in Southern Brazil based on a survey, only care about biomedical sharp waste management without any regards to other medical wastes [216, 238].

Pathological Waste and their AD

What constitutes pathological waste are organs, blood, tissues, human or animal body parts (also called anatomical

waste such as amputated arms or legs and placenta), body fluids, infected animal carcasses, and human fetuses [1, 205, 207, 224, 239-240]. They are products of medical and veterinary hospitals with disease-spreading capabilities, odor pollution and insects and rodents attraction [1, 241]. Presently, based on literature studies, the most used medical waste for biogas production is placenta. Placenta have been digested in Philippines, Tanzania and India to produce biogas. In his studies, Kabbashi et al. (2018)[209] found that the rate of production of waste blood from laboratories is 21.1% and that of expired blood bank is 54.9% in Khartoum hospital, in Sudan. But likened to placenta, blood is least exploited despite being a good feedstock for biogas production.

At Mwananyamala Referral Hospital in Dar es Salaam, Tanzania, a 38.5 m^3 bioreactor produces 800 litres of biogas in a 90 days long digestion of 46 placentas daily feed, weighing 560g each and 1700 litres of gas from food residues in a separate digestion reported by Kellner [242]. From the same hospital, Honest & Saria (2020)[240], found that co-digestion of 83.1±14.7 kg/day of food waste and 25.6±4.5 kg/day of placenta in a 32 m^3 reactor at a pH between 6.3-8.0 and temperature of 30.3°C produces 2.5 m^3/day of biogas after 18 weeks, at a pressure ranging from 5-33 kPa. Also, ten regional referral hospitals, namely, Mbeya, Geita, Dodoma, Ruvuma, Arusha, Iringa, Kigoma, Njombe, Manyara and Mwanza that will make use of food leftovers and placenta in fixed underground round tank to produce biogas, had been targeted by the Tanzanian government, in an attempt to replicate those in Ubungo Municipal Referral Hospital and the Dar es Salaam's Mwananyamala Regional Referral Hospital used for heating and cooking [241]. Moreso, a contract was awarded to MOCUBA Enterprises Company Limited to build a biogas facility at Sinza hospital in Dar es Salaam to produce biogas from placenta and other waste in 2020 [243], as shown in Figure 2.

Jared Escarpe, a Filipino, suggest the construction of biodigester at the Perpetual Succor Hospital, Cebu City in the Philippines for co-digestion of food, garden waste and placenta to generate methane for cooking and powering the hospital. The project was supported by United Nations Development Programme (UNDP) and coordinated by Health Care Without Harm (HCWH) [244]. The District Hospital of Kalikot aims at using pregnant women placenta and degradable waste by building a biogas plant that will yield the needed amount of gas to boil and purify water required by patients in the hospital [245]. In some hospitals, instead of pathological wastes, other non-medical organic waste generated inside the health facility are used to produce biogas. Typically, in 2007, the Newham General Hospital in UK began composting food waste, in accordance with Ignou et al. [227] report. Likewise, the Holy Family Hospital in Bandra, India converts kitchen waste in the facility to biogas (1 cylinder capacity) and biofertilizer using a model created by Kabir Udeshi [246]. First and foremost, unlike food waste, digestion of placenta takes

Figure 2. Biogas plants at two Tanzania hospitals [From Honest & Saria [240] & Kellner [243], with permission from JEP & GHW, respectively].

Table 3. Biogas from biodegradable hospital waste

References	Medical waste	Biogas volume	AD condition
Kellner (2019)[242]	Placenta	800 litres	90 days RT & 38.5 m^3 digester
Honest & Saria, (2020)[240]	Food waste + Placenta	2.5 m^3/day	32 m^3 digester; 6.3-8.0 pH; 30.3°C & 18 weeks RT
Dhakal et al., (2015)[247]	Food leftovers + Amputed pathological waste + Cow dung	5.78 m^3/day	21 m^3 digester
Rahman & Melville (2023)[248]	Hospital waste	62 m^3	500 kg Feed
Kabbashi et al. (2018)[209] & Mohammed et al. (2017)[207]	Red blood cell & Whole blood cell	304-3075 mL CH_4	Different samples and reacting conditions

AD: Anaerobic digestion.

longer fermentation time of around 180 days. Though 1 kg of placenta is equivalent to 31.3 liters of biogas above food sample yield [243]. Limited resources (Table 3) are found in the literature on biogas recovery from medical waste.

BIODEGRADABILITY OF ANTIBIOTICS CONTAMINATED MANURE

Since 1990s, the anaerobic digestion of synthetic drug based effluents and antibiotics had been studied by various researchers using sulphate reducing organisms, iron reducing organisms, methanogenic organisms and nitrate reducing organisms [204, 249]. Before the AD of waste of pharmaceutical origin, their toxicity and biodegradability should be determined in advance for appropriate selection of a suitable operational parameters, as some antibiotics promote AD performance at certain conditions [131, 250-251]. In essence, low drugs concentration positively inhibits biogas yield whereas high drugs concentration stimulates the process [145]. So, it affirms the reason why some antibiotics, PCPs and chemical/pharmaceutical wastewaters are inhibitory, toxic and hard or partially degraded as they affect methanogen's growth and physiology and the efficiency of the process [131, 252, 253]. Typically, carbamazepine, diclofenac and tramadol, based on the SMA, is likely to be stimulating at low concentration while carbamazepine at high concentration – bearing similarity with ciprofloxacin which can either be stimulatory and non-stimulatory at high or low amounts [93, 145]. Such inhibitory properties of antibiotics on methane and biogas synthesis are due to pH reduction, dissolved COD and volatile fatty acids (VFAs) accumulation; such as large accumulation of VFA during AD of 45-50 mg/L of SMX [93, 250].

For example, antibiotics like chloramphenicol, neomycin, tylosin, doxycycline, chlortetracycline and streptomycin caused slight inhibitory effect on CH_4 production while erythromycin concentration of up to 250 mg/L, linear alkylbenzene sulfonate (LAS) and carbamazepine (CBZ) did not inhibit biogas generation according to literature [111, 142, 249]. Similarly, amoxicillin (60 mg/L) reduces the cumulative biogas and CH_4 production by 25%; tetracycline and sulfamethoxazole mixture (0.5–15 mg/L) and penicillin inhibits AD by 25% and 35% respectively in an ASBR; chlortetracycline reduces CH_4 production by 20% when fed to pig and; oxytetracycline strongly inhibits AD completely, except at concentrations around 250 mg/L in some cases [90, 102, 110, 141, 249]. On the other hand, Azizan et al. (2021)[141] found that an 8.5 mg/L concentration of tetracycline in pharmaceutical effluents totally inhibit CH_4 production. Stergar & Konèan [251] states that, equipment selection for anaerobic treatment of certain waste depends on the nature of the substrate and the process limiting steps. Sulfamerazine in pharmaceutical wastewater, as an example was found to be biodegradable and only inhibit methanogenesis at threshold concentration of 90 mg/L in a UASB bioreactor as reported by Li et al. [111]. Also, heavy metals and pharmaceuticals removal from pharmaceutical effluent sludge using steam explosion pretreatment method was found to improve its biodegradability and biogas yield (up to 380 mL CH_4/gVS) when pretreated at 10 bar for 15 minutes [254].

The AD technology offers reliability in treating animal manure, tannery waste, textile, distillery, personal care products (PCP) and both low- and high-strength pharmaceutical industry wastes, as it needs low energy compared to aerobic process which offers a decrease in biomass yield [186, 189, 255–257]. To address the disparity, Iliopoulous et al. [156] explored the joint use of anaerobic MBBR and aerobic MBR for pharmaceuticals removal and municipal wastewater treatment. PCPs such as diclofenac and contraceptives in soils are non-biodegradable and affect live organisms in the ecosystem (the case of reproductive failure in fish caused by 17-α ethinylestradiol) [72, 78, 92, 258-259]. According to Omil, et al. (2007)[258] reports, AD of PCPs have earlier been studied in sewage sludge. Even though antibiotics in animal waste pose significant side effects during their digestion, AD remains the most efficient way of treating antibiotic-contaminated livestock wastes [250]. Antibiotics (e.g., ampicillin, oxytetracycline, tylosin, chlortetracycline, florfenicol and sulfamethazine) are often spotted in cow and pig manure [85, 102, 250]. Oxytetracycline and chlortetracycline in pig manure, for instance, decreases

CH_4 production during AD by 56-62% while at mesophilic conditions, oxytetracycline in cow manure reduces biogas production by 60% [250]. Gaballah et al. [260] mentioned that the veterinary antibiotics present sometimes in mixed form in manures, can be removed before its AD. Nesse et al. [261] reported that the edible mushroom grown on PPR-polluted biogas digestate led to the uptake of the pharmaceuticals present in it in low amounts.

Medical and Pharmaceutical Sludge Digestion

Literature studies shows that pretreated medical cotton industry wastes, after 90 days fermentation yields 26.916 mL/gVS and 51.622 mL/gVS of biogas at respective mesophilic and thermophilic temperature regimes [262]. Biogas plants in the past 10 years have been constructed near healthcare facilities in Madagascar, Nepal and Tanzania by the Health, Environment and Climate Action Foundation of Nepal and the HCWH [263]. Based on digestion carried out by Dhakal et al. [247] for 147 days in Bir Hospital (established in 1889), Kathmandu using a 21m^3 digester, 95 kg/day of mixed healthcare waste produced 5.78 m^3/day of biogas. In an experiment carried out by Mohammed et al. [207] and Kabbashi et al. [209], three Red Blood Cells (RBCs) samples (A: 446.4g RBC with TS = 22.4% operated at 31.57°C for 150 days produces 1470 mL of methane gas, B: 348.4g RBC with TS=28.7% operated at 29.44°C for 110 days produces 2625 mL of methane gas, and C: 348.4g RBC with TS=28.7% operated at 29.25°C for 90 days produces 304 mL of methane gas) and two whole blood samples (D: 274g RBC with TS=36.5% operated at 29.25°C for 20 days produces 3075 mL of methane gas and E: 274g RBC with TS=36.5% operated at 29.89°C for 20 days produces 2570 mL of methane gas) were used at Khartoum, National

Figure 3. Biogas plants for ad of medical and pharmaceutical waste/effluents [From Stringer, Avadhani & Rege; i.e., [263, 266-267], with permission from GPP, Embio Ltd. & IBA].

Centre for Energy Research (NCER) to assess the methane gas yields of biomedical wastes. Dextran pharmaceutical effluent released from Xinlun Pharmaceutical Factory in Jianyang City, China yields 215.85 L/kgCOD of biogas after 28 days in a 635.5 liters of plexiglass digester according to Xu et al. [30]'s findings. Huang et al. carried out AD of 3 kg of dextran contained in activated sludge at 21-33°C for 42 days yielding 201.77 litres (13.31 L/kgCOD) of biogas when there was no mixing and 381.60-455.05 litres (19.81-21.07 L/kgCOD) when mixing was involved in a 750 liters biodigester. Findings of Gustavsson et al. [264] shows that 180 mL/gVS of CH_4 is produced by digesting industrial sludge from Björkborn industrial area in Karlskoga polluted with explosives, nitroaromatic compounds and pharmaceutical residue – whereas co-digestion of the same sludge with oat gives 270 mL/gVS of CH_4. Meanwhile, Yin et al. [164] produced 499.46 mL/gTS of biogas from pharmaceutical sludge at 10.32 inoculum-to-substrate ratio. Adesina & Felix [98] reported that diclofenac was biodegraded only at sludge retention time of 8 days minimum. There was 20.8% improvements in CH_4 generated when anaerobic granular sludge and anaerobic suspended sludge was digested with 1000 mg/L LCM. Biogas yield and CH_4 content during norfloxacin and tetracycline distillery wastewater purifications are 191 cm^3/g and 254 cm^3/g and 70% and 67% respectively [87]. Currently, addition of zero valent iron (ZVI) and granular activated carbon (GAV) to improve AD of pharmaceutical wastewater had been studied [265]. Practical applications of medical and pharmaceutical digesters for biogas production are shown in Figure 3.

Embio Limited, India, converts molasses into drugs intermediate by fermentation which in turn generates substantial amount of waste with biogas potential. Thus, a fixed film digester anaerobically digest the 300 m^3/day of wastewater generated at 50-55°C (& pH=7.2), and yields 14500-15000 m^3/day of biogas which is rerouted to boilers through gas blowers to generate steam for reuse [266]. Also in Figure 1c, Anthem Biosciences Private Limited treats thick slurry fermentation waste broth (pharmaceutical effluent) from tetrose, peptone, soya flour and cultures as raw material to produce 2250-2500 m^3/day of biogas [267].

Biogas from Herbal Pharmaceutical Waste

India has witnessed rapid growth and development in the manufacture and use of herbal pharmaceutical products to treat diseases, while in Poland, the country is regarded as a major herbs producer in Europe between 2012- [192, 268]. Agricultural outputs such as herbs has also been used to formulate pharmaceutical products some of which (real herbs or herbal industry wastes and effluents) can be used to produce biogas and bio-slurry, as they are characterized by high organic content [269, 270]. *Justicia schimperiana* (JS), lemon balm and alder buckthorn are typical examples of herbs or herbal industry waste that has been used as biogas substrates as well as medicinal plants like *Glycyrrhiza glabra* (GG), Mentha, *Cuminum cyminum* (CC), Lavender and Arctium [268, 271-272]. Some of these pharmaceutical herbs are characterized with unpleasant smell and high composition of organic pollutants [192, 272]. GG and CC of 10% TS, produces 13471 mL and 13611 mL of biogas after digesting 250g each of the plants in batch reactors at 35°C, according to results obtained by Fardad et al. [271]. Findings from experiment conducted by Czubaszek (2019) [268], shows that alder buckthorn and lemon balm produces 386±33 and 461±23 NL $kgVS^{-1}$ of biogas within a 35 days retention period. Patel et al. [273] stated that the biogas potential of herbs is naturally favoured by their high organic matter content. Also, from the Laboratory of Ecotechnologies (the biggest biogas lab in Poland), biogas produced is 172.18 m^3/Mg FM (or 526.23 m^3/Mg TS) at pH of 5.4, from a herb of 32.70%TS and 90.13 %VS [274]. Peni et al. [41] recorded 169.4 NL/kgVS and 193.2 NL/kgVS of biomethane from the perennial plant called *Helianthus salicifolius* – from raw biomass and silage respectively.

Human Urine Digestion

Between 70-90% of antibiotics are usually released in their original form via urine and feces [275]. Human urine with pH ranging from 4.8-7.5 contains 99.6% water and 0.4% dry matter, in addition to 2% w/w urea that is readily hydrolyzed to CO_2 and NH_3 [21, 276]. According to Kim et al. (2020)[276], an average human produce 1.4 liters of urine that contributes 50% of phosphorus and 85% of nitrogen in domestic sewage. Accordingly, Haque & Haque [21] shows that mineral nitrogen fertilizer in Sweden have similar effect on plant growth compared to human urine. Mono- and co-digestion of urine with other substrate have also been studied by researchers. For instance, the ratio, "cow dung : human urine : water" (50:10:40) generates 43.42 litres of biogas while 1:1 ratio of human urine and cow dung produces 41.95 liters of biogas at 28-35°C [21]. In another studies, the addition of 250 mL urine to 100 g of boiled rice produced maximum CH_4 and biogas [277]. A practical approach towards generating electricity and cooking gas from biogas plants to be constructed by Lumos Laboratories Nigeria Limited was made by the Federal Government recently, which would be adopted at prisons across the state. The plants would use human urine at correctional centers in the country and was pioneered by Nwosu (Patent No. NG/P/20/2013/699) in 2014 [278].

Kinetics and Optimization of Biogas Production

In the literature, kinetics of biogas production from medical and pharmaceutical wastes is scanty. This study identifies the Monod, Haldane, Andrew, Modified Gompertz, modified Stover-Kincannon and Van der Meer and Heertjes kinetic models as some of the mathematical models for optimizing AD and biogas production associated with pharmaceutical wastewater as feedstock [191, 279–281]. The Andrews' kinetic model is a concentration-dependent process designed to biologically eliminate and reuse liquid toxic waste. The model is best suited for

pharmaceutical waste/wastewater digestion at low substrate concentration, because at high concentrations, toxicity and inhibitory effects are seriously felt [53]. In numerous occasions, there is no proportionate relationship between the inhibitory effect and the inhibitor concentration on the kinetics, hence a non-linear function (modified propionate and acetate consumption model) to study the inhibition of pharmaceuticals on both acetogenesis and methanogenesis can be used [70]. Amin et al. [90] examined the effect of high OTC loadings on acidogens and methanogens using Haldane inhibition kinetics, where they realized that an increase in toxicity at increasing loading rate signifies a decrease of the Haldane inhibition constant. Though the Monod equation does not explain possible inhibitions in AD systems involving pharmaceutical effluents, the model can be applied while studying the kinetics of biogas production using biodegradable pathological waste. Equation 1, 2, 3 and 4 are the Monod, Andrews', the propionate and acetate consumption rate and Haldane kinetic models [53, 70, 90]:

$$\mu = \mu_{max} \frac{S}{K_s+S} \quad (1)$$

$$\mu = \mu_{max} \frac{S}{K_s+S+\frac{S^2}{K_I}} \quad (2)$$

$$r_{modified} = r. \frac{1}{1+\left(\frac{I}{K_I}\right)^m} \quad (3)$$

$$\mu = \frac{\mu_{max}\, S}{(K_s+S)\left(1+\frac{S}{K_I}\right)} \quad (4)$$

where, μ = specific growth rate of biomass (/day), μ_{max} = maximum specific growth rate of biomass (/day), S = biodegradable pollutant (substrate) concentration (mg/L), K_S = half saturation coefficient (mg/L), K_I = inhibition coefficient (mg/L), $r_{modified}$ = propionate or acetates consumption rate, I = concentration of the pharmaceutical and m = constant representing the non-linear dependence of inhibition (when m = 1, the inhibition reduces to non-competitive type; when m < 1, inhibitory effect on the kinetics tends to be insensitive with inhibitor concentration and; when m > 1, inhibitory effect on the kinetics tends to be very sensitive with inhibitor concentration). Human and animal urine from patients at medical centres are typical example of biogas substrates containing inhibitors since they are released unchanged from the sick organism consuming pharmaceuticals. Optimization of biogas production from urine can be carried out using Equations 2 and 3 if the patient animal is known to consume antibiotics and other inhibitory drugs. Nevertheless, such studies is still lacking, kinetics of biogas production using non-contaminated urine has been analyzed by Sau et al. (2013)[277] using the Chen and Hashimoto model.

WASTE MANAGEMENT PRACTICE

In this work, three types of waste have been identified, namely, chemical, medical and pharmaceutical wastes. Before the detection of pharmaceuticals in the aquatic environment, they have been detected on land and studied since the 1970s [66]. In water bodies, many are described as pseudo-persistent, i.e., relatively having short half-lives in the ecosystem with notable effect to humans such as reproductive damage, cell inhibition and behavioral changes [92, 98]. However, they are unlikely to cause significant harm on land (waste of hospital, landfill, agricultural, graveyards and industrial pharmaceutical activities), because they occur at very low concentrations; based on few research conducted to ascertain their toxicity to the ecosystem [75, 92, 131]. Pharmaceutical waste are simply contaminated, prescribed, expired, spilt, vaccines, sera, and propriety drugs and unused pharmaceutical products which is no longer needed and so is disposed off [139, 207, 282]. Unused medications could cost billions of dollars globally, and according to a US report, more than 50% of patients stores unused and expired drugs in their homes which they end-up flushing down the toilet [98, 283]. While studying 427 households in Nigeria, Auta et al. [283] found that 94.1% of households keep unused medicines either for re-use or serve another sick household member with 'probably' a similar medical issue But elsewhere (Lagos), out of 376 healthcare workers, 40.4% returned unused drugs, 41.2% returned expired ones to manufacturers and 19.9% return it to sellers according to a survey carried out by Adesina & Felix [98]. Reasons resulting in unused drugs are change in treatment plan, recommended drug is inappropriate for the need, non-adherence to therapy, medications reaching expiry dates, supply of unwanted quantities, adverse drug event, over prescribing, and nonrecognition of drugs due to a foreign label [283, 284]. In Lagos, Nigeria, absence of proper disposal guidelines for unused medication according to global best practices and a non-return policy (which is only 22.9% according to a US report but a predominant practice in Sweden) are challenges to address [98, 285].

Individuals and households, as a common practice, dispose medicines in garbage; especially in countries like Lithuania, United Kingdom and Kuwait [285]. A study of a pharmacy in Kuwait shows 73% of the respondents use trash to dispose unwanted medicines, 38.6% of doctors in South India dispose pharmaceuticals in dustbin and 24.6% flush them down the sink, a method commonly practiced in the US, and 59.7% of the 22 wards studied in Maiduguri, North-Eastern Nigeria dispose expired drugs using household garbage [98, 285]. Unprofessional disposal of expired pharmaceuticals (e.g., landfill and water bodies) had previously resulted in deaths and medication poisoning in scavengers, children and adult, whereas those emptied into

sewage systems kills bacteria responsible for its treatment [140, 284]. For example, waste drugs could be stolen from storage and sidetracked to the market for misuse and resale [284]. In Nigeria, non-prescribed drugs such as antihypertensives, analgesics, antimalarials, antibiotics, steroids, antacids, anticonvulsants, and antihistamines are highly consumed [66].

Management of chemical and biomedical waste is also a thing of concern. Chemical waste materials entails hormones, antioxidants, plasticizers, detergents, insecticides, pesticides, fire retardants, disinfectants, human and veterinary drugs that originates from homes, agricultural applications, dental, medical, veterinary laboratories and industries [7, 38, 89]. Biomedical waste as explained earlier, emanates as a result of diagnosis, immunization or treatment in medical or research facilities which are categorized into bio-hazardous and non-hazardous waste with characteristic high BOD content (about 234 ppm in Nigerian hospital sewage) [226, 286, 287]. For instance, about 26.5% of waste produced in 5 healthcare institutions in Nigeria's capital territory is hazardous medical waste based on a 2006 study [214]. Fundamental data regarding medical waste disposal emerging from hospitals is not available in Nigeria and the responsibility for its disposal is not assigned to anyone nor is it clearly defined [236, 283]. Except for Lagos in Nigeria, that at least constructed several highly-equipped transfer loading stations at strategic locations in the state [288], as shown in Figure 4.

Previously, moves have been made by the American Hospital Association (AHA) and the Environmental Protection Agency (EPA) in 1998 to reduce health care waste volume from 2005-2010, and in 1980 in Malaysia, clinical waste management began in the country [225, 286]. As a solution to menace posed by medical and pharmaceutical waste, Nyaga et al. [289] suggests the lessening of pharmaceutical waste generation, its collection at designated sites (e.g., Figure 2), implementation of a take-back options, application of recent technology and the lunching of public awareness campaigns together with policies and guidelines for a safe disposal. But in developing African countries, especially Nigeria, only few are well informed about the dangers posed by medical waste to the environment – specifically, in Tripoli, Libya, out of 300 medical waste handlers surveyed, only 7% had received training vis-à-vis its handling [228, 239]. In addition, zero training was given to medical waste personnel in Serbia while only 23% received such training in Bangladesh [228]. But surprisingly, in North-Eastern Nigeria, and based on Okoro & Peter [285]'s survey, more than 80% of the people knew the impacts of improper discarding of medicines to public well-being. Currently, a company based in Amsterdam (Pharmafilter), has developed a technology that digest biodegradable hospital waste anaerobically, thereby combating antibiotic resistance and in turn uses the biogas generated to power itself, as shown in Figure 5.

The use of the Pharmafilter technology would however be based on the time bound planning (scheduling) of various waste management steps of segregation, collection, storage, treatment to disposal [1, 229]. It is worthy of note that segregation is best practiced at place of waste

Figure 4. Special containers for loading medical waste at a transfer loading station in Lagos, Nigeria [From Awodele et al. [288], with permission from BMC Public Health].

Figure 5. Anaerobic digestion of hospital waste in Amsterdam, Netherland. [From Burik [213], with permission from LABIOTECH].

generation (usually the medical arena) by physicians, technicians and nurses [227]. In Africa, Indonesia and Korea, mixed waste containing hospital and municipal waste are commonly found in waste streams, bins, residential waste landfills and road sides [228, 288]. At present, Nigeria has no planned time scheduled healthcare waste management system and only 16.9% of health facility waste had been reported to be segregated in Lagos [229, 236]. Furthermore, syringes and other medical waste are regular contaminants of beaches of Bali in Indonesia [228]. Figure 6 shows medical waste being separated from non-medical waste at a hospital at Tamil Nadu, India.

Waste management practices for medical, pharmaceutical wastes and pharmaceutical process wastewater are inertization, deep burial, waste immobilization, microwaving, encapsulation, secure land filling, burning and incineration [98, 240, 290]. Honkanen [291] carried out a feasibility assessment on incenerating medical waste in South-Asia.

Figure 6. Segregation of medical from nonmedical waste at Christian Medical College, Vellore, Tamil Nadu, India [From Thomas et al. [153], with permission from IJERT]

Incineration is a high temperature (600-1200°C) thermal process that combust waste to gases and inert materials [7, 238]. Controlled air, rotary kiln and excess air incinerators are the three main kinds of incinerators [208]. Novel incinerator technologies are foundries or coal-fired thermal power stations and cement kilns (suited for expired pharmaceuticals) that disperse unwanted gases through tall chimneys [208, 284]. Sometimes these incinerators are made of bricks, especially the De Monfort incinerators [292]. Open burning at low temperatures and incineration release carcinogenic toxic chemical products called furans and dioxins that pollutes the air, thereby causing respiratory illnesses and cancer, as well as residual ashes [7, 209, 212, 284]. Apart from residual ashes, dioxins and furans, incinerators are sometimes left with needles, blades, heavy metals, polycyclic aromatic hydrocarbons (PAHs) and other sharps [3, 209, 238]. China has the highest dioxin emission threshold (5 ng/m^3) than the majority of European nations [209].

Nevertheless, incineration remains the most extensively used technique of treating hospital waste in numerous places around the world [7]. In Africa, Ethiopia, Botswana, Algeria and South Africa use incinerators: Tanzania in recent past constructed 13 pilot small scale incinerators [216]; in Ibadan, Nigeria, only few hospitals use incinerators [226]; and inadequate incineration practices in Tunisia is responsible for 90% of dioxins and furan emissions [229]. Six out of nine private healthcare facilities in Iran having incinerators, face mild operational problems; Korea is witnessing an insufficient controlled practices; incinerators in Malaysia are meant for hazardous and infectious waste treatment and; 67% of hospitals incinerate their infectious waste in the US [225, 226, 228, 288]. Hitherto, it took Eritrea 6 months to burn 7 truckloads of expired aspirin tablets; the act of burning unused and expired medicines at home is commonly practiced in Lithuania; while in Ibadan, Nigeria, approximately 30% of hospitals apply the open burning process [226, 284, 285].

CONCLUSION

Influence of wastes from pharmaceutical industries and medical health centers on the environment is a thing of concern to safeguard the habitat. Researchers are currently looking at ways to dispose or decompose these wastes and their effluents by transforming them to bioenergy or biofuel source by harnessing their unique characteristics. The market for pharmaceutical industry is growing and has witnessed skyrocketed increase in drug mass production due to recorded spread of infectious diseases occasioned by the famous Ebola and COVID-19 virus worldwide. This has also added to the amount of waste emerging from hospitals globally. Though not all these wastes are biodegradable due to their inhibitory or toxic nature, researchers and environmental management experts are constantly trying their luck to produce biogas from both pharmaceutical and medical waste, and had attempted to build biogas plant at close proximity to hospitals in some countries. However, incineration and open-burning remains the most applied waste management practice at the moment with attendant harmful effect to humans.

Construction of biogas plants is therefore encouraged to digest biodegradable pharmaceutical wastewater and medical waste, because it is safer than the combustion processes. Findings show that degradation of pharmaceutically contaminated effluents produces less biogas and requires longer retention times of more than 100 days due to the presence of negligible biodegradable matter or inhibitors. In Sudan, 1470-3075 mL of methane has been obtained from blood biodigestion. Around 2.5m^3 of biogas has been obtained by degrading placenta at Mwananyamala hospital in Tanzania and 5.78m^3/day of biogas has been obtained through the decomposition of pathological waste near Bir Hospital, Nepal. Existing gaps that need to be filled is the practical and purposeful construction of a multipurpose bioreactors (either conventional or nonconventional/membrane reactors) for large scale treatment of pharmaceutical, medical and associated effluent with a biogas and biofertilizer recovery system. Extensive review of literature has shown that such practices are often done at laboratory scale using nonconventional reactors with just few efforts made to recover the potential bioenergy inherent in the wastes. Separation technologies such as liquid-liquid extraction and distillation is recommended where the end-product of the wastewater treatment setup is to produce drinkable water. To sum it up, a biogas or bioenergy recovery system from pharmaceutical, herbal and medical waste has the potential to address environmental pollution, energy security, efficiency, health of (would be affected) humans and animals and unemployment challenges. Chemical or biochemical engineers and biotechnologist should therefore study the kinetics inhibiting anaerobic digestion of some pharmaceutical wastes/effluent in order to discover ways to optimize useful product recovery or suppress their effect during treatment.

ACKNOWLEDGEMENT

Part of this manuscript had been presented in AL-FARABI 4th International Congress on Applied Sciences (Erzurum, Türkiye) between August 19-20, 2022. The organizers are acknowledged for accepting the Abstract and providing the corresponding author the opportunity to present the write-up.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. These data are cited accordingly and can found amongst the list of bibliographies used.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Khobragade DS. Health care waste: Avoiding hazards to living and non living environment by efficient management. Fortune J Health Sci 2019;2:14-29. [CrossRef]

[2] Khaire KC, Maibam PD, Thakur A, Goyal A. Biomedical and pharmaceutical applications of xylan and its derivatives. In: Brienzo M, editor. Hemicellulose Biorefinery: A Sustainable Solution for Value Addition to Bio-Based Products and Bioenergy. Singapore: Springer; 2022. p. 447-465. [CrossRef]

[3] Odumosu BT. Biomedical waste: Its effects and safe disposal. In: Chandra R, editor. Environmental Waste Management. Florida: CRC Press; 2016. p. 81-93.

[4] Grabic R, Ivanová L, Kodešová R, Grabicová K, Vojs Staňová A, Imreová Z, et al. Desorption of pharmaceuticals and illicit drugs from different stabilized sludge types across pH. Water Res 2022;220:118651. [CrossRef]

[5] Farissi S, Ramesh S, Muthuchamy M, Muthukumar A. Biodegradation and photocatalysis of pharmaceuticals in wastewater. In: Shah MP, Rodriguez-Couto S, Kapoor RT, editors. Innovative Microbe-Based Applications for Removal of Chemicals and Metals in Wastewater Treatment Plants. Amsterdam, Netherlands: Elsevier; 2022. p. 69-97. [CrossRef]

[6] Golovko O, Ahrens L, Schelin J, Sörengård M, Bergstrand KJ, Asp H, et al. Organic micropollutants, heavy metals and pathogens in anaerobic digestate based on food waste. J Environ Manage 2022;313:114997. [CrossRef]

[7] Golovko O, Kaczmarek M, Asp H, Bergstrand KJ, Ahrens L, Hultberg M. Uptake of perfluoroalkyl substances, pharmaceuticals, and parabens by oyster mushrooms (Pleurotus ostreatus) and exposure risk in human consumption. Chemosphere 2022;291:132898. [CrossRef]

[8] Sharma N, Sharma S. A review on various treatment methods for treating pharmaceutical wastewater. Int Res J Eng Technol 2020;7:1406–1410.

[9] Ahmad S, Abbasi BK, Nazir MS, Abdullah MA. Metal organic frameworks (MOFs) as formidable candidate for pharmaceutical wastewater treatment. In: Lichtfouse E, Muthu SS, Khadir A, editors. Inorganic-Organic Composites for Water and Wastewater Treatment. Singapore: Springer; 2022. p. 37-63. [CrossRef]

[10] Gandhirajan M, Amarnath G, Kavitha P, Bhagavath R. Characterisation and treatment of pharmaceutical R&D wastewater. J Ind Pollut Control 2008;24:1-8.

[11] Mukesha P, Srinivasamurthyb S, Vigneshkumarb PS, Balamurugana P. A treatment of toxic substance in pharmaceutical industry wastewater: A review. Inform Technol Ind 2021;9:410-417. [CrossRef]

[12] Ismail ZZ, Habeeb AA. Pharmaceutical wastewater treatment associated with renewable energy generation in microbial fuel cell based on mobilized electroactive biofilm on zeolite bearer. J Eng 2015;21:35-44. [CrossRef]

[13] Asplund K. Removal of pharmaceutical compounds by anaerobic digestion of sewage sludge. Dissertation. Florida: NOVA Univ; 2022.

[14] Moghaddam A, Khayatan D, Barzegar EF, Ranjbar R, Yazdanian M, Tahmasebi E, et al. Biodegradation of pharmaceutical compounds in industrial wastewater using biological treatment: A comprehensive overview. Int J Environ Sci Technol 2023;20:5659-5696. [CrossRef]

[15] Khodja M, Debih H, Lebtahi H, Amish MB. New HTHP fluid loss control agent for oil-based drilling fluid from pharmaceutical waste. Clean Eng Technol 2022;8:100476. [CrossRef]

[16] Chelliapan S, Sallis PJ. Application of anaerobic biotechnology for pharmaceutical wastewater treatment. Environ Manage Sustain Dev 2011;2:13-21.

[17] Zhong W, Li G, Gao Y, Li Z, Geng X, Li Y, et al. Enhanced biogas production from penicillin bacterial residue by thermal-alkaline pretreatment. Biotechnol Biotechnol Equip 2016;29:522-529. [CrossRef]

[18] Renita AA, Kumar PS, Srinivas S, Priyadharshini S, Karthika M. A review on analytical methods and treatment techniques of pharmaceutical wastewater. Desalin Water Treat 2017;87:160-178. [CrossRef]

[19] Zhao X, Chen H, Zheng Q, Liu J, Pan P, Xu G, et al. Thermo-economic analysis of a novel hydrogen production system using medical waste and biogas with zero carbon emission. Energy 2023;265:126333. [CrossRef]

[20] Gunnerson CG, Stuckey DC. Anaerobic digestion: Principles and practices for biogas systems. 49th ed. World Bank; 1986.

[21] Haque MS, Haque MN. Studies on the effect of urine on biogas production. Bangladesh J Sci Ind Res.2006;41:23-32. [CrossRef]

[22] Chen L, Neibling H. Anaerobic digestion basics. Available at: http://large.stanford.edu/courses/2017/ph240/huang1/docs/cis-1215.pdf. Accessed on Jul 2, 2024.

[23] Ertekin E. Effect of oxytetracycline on biogas production and microbial communities during anaerobic digestion of cow manure by fluorescence in situ hybridization and real time polymerase chain reaction. Master's Thesis. İstanbul: Boğaziçi Univ; 2011.
[24] Korbag I, Omer SMS, Boghazala H, Abusasiyah MAA. Recent advances of biogas production and future perspective. In: Abomohra AEF, Elsayed M, Qin Z, Ji H, Liu Z, editors. Biogas-Recent Advances and Integrated Approaches. IntechOpen; 2021. p. 1-41. [CrossRef]
[25] Luostarinen S, Normak A, Edström M. Overview of biogas technology. Available at: https://www.build-a-biogas-plant.com/PDF/baltic_manure_biogas_final_total.pdf. Accessed on Jul 2, 2024.
[26] Mingchai C, Sangmane P. Decision process for adoption of biogas technology for the sustainable development in Uttaradit Province, Thailand. World Appl Sci J 2012;19:699-703.
[27] Raja IA, Wazir S. Biogas production: The fundamental processes. Univers J Eng Sci 2017;5:29-37. [CrossRef]
[28] Godhole A, Wadetwar RN, Lawal TO, Mahady GB, Raut NA. Microbiology of waste treatment. In: Raut NA, Kokare D, Bhanvase BA, Randive KR, Dhoble SJ, editor. 360-Degree Waste Management: Fundamentals, Agricultural and Domestic Waste, and Remediation. 1st ed. Amsterdam, Netherlands: Elsevier; 2023. p. 185-211. [CrossRef]
[29] Huang R, Mei Z, Long Y, Xiong X, Wang J, Guo T, et al. Impact of optimized flow pattern on pollutant removal and biogas production rate using wastewater anaerobic fermentation. Bioresources 2015;10:4826-4842. [CrossRef]
[30] Xu L, Yang L, Guo S, Zhou J, Luo T, Ran Y, et al. Experimental and CFD simulation study on anaerobic digestion using dextran pharmaceutical wastewater based on cyclic fluidization hydraulic mixing. Environ Prog Sustain Energy 2021;40:13656. [CrossRef]
[31] Fakhri H, Arabaci DN, Ovez S, Aydin S. *Eichhornia crassipes* root biomass to reduce antibiotic resistance dissemination and enhance biogas production of anaerobic membrane bioreactor. Environ Technol 2022;43:4168–4179. [CrossRef]
[32] Kovacs ED, Kovacs MH. Gas chromatographic: Mass spectrometric mining the volatilomes associated to Rhizobiota exposed to commonly used pharmaceuticals. In: Mendes KF, de Sousa RN, Mielke KC, editors. Biodegradation Technology of Organic and Inorganic Pollutants. IntechOpen; 2022.
[33] Díaz AH. Degradation of pharmaceutical compounds by microalgae: Photobioreactor wastewater treatment, biomass harvesting and methanization. Available at: https://www.tdx.cat/handle/10803/390962#page=1. Accessed on Jul 2, 2024.
[34] Goswami RK, Agrawal K, Verma P. An exploration of natural synergy using microalgae for the remediation of pharmaceuticals and xenobiotics in wastewater. Algal Res 2022;64:102703. [CrossRef]
[35] Hassan S, Meenatchi R, Pachillu K, Bansal S, Brindangnanam P, Arockiaraj J, et al. Identification and characterization of the novel bioactive compounds from microalgae and cyanobacteria for pharmaceutical and nutraceutical applications. J Basic Microbiol 2022;62:999–1029. [CrossRef]
[36] Chandel N, Ahuja V, Gurav R, Kumar V, Tyagi VK, Pugazhendhi A, et al. Progress in microalgal mediated bioremediation systems for the removal of antibiotics and pharmaceuticals from wastewater. Sci Total Environ 2022;825:153895. [CrossRef]
[37] Shashikant M, Bains A, Chawla P, Fogarasi M, Fogarasi S. The current status, bioactivity, food, and pharmaceutical approaches of *Calocybe indica*: A review. Antioxidants (Basel) 2022;11:1145. [CrossRef]
[38] Dalecka B. Wastewater treatment from pharmaceutical substances with filamentous fungi. Doctoral Thesis. Riga: Riga Technical Univ; 2021.
[39] Global Methane Initiative (GMI). The agricultural biogas plants in Poland. Available at: https://www.globalmethane.org/documents/Poland-Ag-Biogas-Plants-April-2014.pdf. Accessed on Jul 3, 2024.
[40] Rahmatzafran A, Rossle D, Rianawati E, Loeksmanto IH, Hilbert J, Alemmu S, et al. Biogas markets and frameworks in Argentina, Ethiopia, Ghana, Indonesia, and South Africa. Available at: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5d1e4e65c&appId=PPGMS. Accessed on Jul 2, 2024.
[41] Peni D, Marcin D, Stolarski MJ. Helianthus salicifolius as a new biomass source for biogas production. Energies 2022;15:2921. [CrossRef]
[42] Hasan G, Roubík H, Mazancova J, Banout J. Biogas energy potential in Syria: Prospects and challenges. Prague: Czech University of Life Sciences; 2016.
[43] Ergür HS, Okumus F. Cost and potential analysis of biogas in Eskisehir. Uludag Univ J Fac Eng 2010;15:155-160.
[44] van der Ann L, Reichel A. Bio-waste in Europe-turning challenges into opportunities. Available at: https://cdn.revolutionise.com.au/cups/bioenergy/files/rviur0yxx2psmjoi.pdf. Accessed on Jul 2, 2024.
[45] Gustafsson N. Biogas production based on park waste-Does Helsingborg have the potential? Master's theisis. Lund: Lund Univ; 2019.
[46] Gittelson P, Diamond D, Henning L, Payan M, Utesch L, Utesch N. et al. The false promises of biogas: Why biogas is an environmental justice issue. Environ Justice 2021;1-10. [CrossRef]
[47] Aydin S, Ince B, Cetecioglu Z, Ozbayram EG, Shahi A, Okay O, et al. Performance of anaerobic

sequencing batch reactor in the treatment of pharmaceutical wastewater containing erythromycin and sulfamethoxazole mixture. Water Sci Technol 2014;70:1625–1632. [CrossRef]

[48] Rana RS, Singh P, Kandari V, Singh R, Dobhal R, Gupta S. A review on characterization and bioremediation of pharmaceutical industries' wastewater: An Indian perspective. Appl Water Sci 2017;7:1-12. [CrossRef]

[49] Zupanèiè GD, Gotvajn AG. Anaerobic treatment of pharmaceutical waste fermentation broth. Chem Biomol Eng 2009;23:485-492.

[50] Chelliapan S, Wilby T, Sallis PJ. Performance of an up-flow anaerobic stage reactor (UASR) in the treatment of pharmaceutical wastewater containing macrolide antibiotics. Water Res 2006;40:507–516. [CrossRef]

[51] Gupta SK, Gupta SK, Hung YT. Treatment of pharmaceutical wastes. 2nd ed. Florida: CRC Press; 2004. p. 63-129. [CrossRef]

[52] Shi X, Leong KY, Ng HY. Anaerobic treatment of pharmaceutical wastewater: A critical review. Bioresour Technol 2017;245:1238–1244. [CrossRef]

[53] Hosseini AM, Bakos V, Jobbagy A, Tardy G, Mizsey P, Mako M, et al. Co-treatment and utilisation of liquid pharmaceutical wastes. Period Polytech 2011;55:3-10. [CrossRef]

[54] Mostofi A. Select issues in designing license contracts of strategic alliances in the pharmaceutical supply chain. Dissertation. Wellington: Victoria Univ; 2022.

[55] Enright AM, McHugh S, Collins G, O'FLaherty V. Low-temperature anaerobic biological treatment of solvent-containing pharmaceutical wastewater. Water Res 2005;39:4587–4596. [CrossRef]

[56] Sundararaman S, Sathiyapriya A. Acclimatization of an industrial pharmaceutical wastewater in an aerobic batch mode of operation. Int J Environ Res Dev 2016;6:1-10.

[57] Nguluka NC, Ochekpe NA, Odumosu PO. An assessment of pharmaceutical waste management in some Nigerian pharmaceutical industries. Afr J Biotechnol 2011;10:11259-11264. [CrossRef]

[58] Muruganandam B, Saravanane R, Lavanya M, Sivacoumar R. Effect of inoculum-substrate ratio on acclimatization of pharmaceutical effluent in an anaerobic batch reactor. J Environ Sci Eng 2008;50:191–196.

[59] Andersson S, Karlsson M. A comparative life cycle assessment of advanced processes for the removal of pharmaceutical residues in wastewater. Master's thesis. Gothenburg: Chalmers Unive; 2022.

[60] Pharmapproach. List of pharmaceutical companies in Nigeria. Available from: https://www.pharmapproach.com/list-of-pharmaceutical-companies-nigeria/4/. Accessed Jul 3, 2022.

[61] Zhu W, Bu F, Xu J, Wang Y, Xie L. Influence of lincomycin on anaerobic digestion: Sludge type, biogas generation, methanogenic pathway and resistance mechanism. Bioresour Technol 2021;329:124913. [CrossRef]

[62] Horner R. Global value chains, import orientation, and the state: South Africa's pharmaceutical industry. J Int Bus Policy 2022;5:68-87. [CrossRef]

[63] Njuguna AW, Mayabi AO, Ndirangu W. An investigation into the management of pharmaceutical wastewater in Kenya. In proceedings of the Sustainable Research and Innovation Conference; 2019 May 8–10; Kenya. 2019. p. 127-32.

[64] SaintyCo Pharma Process & Packaging. Top 200 pharmaceutical companies in South Africa. Available at: www.saintytec.com/pharmaceutical-companies-south-africa. Accessed on Jun 21, 2022.

[65] Justice. Pharmaceutical companies in Kenya (2022 list). Pharmchoices, Nairobi, Kenya; 2022. p. 1-18.

[66] Ogunbanwo OM, Kay P, Boxall AB, Wilkinson J, Sinclair CJ, Shabi RA, et al. High concentrations of pharmaceuticals in a nigerian river catchment. Environ Toxicol Chem 2022;41:551–558. [CrossRef]

[67] World Health Organization (WHO). Africa within the global pharmaceutical market. Available at: https://blog.private-sector-and-development.com/app/uploads/2019/02/PRO-Revue28-UK-key-figures.pdf. Accessed on Jul 2, 2024.

[68] Ussai S, Chillotti C, Stochino E, Deidda A, Ambu G, Anania L, et al. Building the momentum for a stronger pharmaceutical system in Africa. Int J Environ Res Public Health 2022;19:3313. [CrossRef]

[69] Holt T, Lahrichi M, Santos da Silva J. Africa: A continent of opportunity for pharma and patients. Available at: https://www.mckinsey.com/industries/life-sciences/our-insights/africa-a-continent-of-opportunity-for-pharma-and-patients. Accessed Jul 2, 2024.

[70] Fountoulakis MS, Stamatelatou K, Lyberatos G. The effect of pharmaceuticals on the kinetics of methanogenesis and acetogenesis. Bioresour Technol 2008;99:7083-7090. [CrossRef]

[71] Darwish M, Abuhabib AA, Mohammad H. Sustainable membranes with FNMs for pharmaceuticals and personal care products. In: Dutta S, Hussain CM, editor. Membranes with functionalized nanomaterials. 1st ed. Amsterdam, Netherlands: Elsevier; 2022. p. 275-328. [CrossRef]

[72] Fawzy ME, Abdelfattah I, Abuarab ME, Mostafa E, Aboelghait KM, El-Awady MH, et al. Sustainable approach for pharmaceutical wastewater treatment and reuse: Case study. J Environ Sci Technol 2018;11:209-219. [CrossRef]

[73] Mahtab MS, Farooqi IH. An overview of occurrence and removal of pharmaceuticals from sewage/

wastewater. In: Zhang T, editor. Sewage-Recent Advances, New Perspectives and Applications. IntechOpen; 2021.

[74] Mohan H, Rajput SS, Jadhav EB, Sankhla MS, Sonone SS, Jadhav S, et al. Ecotoxicity, occurrence, and removal of pharmaceuticals and illicit drugs from aquatic systems. Biointerface Res Appl Chem 2021;11:12530-12546. [CrossRef]

[75] Karungamye PN. Methods used for removal of pharmaceuticals from wastewater: A review. Appl J Environ Eng Sci 2020;6:412-428.

[76] xCarey DE, McNamara PJ. Altered antibiotic tolerance in anaerobic digesters acclimated to triclosan or triclocarban. Chemosphere 2016;163:22–26. [CrossRef]

[77] Bisognin RP, Wolff DB, Carissimi E, Prestes OD, Zanella R. Occurrence and fate of pharmaceuticals in effluent and sludge from a wastewater treatment plant in Brazil. Environ Technol 2021;42:2292–2303. [CrossRef]

[78] Frascaroli G, Reid D, Hunter C, Roberts J, Helwig K, Spencer J, et al. Pharmaceuticals in wastewater treatment plants: A systematic review on the substances of greatest concern responsible for the development of antimicrobial resistance. Appl Sci 2021;11:6670. [CrossRef]

[79] Luo X, Yang Q, Lin Y, Tang Z, Tomberlin JK, Liu W, et al. Black soldier fly larvae effectively degrade lincomycin from pharmaceutical industry wastes. J Environ Manage 2022;307:114539. [CrossRef]

[80] Campbell AJ. The behaviour of pharmaceuticals in anaerobic digester sludge. Master's thesis. Portsmouth: Univ of Portsmouth; 2013.

[81] Lankila A. Removal of pharmaceutical compounds by adsorption. Master's thesis. Lappeenranta: Lahti Univ; 2022.

[82] Zahedi S, Gros M, Balcazar JL, Petrovic M, Pijuan M. Assessing the occurrence of pharmaceuticals and antibiotic resistance genes during the anaerobic treatment of slaughterhouse wastewater at different temperatures. Sci Total Environ 2021;789:147910. [CrossRef]

[83] Song S, Jiang M, Yao J, Liu H, Dai X. Anaerobic digestion of spectinomycin mycelial residues pretreated by thermal hydrolysis: Removal of spectinomycin and enhancement of biogas production. Environ Sci Pollut Res Int 2020;27:39297–39307. [CrossRef]

[84] Cucina M, Zadra C, Marcotullio MC, Di Maria F, Sordi S, Curini M, et al. Recovery of energy and plant nutrients from a pharmaceutical organic waste derived from a fermentative biomass: Integration of anaerobic digestion and composting. J Environment Chem Eng 2017;5:3051–3057. [CrossRef]

[85] Mitchell SM, Ullman JL, Teel AL, Watts RJ, Frear C. The effects of the antibiotics ampicillin, florfenicol, sulfamethazine, and tylosin on biogas production and their degradation efficiency during anaerobic digestion. Bioresour Technol 2013;149:244–252. [CrossRef]

[86] Cetecioglu Z, Ince B, Gros M, Rodriguez-Mozaz S, Barceló D, Ince O, et al. Biodegradation and reversible inhibitory impact of sulfamethoxazole on the utilization of volatile fatty acids during anaerobic treatment of pharmaceutical industry wastewater. Sci Total Environ 2015;536:667–674. [CrossRef]

[87] Golub N, Ying Z, Kozlovets O, Levtun I, Ranra S. Wastewater purification from antibiotics with simultaneous biogas production. J Microbiol Biotechnol Food Sci 2020;10:170-175. [CrossRef]

[88] Lallai A, Mura G, Onnis N. The effects of certain antibiotics on biogas production in the anaerobic digestion of pig waste slurry. Bioresour Technol 2002;82:205–208. [CrossRef]

[89] Gadipelly C, Perez-Gonzalez A, Yadav GD, Ortiz I, Ibanez R, Rathod VK, et al. Pharmaceutical industry wastewater: Review of the technologies for water treatment and reuse. Ind Eng Chem Res 2014;53:11571-11592. [CrossRef]

[90] Amin MM, Hashemi H, Ebrahimi A, Ebrahimi A, Hashemi EH. Effects of oxytetracycline, tylosin, and amoxicillin antibiotics on specific methanogenic activity of anaerobic biomass. Int J Environ Health Eng 2012;1:1-37. [CrossRef]

[91] Koniuszewska I, Harnisz M, Korzeniewska E, Czatzkowska M, Jastzebski JP, Paukszto L, et al. The effect of antibiotics on mesophilic anaerobic digestion process of cattle manure. Energies 2021;14:1125. [CrossRef]

[92] Alenzi A, Hunter C, Spencer J, Roberts J, Craft J, Pahl O, et al. Pharmaceuticals effect and removal, at environmentally relevant concentrations, from sewage sludge during anaerobic digestion. Bioresour Technol 2021;319:124102. [CrossRef]

[93] Visca A, Barra Caracciolo A, Grenni P, Patrolecco L, Rauseo J, Massini G, et al. Anaerobic digestion and removal of sulfamethoxazole, enrofloxacin, ciprofloxacin and their antibiotic resistance genes in a full-scale biogas plant. Antibiotics (Basel) 2021;10:502. [CrossRef]

[94] Amin MM, Zilles JL, Greiner J, Charbonneau S, Raskin L, Morgenroth E. Influence of the antibiotic erythromycin on anaerobic treatment of a pharmaceutical wastewater. Environ Sci Technol 2006;40:3971–3977. [CrossRef]

[95] Aydin S, Ince B, Ince O. Inhibitory effect of erythromycin, tetracycline and sulfamethoxazole antibiotics on anaerobic treatment of a pharmaceutical wastewater. Water Sci Technol 2015;71:1620-1628. [CrossRef]

[96] Zeng S, Sun J, Chen Z, Xu Q, Wei W, Wang D, et al. The impact and fate of clarithromycin in anaerobic

digestion of waste activated sludge for biogas production. Environ Res 2021;195:110792. [CrossRef]
[97] Lehmann L, Bloem E. Antibiotic residues in substrates and output materials from biogas plants - Implications for agriculture. Chemosphere 2021;278:130425. [CrossRef]
[98] Adesina A, Sanni F. Pharmaceutical wastes management and the presence of pharmaceuticals in the environment of health facilities in Lagos state, Nigeria. Texila Int J Public Health 2018;6. [CrossRef]
[99] Sella CF, Carneiro RB, Sabatini CA, Sakamoto IK, Zaiat M. Can different inoculum sources influence the biodegradation of sulfamethoxazole antibiotic during anaerobic digestion? Braz J Chem Eng 2021;39:35–46. [CrossRef]
[100] Mahlaule-Glory LM, Mathobela S, Hintsho-Mbita NC. Biosynthesized bimetallic (ZnO-SnO2) nanoparticles for photocatalytic degradation of organic dyes and pharmaceutical pollutants. In: Pagano R, Valli L, Syrgiannis Z, editors. Catalysts 2022;12:334. [CrossRef]
[101] Caracciolo AB, Visca A, Massini G, Patrolecco L, Miritana VM, Grenni P. Environmental fate of antibiotics and resistance genes in livestock waste and digestate from biogas plants. Available at: https://kosmospublishers.com/environmental-fate-of-antibiotics-and-resistance-genes-in-livestock-waste-and-digestate-from-biogas-plants-2/. Accessed on Jul 3, 2024.
[102] Guo J, Ostermann A, Siemens J, Dong R, Clemens J. Short term effects of copper, sulfadiazine and difloxacin on the anaerobic digestion of pig manure at low organic loading rates. Waste Manag 2012;32:131–136. [CrossRef]
[103] Oliver JP, Gooch CA, Lansing S, Schueler J, Hurst JJ, Sassoubre L, et al. Invited review: Fate of antibiotic residues, antibiotic-resistant bacteria, and antibiotic resistance genes in US dairy manure management systems. J Dairy Sci 2020;103:1051–1071. [CrossRef]
[104] Yahia MB. An advanced physical modeling of adsorption mechanism of pharmaceutical compound on a biochar. AIP Adv 2022;12:035003. [CrossRef]
[105] Pavithra KG, Kumar PS, Rajan PS, Saravanan A, Naushad M. Sources and impacts of pharmaceutical components in wastewater and its treatment process: A review. Korean J Chem Eng 2017;34:2787-2805. [CrossRef]
[106] Bauer A, Amon T, Winckler C, Gans O, Scharf S. Effects of antibiotic residues in manure on biogas yield. Available at: https://boku.ac.at/fileadmin/data/H05000/H13000/Kooperation_BOKU-U/Poster_StratKoopBOKUU_antibiotics_2013_Bauer.pdf. Accessed on Jul 3, 2024.
[107] Chen Z, Wang H, Chen Z, Ren N, Wang A, Shi Y, et al. Performance and model of a full-scale up-flow anaerobic sludge blanket (UASB) to treat the pharmaceutical wastewater containing 6-APA and amoxicillin. J Hazard Mater 2011;185:905–913. [CrossRef]
[108] Saravanane R, Murthy DV, Krishnaiah K. Bioaugmentation and treatment of cephalexin drug-based pharmaceutical effluent in an upflow anaerobic fluidized bed system. Bioresour Technol 2001;76:279–281. [CrossRef]
[109] Sun H. Antibiotic resistance in biogas processes. Doctoral thesis. Uppsala: Swedish Univ; 2021.
[110] Nuengjamnong C, Rachdawong P, Chalermchaikit T. Effect of amoxicillin on biogas production and the Eschericia coli population in biogas systems treating swine wastewater. Thai J Vet Med 2010;40:57-62. [CrossRef]
[111] Li W, Qigui N, Hong Z, Zhe T, Yu Z, Yingxin G, et al. UASB treatment of chemical synthesis-based pharmaceutical wastewater containing rich organic sulfur compounds and sulfate and associated microbial characteristics. Chem Eng J 2015;260:55-63. [CrossRef]
[112] Gupta S, Chandra TS, Sharma A, Lokhande SK. Ozone-induced biodegradability enhancement and color reduction of a complex pharmaceutical effluent. J Int Ozone Assoc 2015;37:1-8. [CrossRef]
[113] Kayalvizhi N, Asha B. Role of volatile fatty acids in acidogenic and methanogenic reactor for treating pharmaceutical wastewater. Asian J Microbiol Biotechnol Environ Sci 2020;22:479-485.
[114] Kelbert M. Antineoplastic drugs: Effect of doxorubicin on enriched archaea culture from anaerobic digestion and potential degradation via an enzymatic process. Doctoral thesis. Florianópolis: Univ Federal de Santa Catarina; 2022.
[115] Ng KK, Shi X, Tang MKY, Ng HY. A novel application of anaerobic bio-entrapped membrane reactor for the treatment of chemical synthesis-based pharmaceutical wastewater. Sep Purif Technol 2014;132:634-643. [CrossRef]
[116] Pugazhendi A, Jamal MT, Al-Mur BA, Jeyakumar RB. Bioaugmentation of electrogenic halophiles in the treatment of pharmaceutical industrial wastewater and energy production in microbial fuel cell under saline condition. Chemosphere 2022;288:132515. [CrossRef]
[117] VEOLIA. Pharmaceutical manufacturing - Wastewater treatment guide. Available at: https://www.veoliawatertech.com/sites/g/files/dvc3601/files/document/2020/06/Veolia_Pharma_Guide_Wastewater_2020_HR_With_Links_0.pdf. Accessed Jul 3, 2024.
[118] Chittala G, Mogadati PS. Performance studies on a pharmaceutical wastewater treatment plant with a special reference to total dissolved solids removal. Int J Life Sci Biotechnol Pharma Res 2012;1:103-112.
[119] Hrenovic J, Stilinovic B, Dvoracek L. Use of prokaryotic and eukaryotic biotests to assess toxicity

of wastewater from pharmaceutical sources. Acta Chim Slov 2005;52:119-125.
[120] Menacherry SPM, Aravind UK, Aravindakumar CT. Oxidative degradation of pharmaceutical waste, theophylline, from natural environment. Atmosphere (Basel) 2022;13:835. [CrossRef]
[121] Chen YF, Ng WJ, Yap MGS. Performance of upflow anaerobic biofilter process in pharmaceutical wastewater treatment. Resour Conserv Recycl 1994;11:83-91. [CrossRef]
[122] Gulmez B, Ozturk I, Alp K, Arikan OA. Common anaerobic treatability of pharmaceutical and yeast industry wastewater. Water Sci Technol 1998;38:37-44. [CrossRef]
[123] Vijayan DS, Mohan A, Nivetha C, Sivakumar V, Devarajan P, Paulmakesh A, et al. Treatment of pharma effluent using anaerobic packed bed reactor. J Environ Public Health 2022;2022:4657628. [CrossRef]
[124] Ribeiro MHG, de A. Silva MC, Benetti AD. Anaerobic membrane bioreactor to remove pesticides and pharmaceuticals from wastewater: A bibliometric review. ICONASET-2023: Complexity and İmpact of Emerging Contaminants On Environment and Human Health, 2024. [CrossRef]
[125] Kumar V, Bansal V, Madhavan A, Kumar M, Sindhu R, Awasthi MK, et al. Active pharmaceutical ingredient (API) chemicals: A critical review of current biotechnological approaches. Bioengineered 2022;13:4309–4327. [CrossRef]
[126] Malmborg J, Magnér J. Pharmaceutical residues in sewage sludge: Effect of sanitization and anaerobic digestion. J Environ Manage 2015;153:1–10. [CrossRef]
[127] Mejías C, Martín J, Santos JL, Aparicio I, Alonso E. Occurrence of pharmaceuticals and their metabolites in sewage sludge and soil: A review on their distribution and environmental risk assessment. Trends Environ Anal Chem 2021;30:e00125. [CrossRef]
[128] Sreekanth D, Sivaramakrishna D, Himabindu V, Anjaneyulu Y. Thermophilic treatment of bulk drug pharmaceutical industrial wastewaters by using hybrid up flow anaerobic sludge blanket reactor. Bioresour Technol 2009;100:2534–2549. [CrossRef]
[129] Zhan H, Liang X, Wei Y, Zhuang X, Peng H, Zeng Z, et al. Utilization and valorization of pharmaceutical process residues: Current status and future trends. J Clean Prod 2024;438:140751. [CrossRef]
[130] Patel M, Kumar R, Kishor K, Mlsna T, Pittman CU Jr, Mohan D. Pharmaceuticals of emerging concern in aquatic systems: Chemistry, occurrence, effects, and removal methods. Chem Rev 2019;119:3510–3673. [CrossRef]
[131] Fountoulakis M, Drillia P, Stamatelatou K, Lyberatos G. Toxic effect of pharmaceuticals on methanogenesis. Water Sci Technol 2018;50:335-340. [CrossRef]
[132] Robertson KJ, Brar R, Randhawa P, Stark C, Baroutian S. Opportunities and challenges in waste management within the medicinal cannabis sector. Ind Crops Prod 2023;197:116639. [CrossRef]
[133] Wohde M, Berkner S, Junker T, Konradi S, Schwarz L, Düring RA. Occurrence and transformation of veterinary pharmaceuticals and biocides in manure: A literature review. Environ Sci Eur 2016;28:23. [CrossRef]
[134] Arcanjo GS, Dos Santos CR, Cavalcante BF, Moura GA, Ricci BC, Mounteer AH, et al. Improving biological removal of pharmaceutical active compounds and estrogenic activity in a mesophilic anaerobic osmotic membrane bioreactor treating municipal sewage. Chemosphere 2022;301:134716. [CrossRef]
[135] Reddy BV, Sandeep P, Ujwala P, Navaneetha K, Reddy KVR. Water treatment process in pharma industry: A review. Int J Pharm Biol Sci 2014;4:7-19.
[136] Lie M, Rubiyatno, Binhudayb FS, Thao NTT, Kristanti RA. Assessing the impact of pharmaceutical contamination in Malaysian groundwater: Risks, modelling, and remediation strategies. Trop Aquat Soil Pollut 2024;4:43-59, 2024. [CrossRef]
[137] Singh S, Pant A, Dutta K, Rani T, Vithanage M, Daverey A. Phytoremediation of pharmaceuticals and personal care products using the constructed wetland. Environ Chem Ecotoxicol 2024;6:104-116. [CrossRef]
[138] Abdelmigeed MO, Sadek AH, Ahmed TS. Novel easily separable core-shell Fe3O4/PVP/ZIF-8 nanostructure adsorbent: Optimization of phosphorus removal from fosfomycin pharmaceutical wastewater. RCS Adv 2022;12:12823-12842. [CrossRef]
[139] Javid F, Ang TN, Hanning S, Svirskis D, Burrell R, Taylor M, et al. Subcritical hydrothermal deconstruction of two hormones (adrenaline and progesterone) in pharmaceutical waste. J Supercrit Fluids 2022;179:105388. [CrossRef]
[140] Michael I, Ogbonna B, Sunday N, Anetoh M, Matthew O. Assessment of disposal practices of expired and unused medications among community pharmacies in Anambra state southeast Nigeria: A mixed study design. J Pharm Policy Pract 2019;12:12. [CrossRef]
[141] Azizan NAZ, Yuzir A, Abdullah N. Pharmaceutical compounds in anaerobic digestion: A review on the removals and effect to the process performance. J Environ Chem Eng 2021;9:105926. [CrossRef]
[142] Hao F. Research progress in pharmaceutical wastewater treatment technology. E3S Web Conf 2019;118:04019. [CrossRef]
[143] Cha YS, Yoon SU, Kim CG. Studies on influence and fate of carbamazepine in anaerobic digestion of sludge. J Environ Biol 2016;37:37–42.
[144] Vaughan M. uPOPs prevention and chemical awareness: Elements of a general awareness campaign. Available at: https://www.sprep.org/attachments/Reports/GEFPAS_Pollutant_Awareness_Camapign.pdf. Accessed on Jul 3, 2024.

[145] Šefčovičová K, Veronika K, Juraj M, Maros S, Igor B, Andrey K. Influence of selected pharmaceuticals on biogas production in mesophilic anaerobic fermentation. Res Pap 2021;29:149-157. [CrossRef]
[146] Klatte S, Schaefer HC, Hempel M. Pharmaceuticals in the environment-A short review on options to minimize the exposure of humans, animals and ecosystems. Sustain Chem Pharm 2016;5:61-66. [CrossRef]
[147] Colella K. Time trends of pharmaceuticals in wastewater treatment plant effluent with sources from pharmaceutical manufacturing facilities and hospitals. Master's thesis. New York: Stony Brook Univ; 2014.
[148] Danbauchi ES. Evaluation of lower Usuma Dam water quality for domestic supply (FCT) Abuja, Nigeria. Int J Res Sci Innov. 2020;7:219-224.
[149] Ilechukwu I. We found traces of drugs in a dam that supplies Nigeria's capital city. Available at: https://theconversation.com/we-found-traces-of-drugs-in-a-dam-that-supplies-nigerias-capital-city-161927. Accessed on Jul 3, 2024.
[150] Adegbe EA, . Pharmaceutical compounds in wastewater discharged from a University Teaching Hospital liquid waste treatment plant. Niger Res J Chem Sci 2019;7:256-261.
[151] Naveenkumar M, Anantharaj C, Porkodi N, Senthilkumar K, Nandakumar NP. A review on pharmaceutical wastewater treatment using biological process-benefits and opportunities. in Development in Wastewater Treatment Research and Processes-Emerging Technologies for Removal of Pharmaceuticals and Personal Care Products: State of the Art, Challenges and Future Perspectives, 2024, pp. 99-114. [CrossRef]
[152] Ameri B, Salah H. Analogous study of biogas production by anaerobic digestion of sewage treatment plant sludge , proposal of universal dimensionless models. Energy Sci Eng 2022;11;2366–2384. [CrossRef]
[153] Thomas A, et al. Treatment of pharmaceutical waste water treatment. Int J Eng Res Technol 2021;10:733-736.
[154] Mehmood T, Nadeem F, Bilal M, Meer B, Meer, K Qamar SA. Biological treatment of pharmaceutical wastes. In: Singh P, Verma P, Singh R, Ahamad A, Batalhao ACS, editors. Waste Management and Resource Recycling in the Developing World. 1st ed. Amsterdam, Netherlands: Elsevier. 2023. pp. 577-600. [CrossRef]
[155] Kamali M, Aminabhavi TM, Costa MEV, Islam SU, Appels L, Dewil R. Pharmaceutically active compounds in anaerobic digestion processes-Biodegradation and fate. In: Advanced Wastewater Treatment Technologies for the Removal of Pharmaceutically Active Compounds. Cham: Springer; 2023. pp. 91-106. [CrossRef]
[156] Iliopoulou A, Arvaniti OS, Deligiannis M, Gatidou G, Vyrides I, Fountoulakis MS, et al. Combined use of strictly anaerobic MBBR and aerobic MBR for municipal wastewater treatment and removal of pharmaceuticals. J Environ Manage 2023;343:118211. [CrossRef]
[157] Inanc B, Calli B, Alp K, Ciner F, Mertoglu B, Ozturk I. Toxicity assessment on combined biological treatment of pharmaceutical industry effluents. Water Sci Technol 2002;45:135–142. [CrossRef]
[158] Buitrón G, Melgoza RM, Jiménez L. Pharmaceutical wastewater treatment using an anaerobic/aerobic sequencing batch biofilter. J Environ Sci Health A Tox Hazard Subst Environ Eng 2003;38:2077–2088. [CrossRef]
[159] Huang B, Wang H, Cui D, Zhang B, Chen Z-B, Wang A-J. Treatment of pharmaceutical wastewater containing β-lactams antibiotics by a pilot-scale anaerobic membrane bioreactor (AnMBR). Chem Eng J 2018;341:238-247. [CrossRef]
[160] An Z, Junjie Z, Min Z, Yan Z, Xiaomei S, Hongjun L, et al. Anaerobic membrane bioreactor for the treatment of high-strength waste/wastewater: A critical review and update. Chem Eng J 2023;470:144322. [CrossRef]
[161] Schlott DA, Charbonneau SG, Greiner JA, Green RE, Quane DE, Robertson WM. Design, construction & start-up of an anaerobic treatment system for pharmaceutical wastewater. In proceedings of the 43rd Industrial Waste Conference; 1998 May 10–12; Indiana, USA. 1989.
[162] Murugesan MP, Akilamudhan P, Sureshkumar A, Arunkarhikeya G. Treatment of hospital and biomedical waste effluent using HUASB reactor Int J Innov Sci Res 2014;11:379-386.
[163] Li Y, Li C, Wang Z, Liu Y, Jia Y, Li F, et al. Navigating the complexity of pharmaceutical wastewater treatment by 'effective strategy, emerging technology, and sustainable solution. J Water Process Eng 2024;63:105404. [CrossRef]
[164] Yin F, Wang D, Li Z, Ohlsen T, Hartwig P, Czekalla S. Study on anaerobic digestion treatment of hazardous colistin sulphate contained pharmaceutical sludge. Bioresour Technol 2015;177:188–193. [CrossRef]
[165] Liu H, Xu G, Li G. Autocatalytic sludge pyrolysis by biochar derived from pharmaceutical sludge for biogas upgrading. Energy 2021;229:120802. [CrossRef]
[166] Ouyang J, Zhou L, Liu Z, Heng JYY, Chen W. Biomass-derived activated carbons for the removal of pharmaceutical micropollutants from wastewater: A review. Sep Purif Technol 2020;253:117536. [CrossRef]
[167] Saravanane R, Murthy DVS, Krishnaiah K. Assessment of toxicity and anaerobic degradation of anti-osmotic drug based pharmaceutical effluent in an upflow anaerobic fluidized bed system. Glob Nest Int J 2000;2:149-158. [CrossRef]

[168] Mantovani M, Rossi S, Ficara E, Collina E, Marazzi F, Lasagni M, et al. Removal of pharmaceutical compounds from the liquid phase of anaerobic sludge in a pilot-scale high-rate algae-bacteria pond. Sci Total Environ 2024;908:167881. [CrossRef]

[169] Kasulla S, Malik SJ. Unlocking the true energy potential of waste water treatment plants. Available at: https://www.brenstech.com/2023/10/03/unlocking-the-true-energy-potentials-of-waste-water-treatment-plants/. Accessed on Jul 3, 2024.

[170] Freitas RXA, Borges LA, de Souza HF, Colen F, Cangussu ASR, Sobrinho EM, et al. Characterization of the primary sludge from pharmaceutical industry effluents and final disposition. Processes 2019;7:231. [CrossRef]

[171] Ali Q, Zainab R, Badshah M, Sarwar W, Khan S, Mustafa G, et al. Prospecting the biodegradation of ciprofloxacin by Stutzerimonas stutzeri R2 and Exiguobacterium indicum strain R4 isolated from pharmaceutical wastewater. H_2Open J 2024;7:149-162. [CrossRef]

[172] Pandis PK, Kalogirou C, Kanellou E, Vaitsis C, Savvidou MG, Sourkouni G, et al. Key points of advanced oxidation processes (AOPs) for wastewater, organic pollutants and pharmaceutical waste treatment: A mini review. Chemengineering 2022;6:8. [CrossRef]

[173] Carballa M, Manterola G, Larrea L, Ternes T, Omil F, Lema JM. Influence of ozone pre-treatment on sludge anaerobic digestion: Removal of pharmaceutical and personal care products. Chemosphere 2007;67:1444–1452. [CrossRef]

[174] Wen S, Chen L, Li W, Ren H, Li K, Wu B, et al. Insight into the characteristics, removal, and toxicity of effluent organic matter from a pharmaceutical wastewater treatment plant during catalytic ozonation. Sci Rep 2018;8:9581. [CrossRef]

[175] Sörensen M, Zegenhagen F, Weckenmann J. State of the art wastewater treatment in pharmaceutical and chemical industry by advanced oxidation. Pharmind Prax 2015;77:594-607.

[176] Remya RR, Julius A, Suman TY, Mohanavel V, Karthick A, Pazhanimuthu C, et al. Role of nanoparticles in biodegradation and their importance in environmental and biomedical applications. J Nanomater 2022;2022:6090846. [CrossRef]

[177] Zhang J, Peng Y, Li X, Du R. Feasibility of partial-denitrification/ anammox for pharmaceutical wastewater treatment in a hybrid biofilm reactor. Water Res 2022;208:117856. [CrossRef]

[178] Maleki Shahraki Z, Mao X. Biochar application in biofiltration systems to remove nutrients, pathogens, and pharmaceutical and personal care products from wastewater. J Environ Qual 2022;51:129–151. [CrossRef]

[179] Ganesan S, Shanmugam S, Alagarasan JK, Lingassamy AP, Savunthari KV, Lo HM, et al. Novel African tulip fruit waste-derived biochar nanostructural materials for the removal of widespread pharmaceutical contaminant in wastewaters. Biomass Convers Bior 2023;13:13513–13525. [CrossRef]

[180] Ihsanullah I, Khan MT, Zubair M, Bilal M, Sajid M. Removal of pharmaceuticals from water using sewage sludge-derived biochar: A review. Chemosphere 2022;289:133196. [CrossRef]

[181] Sheng X, Wang J, Cui Q, Zhang W, Zhu X. A feasible biochar derived from biogas residue and its application in the efficient adsorption of tetracycline from an aqueous solution. Environ Res 2022;207:112175. [CrossRef]

[182] Al-Samrraie LA, Alrawashdeh KAB, Al-Issa HA, Shakhatreh S, Hussien AA, Qasem I. Improve heavy metals and pollutants removal from the pharmaceuticals wastewater using Washingtonia robusta: New extraction process. Civ Environ Eng 2022;18:340–349. [CrossRef]

[183] Madikizela LM, Pakade VE. Trends in removal of pharmaceuticals in contaminated water using waste coffee and tea-based materials with their derivatives. Water Environ Res 2023;95:e10857. [CrossRef]

[184] Al-Mashaqbeh O, Alsalhi L, Salaymeh L, Dotro G, Lyu T. Treatment of pharmaceutical industry wastewater for water reuse in Jordan using hybrid constructed wetlands. Sci Total Environ 2024;939:173634. [CrossRef]

[185] Gnanavel G, Muthusamy P. Pharmaceutical industry wastewater treatment using atmospheric air and pure oxygen. Int Acad Sci Eng Technol 2018;7:1-6.

[186] Mohan SV, Prakasham RS, Satyavathi B, Annapurna J, Ramakrishna SV. Biotreatability studies of pharmaceutical wastewater using an anaerobic suspended film contact reactor. Water Sci Technol 2001;43:271–276. [CrossRef]

[187] Chen Z, Wang Y, Li K, Zhou H. Effects of increasing organic loading rate on performance and microbial community shift of an up-flow anaerobic sludge blanket reactor treating diluted pharmaceutical wastewater. J Biosci Bioeng 2024;118:284-288. [CrossRef]

[188] Chen Z, Xu J, Hu D, Cui Y, Wu P, Hui Ge, et al. Performance and kinetic model of degradation on treating pharmaceutical solvent wastewater at psychrophilic condition by a pilot-scale anaerobic membrane bioreactor. Bioresour Technol 2018;269:319328. [CrossRef]

[189] Ng KK, Shi X, Ng HY. Evaluation of system performance and microbial communities of a bioaugmented anaerobic membrane bioreactor treating pharmaceutical wastewater. Water Res 2015;81:311–324. [CrossRef]

[190] Dutta K, Lee MY, Lai WW, Lee CH, Lin AY, Lin CF, et al. Removal of pharmaceuticals and organic matter from municipal wastewater using two-stage anaerobic fluidized membrane bioreactor. Bioresour Technol 2014;165:42–49. [CrossRef]

[191] Mullai P, Solaiappan V, Sabarathinam PL. Biogas production kinetics in an anaerobic multiphase hybrid reactor treating antibiotic industry wastewater. Desalin Water Treat 2018;122:247–253. [CrossRef]
[192] Nandy T, Kaul SN, Szpyrkowicz L. Treatment of herbal pharmaceutical wastewater with energy recovery. Int J Environ Stud 1998;54:83-105. [CrossRef]
[193] Svojitka J, Dvorak L, Studer M, Straub JO, Frömelt H, Wintgens T. Performance of an anaerobic membrane bioreactor for pharmaceutical waste-water treatment. Bioresour Technol 2017;229:180–189. [CrossRef]
[194] Wang KM, Zhou LX, Ji KF, Xu SN, Wang JD. Evaluation of a modified internal circulation (MIC) anaerobic reactor for real antibiotic pharmaceutical wastewater treatment: Process performance, microbial community and antibiotic resistance genes evolutions. J Water Process Eng 2022;48:102914. [CrossRef]
[195] Kumar P, Meena M, Kavar AB, Nama P, Pathak A, Varma R, et al. Experimental study to optimise the treatment efficacy of pharmaceutical effluents by combining electron beam irradiation with conventional techniques. Available at: https://arxiv.org/abs/2109.02479. Accessed on Jul 3, 2024.
[196] Kumar P, Mandal MK, Pal S, Chaudhuri H, Dubey KK. Membrane bioreactor for the treatment of emerging pharmaceutical compounds in a circular bioeconomy. In: Varjani S, Pandey A, Bhaskar T, Mohan SV, Tsang DCW, editors. Biomass, Biofuels, Biochemicals, Circular Bioeconomy: Technologies for Waste Remediation. Amsterdam, Netherlands: Elsevier; 2022. pp. 203-221. [CrossRef]
[197] Hu D, Min H, Chen Z, Zhao Y, Cui Y, Zou X, et al. Performance improvement and model of a bio-electrochemical system built-in up-flow anaerobic sludge blanket for treating β-lactams pharmaceutical wastewater under different hydraulic retention time. Water Res 2019;164:114915. [CrossRef]
[198] Nandy T, Kaul SN. Anaerobic pre-treatment of herbal-based pharmaceutical wastewater using fixed-film reactor with recourse to energy recovery. Water Res 2001;35:351–362. [CrossRef]
[199] Fazal S, Zhang B, Zhong Z, Gao L, Lu X. Membrane separation technology on pharmaceutical wastewater by using MBR (Membrane Bioreactor). J Environ Prot 2015;6:299-307. [CrossRef]
[200] Xiao Y, Hazarki Y, de Araujo C, Chun Chau S, Stuckey DC. Removal of selected pharmaceuticals in an anaerobic membrane bioreactor (AnMBR) with/without powdered activated carbon (PAC). Chem Eng J 2017;321:335–345. [CrossRef]
[201] Mestre AS, Viegas RMC, Mesquita E, Rosa MJ, Carvalho AP. Engineered pine nut shell derived activated carbons for improved removal of recalcitrant pharmaceuticals in urban wastewater treatment. J Hazard Mater 2022;437:129319. [CrossRef]
[202] Baaloudj O, Badawi AK, Hamza K, Yasmine B, Raouf H, Noureddine N, et al. Techno-economic studies for a pilot-scale Bi12TiO20 based photocatalytic system for pharmaceutical wastewater treatment: From laboratory studies to commercial-scale applications. J Water Process Eng 2022;48:102847. [CrossRef]
[203] Łubek-Nguyen A, Ziemichod W, Olech M. Application of enzyme-assisted extraction for the recovery of natural bioactive compounds for nutraceutical and pharmaceutical applications. Appl Sci 2022;12:3232. [CrossRef]
[204] Taylor P, Saravanane R, Murthy DVS, Krishnaiah K. Bioaugmentation and anaerobic treatment of pharmaceutical effluent in fluidized bed reactor. J Environ Sci Heal Part A Toxic/Hazardous Subst Environ Eng 2021;36:779-791. [CrossRef]
[205] Abu Mhady AI, Awad MA, Al-Aghah MR, El-Nahhal YZ. Assessment of medical waste Dehghani MH, Azam K, Changani F, Dehghani Fard E. Assessment of medical waste management in educational hospitals of Tehran university medical sciences. Iran J Environ Heal Sci Eng 2008;5:131-136.
[206] Dehghani MH, Azam K, Changani F, Fard ED. Assessment of medical waste management in educational hospitals of Tehran university medical sciences. Iran J Environ Heal Sci Eng 2008;5:131-136.
[207] Mohammed AMA, Kabbashi FMA, Hamad HK. Production of biogas from biomedical waste (blood). Master's thesis. Khartom: Sudan Univ; 2017.
[208] Health Care Without Harm. Non-incineration medical waste treatment technologies in Europe. Available at: https://www.env-health.org/IMG/pdf/altech_Europe_updated_version_10_12_2004.pdf. Accessed on Jul 3, 2024.
[209] Kabbashi FM, Hassan E. Methane production from biomedical waste (blood). Int J Energy Environ Eng 2018;12:642-649.
[210] Kularatne RKA. Biomedical waste generation at Ayurveda hospitals in South Asia: A mini review of the composition, quantities and characteristics. Waste Manag Res 2024;42:95–110. [CrossRef]
[211] Rana A, Sharma N, Hasan I. A review on hospital waste as a potential environmental pollution and their remediation mechanisms. AIP Conf Proc 2023;2535:020014. [CrossRef]
[212] Babu BR, Parende AK, Rajalakshmi R, Suriyakala P, Volga M. Management of biomedical waste in India and other countries: A review. J Int Environ Appl Sci 2009;4:65-78.
[213] Burik A. Microbes come to the rescue to reduce hospital waste. Available at: https://www.labiotech.eu/trends-news/pharmafilter-microbes-reduce-hospital-waste/. Accessed on Jul 3, 2024.
[214] Ejaeta O. National healthcare waste management plan. Federal Ministry of Health, Nigeria; 2008.

[215] Anitha J, Jayraaj IA. Isolation and identification of bacteria from biomedical waste (BMW). Int J Pharm Pharm Sci 2012;4:286-388.
[216] Longe EO, Williams A. A preliminary study of medical waste management in Lagos metropolis, Nigeria. Iran J Environ Heal Sci Eng 2006;3:133-139.
[217] Hou Y, Linlin J, Wenting M, Jian Li H. Analysing the factors affecting medical waste generation in China. Sustain Chem Pharm 2023;32:100975. [CrossRef]
[218] Wei Y, Cui M, Ye Z, Guo Q. Environmental challenges from the increasing medical waste since SARS outbreak. J Clean Prod 2021;291:125246. [CrossRef]
[219] Gao Q, Shi Y, Mo D, Nie J, Yang M, Rozelle S, et al. Medical waste management in three areas of rural China. PLoS One 2018;13:e0200889. [CrossRef]
[220] Zhimin M. Waste mismanagement: China's struggle with medical trash. Available at: https://www.wilsoncenter.org/publication/waste-mismanagement-chinas-struggle-medical-trash. Accessed on Jul 3, 2024.
[221] Marfe G, Perna S, Hermann A. Challenges in healthcare waste management of the UN 2030 agenda in the COVID-19 Era. Am J Environ Sci 2022;18:20-41. [CrossRef]
[222] Thomas TA. Recycling: Wealth from waste. Curr Med Issues 2017;15:252-256. [CrossRef]
[223] Giakoumakis G, Politi D, Sidiras D. Medical waste treatment technologies for energy, fuels, and materials production: A review. Energies 2021;14:8065. [CrossRef]
[224] Mohiuddin A. Medical waste: A nobody's responsibility after disposal. Int J Environ Sci Nat Resour 2018;15:45-51. [CrossRef]
[225] Yi TC, Jusoh MNH. Overview of clinical waste management in Malaysia. Front Water Environ 2021;1:47-57.
[226] Coker AO, Sangodoyin AY, Ogunlowo OO. Managing hospital wastes in Nigeria. Available at: https://wedc-knowledge.lboro.ac.uk/resources/conference/24/Coker.pdf. Accessed Jul 3, 2024.
[227] Ignou AA. Safe management of wastes from health-care activities, 2nd ed. Malta: World Health Organization, 2012.
[228] Stringer R. Medical waste and human rights. Available at: https://noharm-europe.org/sites/default/files/documents-files/1684/MedWaste_Human_Rights_Report.pdf. Accessed on Jul 3, 2024.
[229] Chisholm JM, Zamani R, Negm AM, Said N, Abdel Daiem MM, Dibaj M, et al. Sustainable waste management of medical waste in African developing countries: A narrative review. Waste Manag Res 2021;39:1149–1163. [CrossRef]
[230] Yong Z, Gang X, Guanxing W, Tao Z, Dawei J. Medical waste management in China: A case study of Nanjing. Waste Manag 2009;29:1376–1382. [CrossRef]
[231] Negishi R, Kawahara K. Infectious waste management in Japan: Assessment of current trends in waste measurement and reporting in general and psychiatric hospitals. J Mater Cycles Waste Manag 2022;25:421–429. [CrossRef]
[232] Giacchetta G, Marchetti B. Medical waste management: A case study in a small size hospital of central Italy. Strat Outsource Int J 2013;6:65-84. [CrossRef]
[233] Hassan MF, Shareefdeen Z. Recent developments in sustainable management of healthcare waste and treatment technologies. J Sustain Dev Energy Water Environ Syst 2022;10:1090384. [CrossRef]
[234] Dhanraj K. Perceptions of the pharmaceutical industry and regulators in South Africa towards registration harmonisation in the Southern African development community (SADC). Available at: https://etd.uwc.ac.za/handle/11394/7956. Accessed on Jul 3, 2024.
[235] Savitha KL, Joseph TJ. Efficiency of hospital waste management in Kerala: An analysis based on hospital ownership. Int J Res Anal Rev 2018;5:239-244.
[236] Ezirim I, Agbo F. Role of national policy in improving health care waste management in Nigeria. J Health Pollut 2018;8:180913. [CrossRef]
[237] Manegdeg F, Coronado LO, Paña R. Medical waste treatment and electricity generation using pyrolyzer-rankine cycle for specialty hospitals in Quezon city, Philippines. IOP Conf Ser 2020;463:012180. [CrossRef]
[238] Capoor MR, Bhowmik KT. Current perspectives on biomedical waste management: Rules, conventions and treatment technologies. Indian J Med Microbiol 2017;35:157–164. [CrossRef]
[239] Coker A, Sangodoyin A, Sridhar M, Booth C, Olomolaiye P, Hammond F. Medical waste management in Ibadan, Nigeria: Obstacles and prospects. Waste Manag 2009;29:804–811. [CrossRef]
[240] Honest A, Saria J. Performance of experimental bio-digestion for pathological and biodegradable waste management at Mwananyamala Regional Referral Hospital Tanzania. J Environ Prot 2020;11:838-847. [CrossRef]
[241] Songa SW. Placenta disposal to produce biogas in 10 referral hospitals. IPP Media, Dar es Salaam. 2022 May 12;1-5.
[242] Kellner C. Monitoring the placenta digester at Mwananyamala Referral Hospital; Dar es Salaam. 2019. Available at: http://greenhealthcarewaste.org/wp-content/uploads/2020/12/Tanzania-Monitoring-the-Placenta-Digester-at-Mwananyamala-Referral-Hospital.pdf. Accessed on Jul 3, 2024.
[243] Kellner C. Biogas plants at Sinza hospital; Dar es Salaam. Available at: https://greenhealthcarewaste.org/wp-content/uploads/2020/12/Tanzania-Report-Biogas-Plants-at-Sinza-Hospital.pdf. Accessed on Jul 3, 2024.

[244] Yeo S. Placenta used to generate clean energy in Filipino hospital. Available at: https://www.climatechangenews.com/2013/08/29/placenta-used-to-generate-clean-energy-in-filipino-hospital/. Accessed on Jul 3, 2024.
[245] Shahi PK. Kalikot hospital to run biogas plant. Available at: https://myrepublica.nagariknetwork.com/news/kalikot-hospital-to-run-biogas-plant/. Accessed Jul 3, 2024.
[246] The Indian Express (Express News Service). Holy family hospital starts Rs 13-lakh biogas plant to convert kitchen waste into gas for cooking purposes. Available at: https://indianexpress.com/article/cities/mumbai/holy-family-hospital-starts-rs-13-lakh-biogas-plant-5201852/. Accessed on Jul 3, 2024.
[247] Dhakal N, Karki AB, Nakarmi M. Waste to energy: Management of biodegradable healthcare waste through anaerobic digestion. Nepal J Sci Technol 2015;16:41-48. [CrossRef]
[248] Rahman KM, Melville L. An investigation into the conversion of non-hazardous medical wastes into biogas-A case study from the Health and Family Planning Sector in Bangladesh. Processes 2023;11:1494. [CrossRef]
[249] Fáberová M, Ivanová L, Szabová P, Štolcová M, Bodík I. The influence of selected pharmaceuticals on biogas production from laboratory and real anaerobic sludge. Environ Sci Pollut Res Int 2019;26:31846–31855. [CrossRef]
[250] Wang C, Jianfeng L, Qiumin L, Li H, Changmei W, Kai W, et al. A review of the effects of antibiotics on the anaerobic digestion of swine waste. Curr Opin Environ Sci Health 2021;25:100312. [CrossRef]
[251] Stergar V, Konèan JZ. The determination of anaerobic biodegradability of pharmaceutical waste using advanced bioassay technique. Chem Biomol Eng 2002;16:17-24.
[252] Díaz-Cubilla M, Letón P, Luna-Vázquez C, Marrón-Romera M, Boltes K. Effect of carbamazepine, ibuprofen, triclosan and sulfamethoxazole on anaerobic bioreactor performance: Combining cell damage, ecotoxicity and chemical information. Toxics 2022;10:42. [CrossRef]
[253] Nacheva PM, Peña-Loera B, Moralez-Guzmán F. Treatment of chemical-pharmaceutical wastewater in packed bed anaerobic reactors. Water Sci Technol 2006;54:157–163. [CrossRef]
[254] Aski AL, Borghei A, Zenouzi A, Ashrafi N, Taherzadeh MJ. Steam explosion pretreatment of sludge for pharmaceutical removal and heavy metal release to improve biodegradability and biogas production. Fermentation 2020;6:34. [CrossRef]
[255] Chen Z, Li X, Hu D, Cui Y, Gu F, Jia F, et al. Performance and methane fermentation characteristics of a pilot scale anaerobic membrane bioreactor (AnMBR) for treating pharmaceutical wastewater containing m-cresol (MC) and iso-propyl alcohol (IPA). Chemosphere 2018;206:750–758. [CrossRef]
[256] Gogoi M, Goswami R, Hazarika S. Membrane-based treatment of wastewater generated in pharmaceutical and textile industries for a sustainable environment. In: Verma S, Khan R, Mili M, Hashmi SAR, Srivastava AK, editors. Advanced Materials from Recycled Waste. 1st ed. Amsterdam, Netherlands: Elsevier; 2023. p. 87–109. [CrossRef]
[257] Ricky R, Shanthakumar S. Phycoremediation integrated approach for the removal of pharmaceuticals and personal care products from wastewater - A review. J Environ Manage 2022;302:113998. [CrossRef]
[258] Carballa M, Omil F, Ternes T, Lema JM. Fate of pharmaceutical and personal care products (PPCPs) during anaerobic digestion of sewage sludge. Water Res 2007;41:2139–2150. [CrossRef]
[259] Perez-Lemus N, Lopez-Serna R, Perez-Elvira SI, Barrado E. Analysis of 60 pharmaceuticals and personal care products in sewage sludge by ultra-high performance liquid chromatography and tandem mass spectroscopy. Microchem J 2022;175:107148. [CrossRef]
[260] Gaballah MS, Chand H, Guo J, Zhang C. Mixed veterinary antibiotics removal and effects on anaerobic digestion of animal wastes: Current practices and future perspectives. Chem Eng J 2024;483:149131. [CrossRef]
[261] Nesse AS, Jasinska A, Stoknes K, Aanrud SG, Ogner Risinggård K, Kallenborn R, et al. Low uptake of pharmaceuticals in edible mushrooms grown in polluted biogas digestate. Chemosphere 2024;351:141169. [CrossRef]
[262] Ismail ZZ, Talib AR. Recycled medical cotton industry waste as a source of biogas recovery. J Clean Prod 2016;112:4413–4418. [CrossRef]
[263] Stringer R. A win-win for disposing medical waste with biodigestion. Available at: https://www.greenpolicyplatform.org/blog/win-win-disposing-medical-waste-biodigestion. Accessed on Jul 3, 2024.
[264] Gustavsson LK, Heger S, Ejlertsson J, Ribe V, Hollert H, Keiter SH. Industrial sludge containing pharmaceutical residues and explosives alters inherent toxic properties when co-digested with oat and post-treated in reed beds. Environ Sci Eur 2014;26:8. [CrossRef]
[265] Dai C, Yang L, Wang J, Li D, Zhang Y, Zhou X. Enhancing anaerobic digestion of pharmaceutical industries wastewater with the composite addition of zero valent iron (ZVI) and granular activated carbon (GAC). Bioresour Technol 2022;346:126566. [CrossRef]
[266] Embio Limited. Energy recovery from pharmaceutical waste. Available at: https://www.mahaurja.com/meda/data/off_grid_bio_energy/Success%20Pharmaceutical%20Waste.pdf. Accessed on Jul 3, 2024.

[267] Krishna N. Case study-Biogas production from pharmaceutical waste. Available at: https://biogas-india.com/case-study-biogas-production-from-pharmaceutical-waste/. Accessed on Jul 3, 2024.
[268] Czubaszek R. The assessment of the suitability of lemon balm and alder buckthorn wastes for the biogas production. J Ecol Eng 2019;20:152-158. [CrossRef]
[269] Adetunji CO, Olaniyan OT, Anani OA, Bodunrinde RE, Osemweige OO, Ubi BE. Integrated processes for production of pharmaceutical products from agro-wastes. Biomass Biofuels Biochem; 2022;2022:439–461. [CrossRef]
[270] Sienkiewicz A, Piotrowska-Niczyporuk A, Bajguz A. Herbal industry wastes as potential materials for biofuel production. Proceedings 2020;51:6. [CrossRef]
[271] Fardad K, et al. Biodegradation of medicinal plants waste in an anaerobic digestion reactor for biogas production Document Other les. Comput Mater Contin 2018;55:318–392.
[272] Yitayal A, Mekibib D, Araya A. Study on biogas production potential of leaves of Justicia schimperiana and macro-nutrients on the slurry. Int J Waste Resour 2017;7:294.
[273] Patel S, Das P, Priyadarshi M, Babbar M, Hussain A, Bharat TV. Anaerobic digestion of herbal waste: A waste to energy option. Environ Monit Assess 2024;196:600. [CrossRef]
[274] Lewicki A, Piotrowska-Niczyporuk A, Bajguz A. The biogas production from herbs and waste from herbal industry. J Res Appl Agric Eng 2013;58:114-117.
[275] Zhang H, Yin M, Li S, Zhang S, Han G. The removal of erythromycin and its effects on anaerobic fermentation. Int J Environ Res Public Health 2022;19:7256. [CrossRef]
[276] Kim H, Choi H, Lee C. The potential use of human urine as a solvent for biogas upgrading. J Water Process Eng 2020;36:101343. [CrossRef]
[277] Sau SK, Mann TK, Giri A, Nandi PK. Effect of human urine during production of methane from boiled rice. Int J Sci Res 2013;2:60-64.
[278] The Guardian Newspapers. Generating electricity, cooking gas from urine, biodegradable waste. Available at: https://guardian.ng/generating-electricity-cooking-gas-from-urine-biodegradable-waste/#google_vignette. Accessed on Jul 3, 2024.
[279] Appala VNSG, Pandhare NN, Bajpai S. Mathematical models for optimization of anaerobic digestion and biogas production. In: Nandabalan YK, Garg VK, Labhsetwar NK, Singh A, editors. Zero Waste Biorefinery. Singapore: Springer; 2022. p. 575-591. [CrossRef]
[280] Rorke DCS, Lekha P, Kana GEB, Sithole BB. Effect of pharmaceutical wastewater as nitrogen source on the optimization of simultaneous saccharification and fermentation hydrogen production from paper mill sludge. 2022;25:100619. [CrossRef]
[281] Etheridge SP. Biogas applications. Available at: https://cetesb.sp.gov.br/biogas/wp-content/uploads/sites/3/2014/01/aplicacoes_do_biogas_na_europa_stephen_etheridge.pdf. Accessed on Jul 3,2024.
[282] Sapkota B, Pariatamby A. Pharmaceutical waste management system - Are the current techniques sustainable, eco-friendly and circular? A review. Waste Manag 2023;168:83–97. [CrossRef]
[283] Auta A, Omale S, Shalkur D, Abiodun AH. Unused medicines in Nigerian households: Types and disposal practices. J Pharmacol Pharmacother 2011;2:195–196. [CrossRef]
[284] WHO. Safe disposal of unwanted pharmaceuticals in and after emergencies. Pan Am J Public Heal 2000;7:205-208. [CrossRef]
[285] Okoro RN, Peter E. Household medicines disposal practices in Maiduguri, North-Eastern Nigeria. Int J Heal Life Sci 2019;6:e97085. [CrossRef]
[286] Gerwig K, Permanente K. Waste management & healthcare. Available at: https://noharm-global.org/hcwh-content-tags/health-care-waste-management?page=1. Accessed on Jul 3, 2024.
[287] Obayomi KS, Lau SY, Mayowa IE, Danquah MK, Zhang JC, Tung M, et al. Recent advances in graphene-derived materials for biomedical waste treatment. J Water Process Eng 2023;51:103440. [CrossRef]
[288] Awodele O, Adewoye AA, Oparah AC. Assessment of medical waste management in seven hospitals in Lagos, Nigeria. BMC Public Health 2016;16:269. [CrossRef]
[289] Nyaga MN, Nyagah DM, Njagi A. Pharmaceutical waste: Overview, management, and impact of improper disposal. Med Pharmacol. Available at: https://www.peerscientist.com/volume3/issue2/e1000028/pharmaceutical-waste-overview-management-and-impact-of-improper-disposal.pdf. Accessed on Jul 3, 2024.
[290] T. Honkanen, "Medical waste management in Thailand and Vietnam," Licentiate Master of Tech. Thesis on Sustainability Sciences, LUT School of Energy Systems, Lappeenranta-Lahti University of Technology (LUT), 2024.
[291] Tóth AJ, Gergely F, Mizsey P. Physicochemical treatment of pharmaceutical process wastewater: Distillation and membrane processes. Chem Eng 2011;55:59-67. [CrossRef]
[292] Clarke E, Hottor J. Health care waste management in Ghana-MOH policy and guidelines for health institutions. Available at: https://www.moh.gov.gh/wp-content/uploads/2016/02/Health-Care-Waste-Management-Policy-and-Guidelines.pdf. Accessed on Jul 3, 2024.

Sigma J Eng Nat Sci, Vol. 42, No. 4, pp. 1292–1314, August, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2023.00133

Review Article

Turkish sentiment analysis: A comprehensive review

Ayşe Berna ALTINEL GİRGİN[1,*], Gizem GÜMÜŞÇEKİÇÇİ[2], Nuri Can BİRDEMİR[1]

[1]*Department of Computer Engineering, Faculty of Technology, Marmara University, İstanbul, 34722, Türkiye*
[2]*Department of Computer Engineering, Işık University, İstanbul, 34980, Türkiye*

ARTICLE INFO

Article history
Received: 17 July 2023
Revised: 18 November 2023
Accepted: 29 December 2023

Keywords:
Deep Learning; Machine Learning; Natural Language Processing; Polarity Detection; Sentiment Analysis; Sentiment Classification

ABSTRACT

Sentiment analysis (SA) is a very popular research topic in the text mining field. SA is the process of textual mining in which the meaning of a text is detected and extracted. One of the key aspects of SA is to analyze the body of a text to determine its polarity to understand the opinions it expresses. Substantial amounts of data are produced by online resources such as social media sites, blogs, news sites, etc. Due to this reason, it is impossible to process all of this data without automated systems, which has contributed to the rise in popularity of SA in recent years. SA is considered to be extremely essential, mostly due to its ability to analyze mass opinions. SA, and Natural Language Processing (NLP) in particular, has become an overwhelmingly popular topic as social media usage has increased. The data collected from social media has sourced numerous different SA studies due to being versatile and accessible to the masses. This survey presents a comprehensive study categorizing past and present studies by their employed methodologies and levels of sentiment. In this survey, Turkish SA studies were categorized under three sections. These are Dictionary-based, Machine Learning-based, and Hybrid-based. Researchers can discover, compare, and analyze properties of different Turkish SA studies reviewed in this survey, as well as obtain information on the public dataset and the dictionaries used in the studies. The main purpose of this study is to combine Turkish SA approaches and methods while briefly explaining its concepts. This survey uniquely categorizes a large number of related articles and visualizes their properties. To the best of our knowledge, there is no such comprehensive and up-to-date survey that strictly covers Turkish SA which mainly concerns analysis of sentiment levels. Furthermore, this survey contributes to the literature due to its unique property of being the first of its kind.

Cite this article as: Altınel Girgin AB, Gümüşçekiççi G, Birdemir NC. Turkish sentiment analysis: A comprehensive review. Sigma J Eng Nat Sci 2024;42(4):1292–1314.

INTRODUCTION

Sentiment Analysis (SA), also referred to as Opinion Mining (OM), encompasses the process of contextually mining the text which includes detecting, identifying, and extracting properties [1]. To realize and automate sentiment analysis, NLP is used. Sentiment analysis has several applications. For instance, it can be used to determine the polarity of a text, meaning that it can be categorized as positive,

*Corresponding author.
*E-mail address: berna.altinel@marmara.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

neutral, or negative, and to identify individuals' opinions, attitudes, and emotions towards an entity or an event [2]. Since sentiment analysis is such a versatile tool, it can be applied in a variety of fields, including marketing, consumer information, politics, and social networks. An increasing amount of people have started to share their opinions and ideas about significant concepts, events, situations, etc. on social networks, which has led to the vast popularity of SA as a research topic [1]. Anecdotally, some of the most widely used social networks are Twitter, Facebook, Instagram, etc. Collecting data is a significant issue for every type of study. Studies that use more data provide more realistic and accurate results but finding proper data is generally challenging. But the increase in the usage of social networks provides significant sources of versatile data for sentiment analysis and this can be considered a reason behind its popularity. The general process of implementing a sentiment analysis model is given in Figure 1.

In this survey, studies about Turkish sentiment analysis are collected, analyzed, and summarized. All the collected studies are analyzed and summarized in a similar structure which consists of their approach, methodology, and performance. This survey categorizes studies considering the approaches employed and their level of sentiment analysis. There are three primary categories of sentiment analysis. These are Dictionary-based sentiment analysis, Machine Learning-based, and Hybrid-based sentiment analysis. This study also compares different types of SA studies and forms a table accordingly within each category based on the level of sentiment analysis, amount of data used and accuracy. The comparison tables are presented for each category. The advantage of these comparison tables are their efficiencies for presenting the main properties of studies in each category from different aspects. This survey also presents a list of public datasets that can be accessed and used for future sentiment analysis research. This list includes, the type of dataset, the year it was created, its language, size and source link can be found. Additionally, the list of the most used lexicons including their language, the size and information on sentiment polarity is also presented in this survey. This research can be useful for future researchers who are interested in Turkish Sentiment Analysis as it covers many different applications of Turkish Sentiment Analysis in one publication. This survey uniquely categorizes different approaches used in Turkish Sentiment Analysis. The contribution of this survey is significant for various reasons. Firstly, this survey reviews and summarizes a large number of previous and recent articles according to used

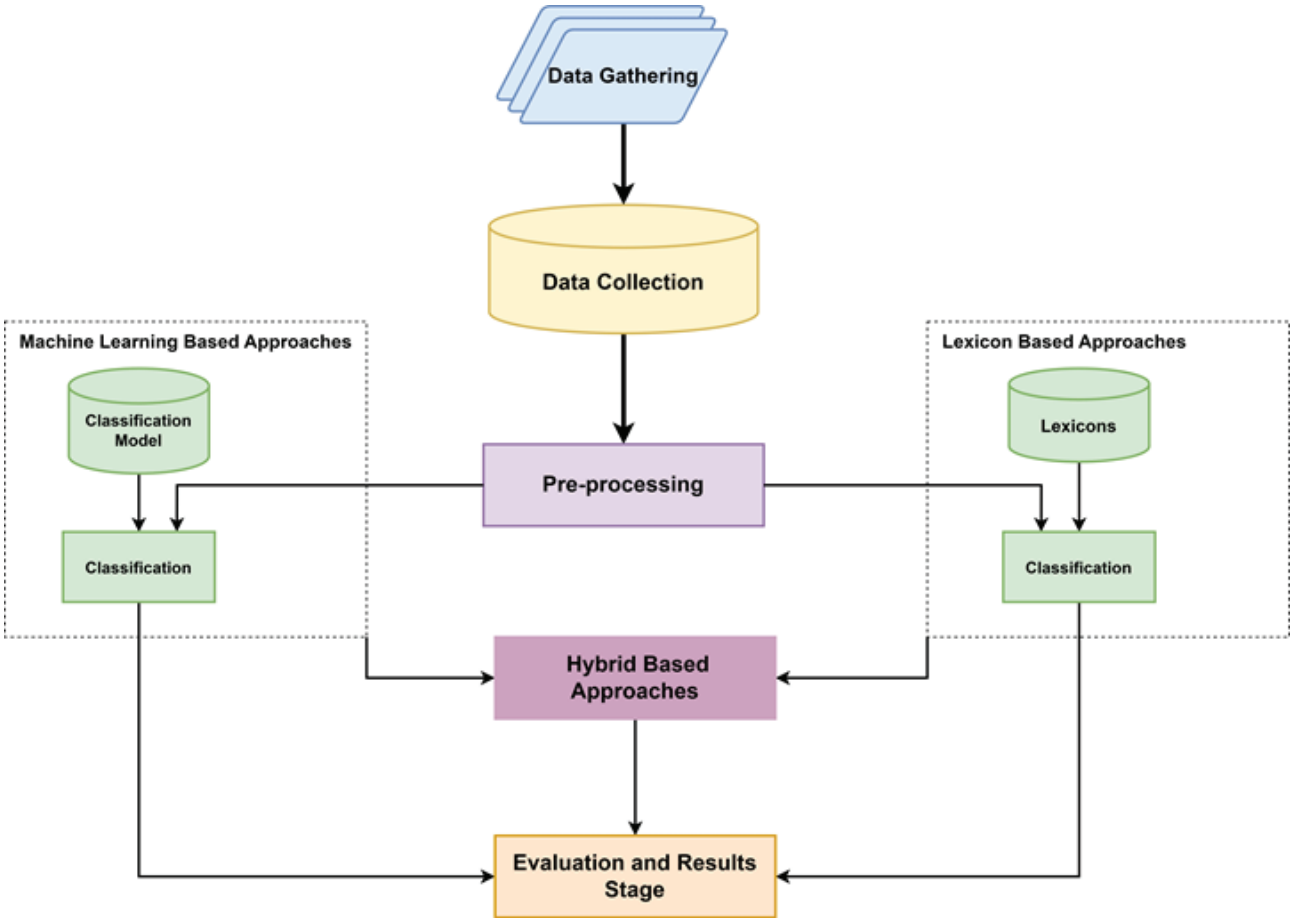


Figure 1. General Process of Turkish Sentiment Analysis.

approaches. This can help interested researchers to easily access the properties of various Turkish Sentiment Analysis studies from past to present and maybe choose the best approaches and techniques for specific research or applications. Second, comparisons of studies from each category are illustrated using tables. This helps visualize the different aspects of studies and allows the easy detection of studies' main features. Third, public datasets and the most frequently used lexicons are collected and listed, which can be accessed by interested researchers to source new research.

This survey is organized as follows: In subsection 1.2, the challenges of Turkish sentiment analysis are presented. In subsection 1.3, an overview of Turkish sentiment analysis is provided. In section 2, the studies using dictionary-based approaches in sentiment analysis are explained. In section 3, the studies using machine learning-based approaches in sentiment analysis are explained. In section 4, the studies using hybrid-based approaches in sentiment analysis are explained. In section 5, different methodologies for constructing a sentiment analysis system are explained. In section 6, the most popular dictionaries and datasets used in Turkish sentiment analysis are presented. In section 7, a general discussion is made about sentiment analysis, including the key points to consider when implementing a sentiment analysis system and current challenges in the field. In section 8, the conclusion of this research is presented.

The Challenges in Sentiment Analysis and Comparison of Turkish Sentiment Analysis with Different Languages

Since sentiment analysis is a language-dependent process, the degree of difficulty associated with performing sentiment analysis in different languages varies. Many linguistic or other types of issues can occur while performing sentiment analysis [3]. In sentiment analysis tasks, preprocessing techniques are generally applied to the data. In the preprocessing stage, the text can be normalized, stemmed, lemmatized, etc. The agglutinative structure of Turkish makes the preprocessing stages generally difficult which complicates Turkish sentiment analysis. In Turkish, the form of a word changes by attaching suffixes to the base (root) word which can change the semantic orientation of the word. This also creates additional challenges in Turkish sentiment analysis [4,5]. In contrast, the English language has a relatively low degree of complexity and inflection, which makes it easier to perform sentiment analysis. Due to the morphology of the Turkish language, creating or finding a proper sentiment lexicon that includes all variants of words can be impossible compared to other languages such as English. Turkish has a large number of unique words and idiomatic expressions that may not be found in other languages. This can make the sentiment analysis process more difficult since the polarities or meanings of these expressions cannot be extracted automatically [3]. Another challenge is that some Turkish characters do not exist in the English alphabet. The transformation of those characters adds difficulties to the sentiment analysis process [4,5,6]. Another important challenge in Turkish sentiment analysis is the limited resources in sentiment lexicons. There may be fewer annotated Turkish texts available for training and evaluating sentiment analysis algorithms compared to English texts [5,6,7]. In addition to the limited resources, the size of the already-limited Turkish lexicons is narrow. Since sentiment analysis is language-dependent, lexicons used in sentiment analysis differ by the language of the sentiment analysis. There are a wide variety of sentiment lexicons available for use in other languages, especially English [6,7]. Lastly, the way that people express sentiment and the words and phrases they use to do so can vary across cultures. Therefore, variations in the way sentiments are expressed can impact the methods of conducting sentiment analysis in different languages. Due to all of these reasons, Sentiment Analysis is most popularly employed in English due to its convenience and there is very limited research published on SA in other complex languages such as Turkish, French etc. [6]. To summarize, the most important challenges of performing Turkish Sentiment Analysis are;

- Language dependency of the sentiment analysis process.
- Language complexity of Turkish.
- Linguistic issues due to the structure of the Turkish language.
- Vocabulary and idiomatic expressions present in Turkish.
- Word Level Sentiment Analysis: Sentiment analysis is conducted on individual words.
- Limited resources in Turkish sentiment lexicons (Annotation availability).
- Narrow capacities of available sentiment lexicons in Turkish.
- Cultural differences between different languages.

The Overview of Turkish Sentiment Analysis

Sentiment Analysis can be performed on many different levels using many different approaches. The levels that the sentiment analysis is performed are aspect level, document level, sentence level and word level.

Aspect Level Sentiment Analysis: In aspect level sentiment analysis, keywords are chosen as entities and sentiment analysis is performed accordingly. Aspect-level sentiment analysis is also referred to as targeted sentiment analysis. Aspect-level sentiment analysis operates under the premise that sentiment is dependent on entities.

Document Level Sentiment Analysis: Document-level sentiment analysis involves analyzing the overall sentiment of an entire document, which may be composed of multiple sentences. The goal of this type of analysis is to understand the sentiment of the entire document.

Sentence and Document Level Sentiment Analysis: Sentence-level sentiment analysis involves analyzing the sentiment of a single sentence, which is made up of multiple words. The goal of this type of sentiment analysis is to understand the sentiment of the entire sentence.

Word Level Sentiment Analysis: Sentiment analysis is conducted on individual words. In Sentiment Analysis, different approaches can be used. These approaches are categorized as dictionary-based, machine learning (ML)-based, and hybrid-based.

Dictionary-Based Sentiment Analysis Approaches: An external lexicon is used to perform sentiment analysis.

ML-Based Sentiment Analysis Approaches: Supervised machine learning, unsupervised machine learning, or deep learning methods are used to perform sentiment analysis.

Hybrid-Based Sentiment Analysis Approaches: In hybrid approaches, instead of using one approach, combinations of different approaches are used.

The overview of sentiment analysis is presented in Figure 2.

TURKISH SENTIMENT ANALYSIS WITH DICTIONARY BASED APPROACHES

In this section, Turkish sentiment analysis studies that use dictionary-based approaches are collected, analyzed, and reviewed. Sentiment analysis can be realized at a variety of levels. These are word, sentence, aspect, and document levels. Regardless of the level of sentiment analysis, the dictionary-based approach relies on a sentiment lexicon and a collection of known and precompiled sentiment terms [8]. This means an external lexicon is utilized to perform the sentiment analysis. Generally, if the lexicon used contains polarity scores of words, the word polarities are used to calculate the sentiment value. The lexicons that contain information about word polarities are called polarity lexicons. Additionally, there are important usages of lexicons in sentiment analysis other than encompassing word polarities.

Aspect Level Sentiment Analysis

This section provides an evaluation of a study that applies dictionary-based approaches to perform an aspect-level Turkish SA, focusing on its approaches, methodology, dataset and performance. In aspect level SA, which is also referred to as targeted SA, special keywords are selected as entities from the given text and the SA is performed towards the targeted entities.

Dehkharghani et al. [3] studied SA in and Turkish sentiment analysis was performed at different granularity levels. In this study, a comprehensive sentiment analysis system was built for Turkish. In the study, a large dataset was used, which was collected from various Turkish movie review websites. The dataset that is used contains 60,000 documents. They only used a subset from the Turkish movie dataset. First, they started by manually labeling 2,700 sentences and 1,000 randomly selected documents. This label annotation process was performed manually by three people. They labeled the data as “positive”, “negative”, or ”neutral”. Following the process of labeling, the distribution of labels for sentences was 50% positive, 30% neutral, and 20% negative, and for documents, it was 52% positive, 29% neutral, and 19% negative. Since they made sentiment analysis

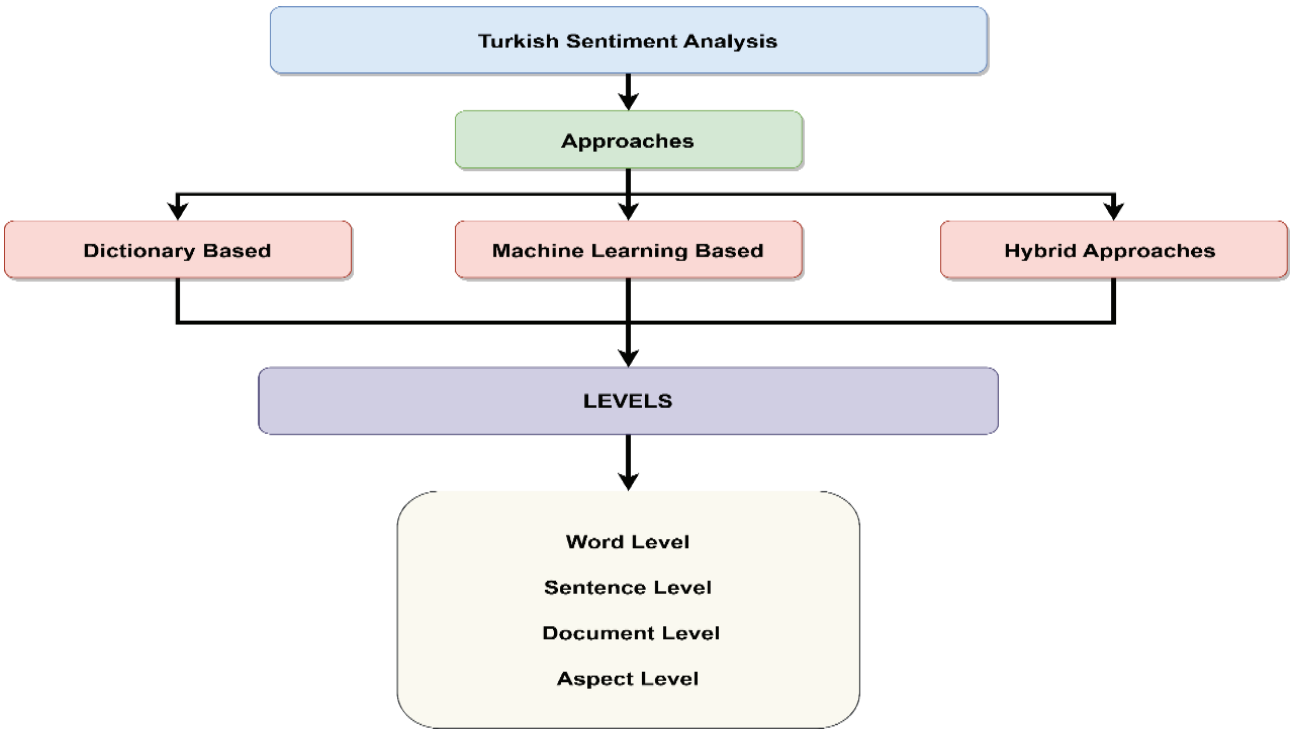


Figure 2. Overview of Turkish Sentiment Analysis.

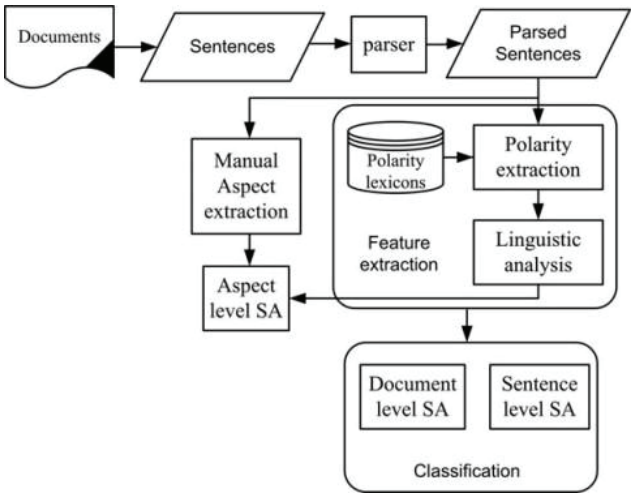


Figure 3. Overview of the system.

at different granularity levels, the proposed system has several different components. The different granularity levels are word level, aspect level, and lastly sentence and document levels. First, the document was segmented into sentences, then they used tokenization to parse the sentences. Following this process, they used a morphological analyzer tool for each word. At the end of this process, they assigned polarity scores to the words' n-grams which are unigram, bigram, and trigram. For the processes, they used different data and NLP tools. They used the ITU Turkish Parser for tokenization and morphological analysis. The SentiTurkNet lexicon was used for finding the polarity scores of words. For the classification, the Logistic Regression classification method was trained and used for different granularity levels. For sentence-level classification, 16 different features were used to train the classifier, for document classification, 20 features were extracted and used and for aspect- level classification, average polarity and number of tokens were used as features. Specifically, word polarities, polar words, sentence types, sentence polarities, emoticons, and linguistic issues such as the number of adjective verbs, number of initial capital words and number of domain-specific indicative terms were used as features. When the proposed system is applied to the chosen data, by utilizing all of the features, they obtained 73.42% and 79.06% accuracy in sentence and document classification. In future works, it will be possible to resolve every issue and sub-problem associated with Turkish sentiment analysis. Additionally, it can be attempted to extend the complexity of the proposed system in the future by exploring the phrase-level sentiment analysis thoroughly. Furthermore, the sentiment analysis process of the study [3] is given in Figure 3.

Sentence and Document Level Sentiment Analysis

In this section, studies on Turkish SA on the sentence-level are reviewed based on methodology, dataset, and performance. All the reviewed studies below use dictionary-based approaches. In sentence-level SA, the sentiment of sentences is extracted based on the information obtained from the whole sentence.

Another recent study had been proposed by Suat and Çınar in [9], wherein the relationships between news about companies and company values were analyzed in the year 2014 using text mining and sentiment analysis. Company publications, news in the media, and social media were used as data sources. In the study, it was stated that in the digitization of textual data, an approach based on sentiment dictionaries would be used to detect the sentiment contents of textual data. For the analysis stage of this study, a polarity lexicon was chosen and used to perform sentiment analysis. The analysis was performed at document-level. According to sentiment analysis, it has been detected that company values are also affected by perceptions and prejudices, unlike priorities such as market investment. This suggests that the quality with which work is accomplished in a company does not hold much significance on its own, without proper marketing and advertising, the real value and effects of the work cannot be reached within the company. No matter how well companies perform, if the performed work cannot be promoted effectively and supported by advertisements, it won't make a big impact and may be forgotten in a short amount of time.

In a new approach proposed by Dehkharghani et al. in [3], sentiment analysis was performed on different granularity levels. The levels that the sentiment analysis performed were aspect level, word level, sentence level, and document level. The polarity of the sentences was extracted from documents and classified with a lexicon-based approach. The details of this sentence-level polarity classification can be found under section 2.1.

The authors in [10] attempted to develop a sentiment lexicon using existing approaches. In order to create the sentiment lexicon, Twitter data was used to conduct sentiment analysis. Sentiment analysis was realized for two distinct topics; the first topic was the effects of weather change on people's feelings. The second topic was to analyze the feelings of people regarding a specific tv-show. To perform sentiment analysis on these two different topics, lexicon-based approaches were used. The details of this study can be found under section 5.

The authors presented a new framework for sentiment analysis classification in [11], using a relatively large Turkish movie site dataset that consists of 60k movie reviews. The dataset is collected from a website called Beyazperde. To create this framework, the authors have customized the SentiStrength sentiment analysis library and used it accordingly. First, the SentiStrength sentiment lexicon is translated into Turkish to be used in sentiment analysis. Then, this newly formed Turkish sentiment lexicon was used to classify the polarities of Turkish movie reviews. The system was evaluated depending on the accuracy scores. The accuracy was calculated by the ratio of the number of reviews

that have the correctly predicted polarity scores and the number of reviews that have incorrectly predicted scores.

In [12], Çoban et al. proposed a sentiment analysis system to classify the polarity scores of tweets. First, the data was gathered and prepared. For the intended purpose, the words in the dataset were stemmed and the stop words removed. Lastly, term frequency methods were used to determine which words in the dataset are more dominant. In this study, three different term frequency methods, TF (Term Frequency), Boolean, and TF-IDF (Term Frequency-Inverse Document Frequency). The authors used both machine learning-based and lexicon-based approaches were separately used to perform sentiment analysis. In the lexicon-based approach, the sentiment analysis was performed on the sentence level using the polarity information of each word in the sentence. The proposed system using the lexicon-based approach achieved a 77.1% accuracy rate while the proposed system that uses a machine learning-based approach to perform sentiment analysis achieved 89% accuracy. A detailed review of this study can be found under section 3.2.

The primary purpose of the study conducted by Albayrak et al. [13] was to analyze and interpret ideas on Twitter, where people share their opinions on certain issues. The tweets that were posted with the hashtag "bedelliaskerlikgeliyor" were collected from Twitter using the "TwitterAPI" with the "Tweepy" library. From this process, a dataset containing 12739 pieces of data was created from the tweets posted on this subject between October 10-12, 2017, and the data were preprocessed using the" NLTK" library. For preprocessing, the punctuation marks, etc. were removed from the dataset. The remaining words in the dataset were analyzed and interpreted by using the" SentiTurkNet" sentiment analysis dictionary, and how people felt about the hashtag topic were analyzed. According to the results, the benefits obtained by combining data analysis and sentiment analysis were determined. One of the advantages of the study is that people's thoughts can be identified about a certain subject with a solution model that is more suitable to the sociological sensitivity of the public.

Karaöz and Gürsoy proposed a new approach in [14] with the aim of indicating that people who work in the social science field with no sufficient software knowledge can also perform sentiment analysis on relatively large data. For this purpose, sentiment analysis was performed by using two dictionaries. Dictionaries were used to determine the polarity values of the words in the dataset. The data used in the study consists of different tweets which are collected over a period longer than eight-months about a TV channel. During this period, a total of 1,200,000 tweets were collected. In this study, only R language and Excel-vba were used. First, the dataset was divided to test and train the system. The division of the used dataset was as follows: 80% for testing of the system and 20% for the experiment. According to the results of this study, the average accuracy rate is calculated as 68.12%. From these results, we can claim that this study underperformed in terms of accuracy. Thus, we can interpret that the studies that include other methodologies such as Machine Learning, etc. to perform sentiment analysis generally possess higher accuracy rates. As a future process, it is aimed to perform better text preprocessing and better usage of sentiment dictionaries to increase the accuracy rate of the system. The advantage of the study is to show that anyone can perform sentiment analysis without having advanced programming knowledge.

Another SA methodology is implemented by Yüksel and Tan in [15]. This methodology proposes to analyze and classify restaurant reviews as positive, negative or neutral. They used a self-collected dataset which was gathered from the Foursquare application and comprises 7086 Turkish reviews from 128 different restaurants. They used the ITU Turkish NLP Web Service, Zemberek and Google Translate API. Their presented approach, Social Information Discovery Algorithm (SIDA), makes decisions in classification based on the presence of some special words which alerts the polarity of the sentiment of a review. According to the experimental results they reported, their algorithm achieved an 81,97% accuracy rate while NB algorithm achieved a 73% accuracy rate in Turkish reviews.

Authors in [16] conducted a sentiment analysis on Twitter data. In this study, the sentiment values of tweets were determined and labeled as positive, negative, or neutral in addition to extracting keywords from tweets and thus creating a sentiment dictionary. In this study a dictionary-based approach with an n-gram model is used for classification. As a dataset, it was created by collecting approximately 7k Turkish tweets from Twitter over a 4-month period. First, all the tweets in the dataset were preprocessed by removing redundant characters and processing special Turkish characters to prepare the data for the classification process. The number of repetitions of words along with word frequencies was found. For the N-gram model, 2,3,4 grams were used. In the dictionary-based approach, the words were classified and grouped as positive, negative, or neutral according to the information in the dictionary. The system was evaluated. As a result, the proposed system achieved the highest accuracy rate of 72%.

Word Level Sentiment Analysis

In this section, Turkish SA studies which employ a word-level SA are reviewed. These studies are reviewed based on methodology, dataset, and performance. The reviewed studies use dictionary-based approaches to perform word-level SA. In word-level SA, the polarities of single words are detected. In the study conducted by Dehkharghani et al. [3], sentiment analysis was performed on different granularity levels. The levels that the sentiment analysis performed were aspect level, word level, sentence level and document level. Word level polarity classification was performed. The word n-grams were assigned polarity scores using lexicon

based approaches. All of the details about this word polarity assignment can be found in section 2.1.

The study by Aydın et al. [17] proposed an approach to generate word and document embeddings for sentiment analysis. The sentiments of words were unstable. Sentiments can differ from one corpus to another. The reason for this instability is the usage of various methods and approaches in SA. This instability causes unbalance in system evaluations. In this study, the authors combined contextual and supervised features with the general semantic representation of words that occur in the dictionary. This research tried to create word vectors while using semantic and sentimental features of words in the vector generation process. The proposed model has many different components in its methodology. In the dictionary approach, they used the TDK lexicon which contains 616.767 different words. From this lexicon, the polarities of words have been obtained. But since the TDK lexicon was not a sentiment lexicon, TDK was combined with the domain-specific scores gathered from the corpus to generate word vectors. In the final component, supervised contextual 4 scores, four supervised scores are assigned to each word obtained from the corpus. Considering this component, a more precise polarity score can be achieved since four scores are analyzed rather than only one score which is the self-score. After different components are created distinctly, they combined the generated output together to receive the best possible result. Lastly, the document vectors were generated accordingly. The flowchart of the proposed system can be seen in Figure 4.

Erşahin et al. proposed a hybrid approach in [18] to apply in Turkish sentiment analysis. A detailed review of this study can be found in section 4.2. The detailed review study by Türkmenoğlu and Tantuğ [19] can be found under 3.2. The authors in [19] used lexicon- and machine-based learning approaches separately.

Performance Comparisons of Dictionary Based Approaches

Table 1 presents Dictionary-Based Turkish Sentiment Analysis studies, categorizing them by year of publication, the utilized approach, analytical level, dataset size, and highest accuracy reported. As presented in Table 1, [20] achieves 91% accuracy, leveraging a substantial dataset of 43,000 samples. Notably, studies [11] and [19] both employ the SentiStrength dictionary; [11] achieves a 73.7% accuracy rate with a dataset comprising 60,000 samples, while [19] demonstrates a slightly higher accuracy of 77.1% using a dataset of 20,000 samples. However, it's worth noting that for a more comprehensive performance comparison between these studies, additional details concerning their respective datasets and approaches are required.

Various approaches can be used to find the sentiment of a given text. These are dictionary-based, ML-based and hybrid based. In this chapter, we analyzed sentiment studies using dictionary-based approaches. While dictionary-based sentiment analysis has its benefits such as offering simplicity and transparency, it is considered outdated when compared to hybrid ML approaches. Dictionary-based methods have limitations as they don't have the ability to recognize context well, making them less effective in capturing subtleties in language [8] which leads to less effective results in performance. They struggle to perform

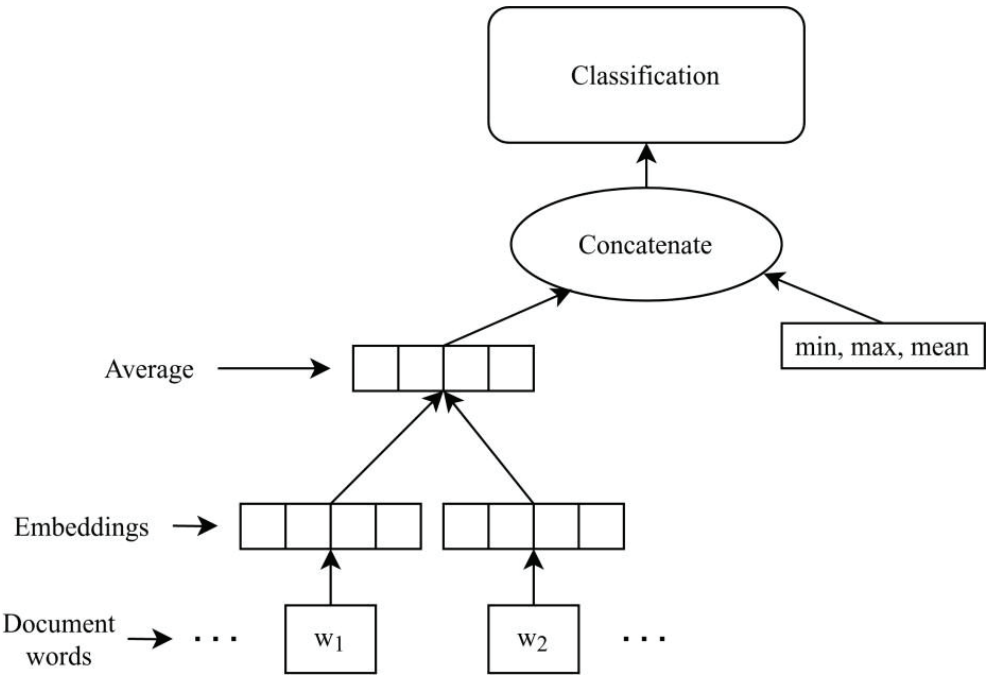


Figure 4. The flowchart of the system.

Table 1. Comparisons of Dictionary-Based Turkish Sentiment Analysis Studies

Study	Year	Level	Data size	Dictionary	Performance metric	Result
Vural et al. [11]	2013	Sentence	60k	SentiStrength	Accuracy	73.7%
Türkmenoglu and Tantug [19]	2014	Sentence	20k	Sentistrength	Accuracy	77.1%
Akgül et al. [16]	2016	Sentence	7k	TDK	Accuracy	72%
Karaöz and Gürsoy [14]	2018	Sentence	1.2M	Custom	Accuracy	68%
Yüksel and Tan [15]	2018	Sentence	7k	Custom	Accuracy	81%
Erşahin et al. [18]	2019	Sentence	220k	SentiTurkNet	Accuracy	74.90%
Toçoğlu and Alpkocak [20]	2019	Sentence	43k	Custom	Accuracy	91%
Aydın et al. [17]	2020	Word	10k	TDK	Accuracy	78.3%

sufficiently with complex languages like Turkish. Besides, hybrid approaches combine the simplicity of dictionaries with the contextual understanding of ML, offering a balanced solution. The evolution of ML-based techniques has significantly improved accuracy and adaptability, making dictionary-based methods less popular in today's dynamic world of sentiment analysis. In particular, the usage of deep-learning techniques has evolved over the years and can provide outstanding results. In chapters 3 and 4, ML and hybrid-based approaches are further explained.

TURKISH SENTIMENT ANALYSIS WITH ML BASED APPROACHES

In this section, Turkish sentiment analysis studies that use ML-based approaches are presented. ML is considered an important concept and a versatile tool that has gained popularity in recent years and is now applied in many different fields. Due to its versatility, ML can be classified as the most popular approach when compared with other methodologies [8]. There are various ML techniques that can be used in Turkish sentiment analysis. ML techniques can be classified as supervised, unsupervised learning and deep learning techniques. All categories of ML techniques are useful in solving a variety of NLP tasks. In sentiment analysis (SA), using ML approaches has many advantages and provides successful results when applied to a well-structured large dataset. Machine learning and deep learning techniques involve training models on large datasets to identify patterns and context in Turkish text. These models then assign sentiment labels to the text based on their learned knowledge. In ML-based Turkish SA, several machine techniques are generally utilized in a single study. Different levels of ML-based Turkish SA can be performed. These levels are aspect level, document level, sentence level, and word level.

Aspect Level Sentiment Analysis

In this section, studies that perform aspect level SA are reviewed based on methodology, dataset, and performance. The studies reviewed below use machine learning-based approaches to perform aspect-level SA.

Another recent study has been conducted by Bayraktar et al. in [21] in which a holistic method has been researched for Turkish aspect-based SA. As a dataset, many restaurant reviews were collected from various resources. Prior to the SA process, preprocessing steps were applied to the dataset. For preprocessing, spelling errors were removed and corrected using Zemberek and Yandex. Then, all the words in the dataset were converted to lowercase and inflectional suffixes were also removed from the words in the dataset using XML tools. This way the preprocessing is done and the root words in the dataset are gathered without losing their meaning. In aspect-based SA, the Latent Dirichlet Allocation (LDA), Pointwise Mutual Information (PMI), C-value, and WSBFE (Web Search Based Feature Extraction) are used during the aspect extraction. In [21], PMI was used to measure the association between two terms P (word1, word2) and indicated the probability of word1 and word2 coexisting, and the formula P(word1) P(word2) represents the probability that the two terms coexist when they are statistically independent. Overall, this research received 56,28% accuracy in the aspect extraction process, and received 52,05% accuracy in sentiment classification, which is relatively low. The main reason that the accuracy is low is the unpredicted aspect-sentiment pairs which are misclassified by the system. As future work, the authors in [21] are considering applying a double propagation method to increase the success of the aspect sentiment matching.

$$PMI(word_1, word_2) = \frac{P(word_1, word_2)}{P(word_1)P(word_2)} \quad (1)$$

$$P(word) \equiv hits(word) \quad (2)$$

$$P(word1, word2) \equiv hits(near(word1, word2)) \quad (3)$$

Ekinci and Omurca proposed a new approach in [22] in which it was aimed to classify the comments written about a product via a subject modeling method according to product features. In this study, GDA was used as the

topic modeling algorithm. The dataset used consisted of 1,000 user comments regarding hotels. These reviews were collected from the website "www.otelpuan.com". The data consist of a total of 5364 sentences. First, the dataset was preprocessed using the Zemberek library. Then, the classification was performed and the system was evaluated. According to the results, the success rate of the system was exactly 99%. The advantage of this study is that it achieves high accuracy by using the LDA algorithm. The disadvantage of this study is that it used a very small dataset to perform classification. So, the rating of the system's accuracy, which is a 99% accuracy rate, may not be reliable.

Mutlu and Özgür proposed a recent study in [23] that focuses on performing a targeted SA on Turkish text. The difference of targeted SA over normal SA is that targeted SA tries to predict the sentiment of a text according to a specific text rather than identifying the overall sentiment. To perform targeted SA, Bert-based models were implemented using different architectures. Bert is a neural network-based model which is widely used in Natural Language processing. The dataset was collected from Twitter. 19% of the dataset includes sentimentally positive tweets, 58% includes negative tweets and 23% includes neutral tweets. The model is trained and tested in the collected dataset. Lastly, the results of the proposed Bert model and different baseline Bert models were evaluated. According to the results, the proposed model outperforms other baseline models by achieving a 67% F1-score when tested on the same dataset.

Sentence and Document Level Sentiment Analysis

In this section, studies performing sentence-level SA are reviewed based on methodology, dataset, and performance. The studies reviewed below use machine learning-based approaches to perform sentence level SA.

Kilimci proposed a new SA system in [24] in which the direction of the Borsa Istanbul index is predicted using SA. For datasets, two distinct datasets were created by utilizing Turkish and English tweets on the tags BIST100 and XU100 in Twitter and used in this research. These datasets were enriched by using word embedding methods such as Word2vec, Glove etc. Although more than one method was emphasized in the study, ensemble learning was unanimously endorsed in the end. Since it is necessary to use different classifiers in ensemble learning, the heterogeneous ensemble system is emphasized. In the study, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM) algorithms were used as basic learners to ensure community diversity. The MV and "STCK" methods were used to combine community decisions. Firstly, ant colony optimization and selection of the features in the dataset were performed for the dataset. Then, these obtained features were embedded using the word2vec and glove methods. Following this step, document representations were created using Term Frequency Inverse Document Frequency (TF-IDF) with Avg(Word2vec), Avg(Glove), Avg(Word2vec) +Avg(GloVe), TF-IDF+Avg(Word2vec), TF-IDF+Avg(GloVe). Ensemble classification is obtained by applying CNN, RNN, and LSTM algorithms to the obtained document representations. The results obtained in the community classification have been achieved utilizing the majority voting and heap community strategy methods. As a result, the study achieved a 78.07% classification performance rating for the Turkish dataset [24].

The author in [25] proposed an approach to perform SA on Turkish tweets from Twitter based on the Latent Dirichlet allocation algorithm. In this study, the data from Twitter were tried to be classified as positive, negative, or neutral. Two approaches can be used to perform SA, machine learning-based approaches and dictionary-based approaches. The dataset used in this study was collected from Twitter. A total of 10600 tweets were retrieved from Twitter. 5300 out of 10600 tweets were positive tweets and the remaining 5300 tweets were negative. First, the data were preprocessed, then machine learning methods were utilized in tweet classification. The Naive Bayes algorithm was chosen as the main classification algorithm. The proposed model was evaluated using the "wekada" cross-validation method. According to the system evaluation, the proposed system achieved a 78.34% accuracy rate in tweet classification.

Kaynar et al. proposed a model in [26] for Turkish SA, performed using machine learning and deep learning methods. In this study, Turkish data was collected from social platforms as a dataset. These data were then classified as positive or negative. For the dataset, 2000 movie reviews were collected from an online resource of which 1000 these reviews were negative, and the remaining were positive. For SA, artificial neural networks(ANN), support vector machines, Naive Bayes, and center-based classifier methods were used. This study mostly focuses on ANN. The ANN structure used is given in Figure 5.

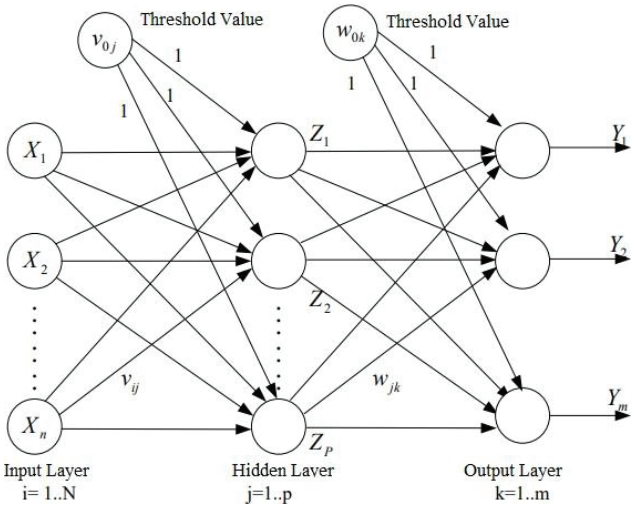


Figure 5. The structure of artificial neural networks.

Aytekin Keskin proposed a SA method in [27] on Turkish texts regarding interest-free finance systems. Specifically, perceptions of potential customers towards interest-free financial systems were aimed to be detected. The dataset used consisted of Turkish reviews of customers about the interest-free finance system in January 2019. The content of the dataset was collected from various internet resources and social media. In this study, a sentence-level SA was realized using a Machine learning approach on Turkish texts. The SA was performed using an online program from the "www.mediatoolkit.com" website. In the classification process, the most common concepts related to interest-free finance systems were identified and the dataset was examined using keywords to classify information about interest-free finance systems. The SA of the texts in the dataset that mentioned concepts related to interest-free finance systems resulted in them being classified as positive, negative, or neutral. These results were analyzed and recommendations were put forward based on the findings. This study aimed to understand customer perceptions of interest-free financial systems. The results showed that the mention of "participation banks" alongside the concept of "interest" in the press creates a negative bias towards these institutions. Additionally, the association of commercial institutions that provide interest-free financing systems with various concepts, namely banks, leads to varied feelings about these institutions within society.

The authors in [12], proposed a SA system to classify the polarity scores of tweets. In this study, the dataset was collected from Twitter. Twitter API was used in the data collection process to retrieve tweets efficiently. This process resulted in the creation of a dataset that consists of 20k tweets. To classify tweets, machine learning approaches were used. Specifically, SVM, Multinomial NB, and K-nearest Neighbors (KNN) were used in the classification process. In this system, it is stated that since each tweets in the dataset can be classified into different categories such as positive, negative, etc., and the basic document classification algorithms were applied. Before the classification process began, preprocessing methods were applied to the content of the tweets. Then, the attribute selection was performed. The system was evaluated for every machine learning approach used in the classification. According to the results, it is seen that different machine learning approaches have different accuracy yields in this system. The accuracies of machine learning approaches are NB:62.04%, MNB:66.06%, SVM: 62.25%, KNN:65.79%. Thus, based on the accuracies achieved in the system, the Multinomial NB outperforms other approaches. In conclusion, this study used widely known classification methods to perform SA on Turkish tweets. For future improvements to the system, an external lexicon could be used besides Machine learning methods. The authors stated that adding an additional lexicon would increase the system's accuracy. The major decrease in the accuracy of the system is the result of using random sentiment score based methods to determine the sentiment value of tweets.

In the study by Türkmenoglu and Tantug [19], opinion extraction was performed and the polarity of these opinions were analyzed. To achieve these, two different approaches were used. These approaches were Machine Learning-based and Lexicon-based SA methods. In this study, two distinct datasets were used. One of these datasets consisted of tweets collected from Twitter. The other dataset consisted of movie reviews which were collected from the website "beyazperde.com". In the machine learning approach, SVM and NB methods were used in the classification of opinions. To use these machine learning methods, a feature set which contains word root words and n-grams was represented with a bundle of words. For the lexicon-based approach, it is attempted to predict the sentimental orientation of an input text using sentiment scores of words and phrases in text using the information in the sentiment lexicon. As a result, the lexicon approach achieved an accuracy of 77.1% and the machine learning approach achieved an accuracy of 89%. From these results, it is detected that the machine learning approach outperforms the lexicon approach.

In the study conducted by Kaya et al. [28], SA in the Turkish language was investigated using different sentiment classification techniques. These techniques were different supervised machine approaches. A total of four different supervised machine learning algorithms were used. These are Naive Bayes, Maximum Entropy, SVM and the character-based N-Gram Language Model. The N-gram language models are used to create the N-gram based character language model. Instead of words, this model uses characters as the basic unit in the algorithm. For strings "s," the model provides p(s). The chain rule is given for a character (c) and a string (s). The context is limited to the previous (n-1) characters due to the N-gram Markovian assumption. As a result, the maximum likelihood estimator for N-grams is given, where C(sc) is the number of occurrences of the sequence in the training data and the denominator is the number of single-character extensions of sc.

$$p(sc) = p(s) \times p(c|s) \tag{4}$$

$$p\left(c_n|s_{c1\cdots c_n-1}\right) = p(c_n|c_1 \cdots c_{n-1}) \tag{5}$$

$$p(cs) = \frac{C(sc)}{\sum_{c} C(sc)} \tag{6}$$

For this study, a dataset was created containing Turkish political news articles collected from different news sites' political sections. Before the sentiment classification was performed, some preprocessing steps were applied to the dataset. These preprocessing steps were removing the HTML tags and stemming the words in the dataset using the Zemberek tool. Then possible roots and suffixes are

found automatically. Additionally, a list of words which people tend to use to express strong sentiments were created to be used in this study. Then, the classification process began. To evaluate the system, the K-fold-cross-validation was used. The experiments of this study were conducted with Bigram and Unigram features. However, results show that bigram information is not as useful as information obtained through unigram for the Sentiment Classification of news. Overall the system received a 76% accuracy rate. For future work, the authors thought of adding Named Entity Recognition to the system which might be used to identify which columnists write about which political party or politician etc.

In this research conducted by Ciftci and Apaydın [29], a modern deep learning method called RNN was developed using LSTM units on the dataset instead of machine learning methods that use the word bundle model such as Logistic Regression and Naive Bayes for SA and comparing the results of RNN based methods with the results obtained from traditional machine learning methods. It was revealed how much improvement modern natural language processing and deep learning methods can bring about in emotion analysis results. An external library was not used for this research. First of all, a 355-thousand-word dataset consisting of 283 thousand positive and 72 thousand negative words was collected from Turkish shopping and movie reviews. After the tags were removed and normalization was applied, punctuation marks and stop words were removed. Then, the datasets were separated as 80% training and 20% testing. After Naive Bayes and Logistic Regression models sentences were vectorized using TF-IDF, a hyper parameter search was conducted for training. TF-IDF was not used for the RNN based algorithm, word vectors directly fed the RNN architecture. The results showed that the RNN-based deep learning method improved the classification accuracy. As a result, the biggest advantage of this research is that despite the unbalanced data problems of modern deep learning methods and the dominance of positive comments, RNN-based algorithms have been shown to improve accuracy in emotion analysis, but the long training times of deep learning methods can be viewed as a disadvantage.

A hybrid machine learning approach is proposed by Shehu and Tokat in [30] using "Random Forest (RF)" and "Support Vector Machine (SVM)" algorithms used for Turkish SA. Turkish letters of 3,000 and 10,500 words were used as a dataset. A sentiment dictionary translated from 27,000 English words was used as the SA dictionary. According to this proposed hybrid model, "Zemberek" was used as a resource to find the roots of the words as a first step. According to the authors, it has been observed that the RF algorithm gives better results than the SVM algorithm in the classification of positive words, while the SVM algorithm gives better results than the RF algorithm in the detection of negative and neutral words. Based on these two results, it was planned to classify the data according to two classes, 'Positive' and 'Others', using the RF algorithm. The data in the 'Others' class is intended to classify the data according to three classes, positive, negative, and neutral, using the SVM algorithm. While SVM achieved 76.4% and RF achieved 75.9% accuracy in the dataset with 3,000 items of content, the mixed method achieved an accuracy rate of 86.4%. In the large data set consisting of 10,500 items of content, SVM achieved an accuracy of 67.6% and RF 71.2%, while the mixed method achieved 82.8% accuracy. The accuracy of this approach can be improved and tested with the use of other data sets. The biggest advantages of the study are that the accuracy rates in classification can be increased by using larger emotion dictionaries, it is open to development due to the improvement of classification algorithms, and it is clear that the mixed model has achieved higher accuracy compared to other methods.

A new SA study was proposed by Acikalin et al. in [31], using various data written in Turkish. In this study, two models based on the BERT model were proposed. These were developed by Google using the transformer architecture. In the first model, a multilingual BERT was adapted to the proposed system. In the second model, Turkish texts were translated to English after which the main model of BERT, which was developed for English, was adapted to this Turkish dataset. The mTranslate library was used in the translation process of the system. Two distinct datasets were used in this proposed system, these datasets included movie and hotel reviews collected from online resources. The size of the movie reviews dataset includes 53,400 words, of which 26700 are positive and 26700 are negative. The size of the hotel review dataset includes 11,600 words, 5800 of which are positive and 5800 are negative. The models that were developed in this study were compared with 4 different models which are fastText-Tr, and fastText-En. Experiments show that when enough data is obtained, the BERT model can learn enough features to successfully classify the given data. In the future, adapting other general models for the purpose of text representation on different datasets and combining the results with BERT are considered. Additionally, developing a Turkish version of the BERT model is being given considerable thought.

The study by Balli et al. [33] aimed to find the sentimental analysis of Turkish tweets. The main purpose of this study was to detect the emotional states of Twitter users by classifying the tweets they wrote as positive, negative, or neutral. For this purpose, 2 different Twitter datasets were used. The first dataset was a public dataset and the second one was a custom-made dataset named SentimentSet which included 11k tweets and is publicly available. After the dataset preparations, preprocessing was applied to the datasets to clean and prepare the data for classification. For classification machine learning models were used. Before the classification, the data was vectorized using the TF-IDF vectorizer to convert the text into numerical form to be used by the machine learning models. Logistic Regression, Random Forest, Naive Bayes, SVM were implemented for machine learning models, and for deep learning, LSTM

models were implemented. Then, the models were trained and tested with the prepared data. According to the results, random forest and LSTM models performed the best by achieving an 85% accuracy rate.

The authors in [34] tried to classify tweets about music, art, etc. The analysis was conducted within categories. In the study, the tweets were taken from Twitter using the Twitter API. In this study, the libchart library was used to graphically display the output of SA and categorical classification. The Bayesian algorithm was used for the classification of tweets. In this study, the tweets of a Twitter user are categorically classified and SA is performed on those tweets. In future studies, it is aimed to increase the number of tweets used in the study and increase the accuracy of the system accordingly. The advantage of the study is that SA and classification of any given Twitter user can be performed through this created interface. The main disadvantage of the study is that only 100 tweets from a selected user can be processed at a time.

In this study by Tuncer and Çetintaş [35], the authors tried to sentimentally classify the tweets on Twitter as positive, neutral and negative and also to detect malicious tweets. For classification, the Decision Tree algorithm and Naive Bayes algorithm were used. For the dataset, 20,000 tweets were collected and used in the system. Knime program and excel macro were used as utility programs. According to the evaluation of the system, the system achieved an average of 75.2% accuracy rate using the Decision Tree algorithm and 56% with using the Naive Bayes algorithm. The advantage of the study was that the classification stages were performed in pairs, which resulted in a more observable result. The disadvantage of the study is that some Turkish translation processes were incomplete.

In [36], Güran et al. studied the sentiment polarity detection problem with social media data. They applied a grid search method in order to discover the most suitable kernel function in SVM. They used 3 datasets: 1) VS1(3 classes, 3000 data), 2) VS2(4 classes, 157 data), 3) VS3(3 classes, 105 data). According to the experimental results they reported, their proposed model received an average of 75.2% accuracy.

A SA model was proposed in [37] for Turkish text. For this purpose, three different types of data were used, which were reviews of products, movies, and books. A machine learning model was used to process and perform SA on all of these three types of data. The data was attempted to be classified as positive, negative, or neutral. First, the most frequently used 20,000 words in the dataset were determined. Scores were assigned between 0 and 1 by normalization. By removing the missing and neutral data, 105,220 data were obtained from each dataset. A mixed set of data was created by taking equal amounts of data from all three datasets. TensorFlow was used for the models. With the Sklearn library, the data was split as 90% training and 10% testing. While each model provided a very high (85-95%) accuracy in its own channel, the success rate was greatly reduced in other channels (50%). While the mixed model had success rates of 79.8% and 85.8% in mixed data, it also achieved a high success rate of 77.8% and 85.8% in other channels.

In the study by Aytuğ [38], SA was performed by analyzing tweets collected from Twitter. Three different machine learning models were used in the classification process of tweets. A special dataset was created for this research by collecting tweets from Twitter using the Twitter API with the Python programming language. This created dataset contained a total of 10600 Tweets, 5300 of these tweets were positive and 5300 are negative. Before the data classification process, preprocessing was applied to the dataset. For preprocessing, the stop words and characters such as "@", "#" were cleaned and all the words in the dataset were rooted using the Zemberek Library and duplicate words were removed from the dataset as well. The N-gram model was used to determine the appropriate features of the data. For classification, NB, SVM, and (Logistic Regression) LR models were used. The Weka software was used to perform a classification procedure on these three classification models. The proposed SA system was evaluated using the 10-fold cross validation method. According to the results, different classification models received different rates of accuracy. The LR model received 77.23% accuracy, the SVM model received 73.68% accuracy and the model NB received 77.78% accuracy. Even though the accuracy rates of the three models are close to each other, it is seen that the NB model achieved the best accuracy rate compared to other models.

In [39], Sarıman and Mutaf have attempted to analyze people's feelings about Coronavirus through social media from the date of the spread of the virus until the present day. People's feelings have been attempted to be analyzed with the help of textual material shared on social media, specifically, tweets shared on Twitter. As of March 11, 2020, 2 million tweets were collected using the Twitter API to form a dataset. To process the data, Python's "pandas", "numpy" and "sklearn" libraries have been used. Preprocessing was applied to the created dataset. Preprocessing is applied by removing unnecessary Turkish characters in the dataset and also by deleting the recurring tweets in the dataset. The tweets underwent SA using machine learning methods. The meaning of the tweets in the dataset was obtained through word sequences, sentence analysis, and emotion analysis. Logistic regression was used in the classification of tweets. The system was evaluated and the AUC metric was used as the basis metric for success. As a result of the classification process, five main classes were created as follows: Eba, Mask, State Support, Curfew, and Short Working Allowance. After predicting the sentiment of the tweets, the AUC value was obtained. The AUC classification results are as follows: Mask (0.97), Eba (0.94), Curfew (0.98), State Support (0.86) and Short Working Allowance (0.91) results were obtained.

The authors in [40] developed an application to analyze tweets using the Tweepy, Odoo, and NLTK modules in the Python programming language. Analyses were made by accessing the tweets posted over the hashtags on the Twitter platform. The Odoo module was used to display data and results in an organized structure. The NLTK module was used to apply natural language processing techniques to the data. The dataset used was created by collecting tweets from Twitter. As a result of the system, it was observed that Twitter tags can be analyzed in a versatile way within a single interface.

In the study by Kaynar et al. [41], a comparison was made regarding the Feature Reduction Methods with Deep Autoencoder Machine Learning in SA. The autoencoder architecture is given in Figure 6. In the autoencoder architecture, the first layer is the input layer, the second is the hidden layer and the final one is the output layer. The space between the input layer and the hidden layer is called the encoder. The space between the hidden layer and the output layer is called the decoder. The results of using linear and nonlinear dimension reduction techniques in combination with machine learning methods were compared to the results of using machine learning methods by themselves. It has been observed that the model that uses linear and nonlinear methods with machine learning methods performs better than the model that only uses machine learning models.

In another recent study proposed by Tuzcu in [43], Turkish texts were classified through SA. In this study it was aimed to receive high accuracy for Turkish SA. A Sentence and document-level SA was performed using machine learning-based approaches. For the dataset, a total of 91309 book reviews were retrieved from an online website. Prior to the sentiment classification, all of the data in the dataset was preprocessed using the methods of the NLTK library. Various machine learning algorithms were used in the classification process of SA. These classifiers are MLP, NB, SVM and LR. The sentence- and document-level SA was performed on the same dataset using all these classifiers and were compared according to the classification success. The system was evaluated using accuracy as a metric. According to the evaluation, it can be observed that different machine learning algorithms received different accuracy rates when tested on the same dataset. The MLP algorithm received an 89% accuracy rate, the LR algorithm received 84% accuracy rate, the SVM algorithm received 80% accuracy rate and the NB algorithm received a 77% accuracy rate. According to these results, the MLP algorithm performs the best, achieving an 89% accuracy rate, when compared with other classification algorithms.

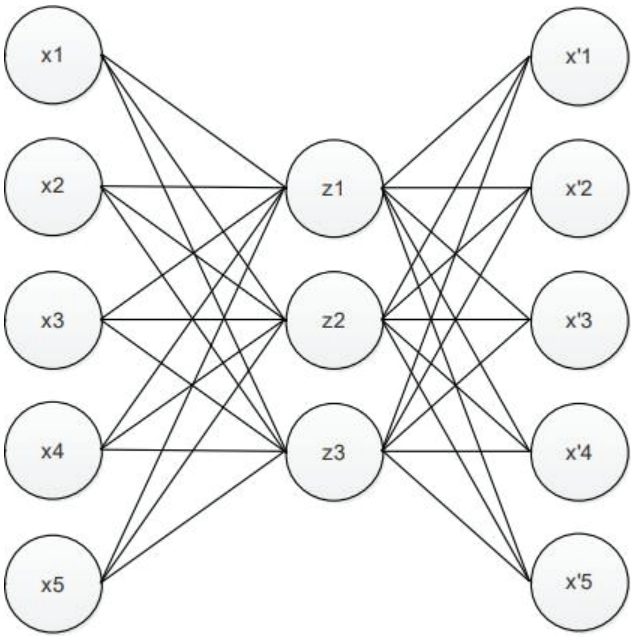


Figure 6. The Architecture of AutoEncoder.

In [44], the authors conducted a SA of Turkish language text using a dataset collected from various social media platforms. The study employed three algorithms: Random Forest, Logistic Regression, and LSTM. The dataset consisted of 28,189 data points collected from five social media platforms and manually labeled as positive, negative, or neutral. Of these, 5,712 were labeled as positive, 11,567 as negative, and 11,247 as neutral. The experimental results showed that deep learning models outperformed machine learning approaches in terms of performance. In the study, the LSTM model achieved the highest accuracy with a rate of 84.46%.

Performance Comparisons of ML Based Approaches

Table 2 presents the Machine Learning-Based SA studies including their approaches and performance results. In this table, ML-based Turkish SA studies are listed based on their year of publication, approach, sentiment level, the size of the data used, and highest results achieved in the study.

According to the results, several parameters are required to draw a meaningful comparison; approach, data, data size, performance metric etc. In this case, to make this comparison between the given studies, we should choose the studies that used the same performance metric since we don't have the complete information about their dataset. According to the given information, the results on Table 2 shows a wide range of performance in ML-based Turkish SA. Studies like Acikalin et al. [31] and Kemaloğlu et al. [44] achieve notably high accuracy rates, while others, such as Çoban et al. [12] and Tuncer et al. [35], report comparatively lower results. The choice of the machine learning algorithm, dataset size, and specific problem domain can significantly affect the outcomes. ML techniques, particularly deep learning approaches, have gained popularity in recent years to provide solutions for many specific NLP tasks, especially in SA [45]. Additionally, the recent utilization of deep learning techniques, as seen in studies like Acikalin et al. [31], has demonstrated impressive accuracy, underlining

Table 2. Comparisons of ML-Based Turkish sentiment analysis studies

Study	Year	Level	Data Size	Performance Metric	Result
Kaya et al. [28]	2012	Sentence Level	400	Accuracy	NB:72.05%, ME:76.78%, N gram:76.78%, SVM:76.31%
Güran et al. [36]	2014	Sentence Level	105	Accuracy	SVM: 75.2%
Türkmenoglu and Tantuğ [19]	2014	Sentence Level	20k	Accuracy	SVM:89% , NB:89%
Çoban et al. [12]	2015	Sentence Level	20k	Accuracy	NB:62.04%, MNB:66.06%, SVM: 62.25%, KNN:65.79%
Kaynar et al. [26]	2016	Sentence Level	2k	F-measure Sensitivity	83%
Onan [25]	2017	Sentence Level	10k	Accuracy	NB:78%, SVM:77%, KNN:72%, RF:72%
Alpkoçak et al. [42]	2019	Sentence Level	27k	Accuracy	SVM:86.2%, ANN:86.6%, RF:82.5%, KNN:70,9%
Erşahin et al. [18]	2019	Sentence Level	220k	Accuracy	NB:82.07%, SVM:85.4%, J48:76.7%
Tuncer et al. [35]	2019	Sentence Level	20k	Accuracy	DT:75.2% , NB:56%
Acikalin et al. [31]	2020	Sentence Level	11k	Accuracy	Bert: 93.32%
Aksu et al. [32]	2020	Sentence Level	49k	F-Score	SVM:91%, NB:86%
Sarıman and Mutaf [39]	2020	Sentence Level	2M	AUC	LR:93%
Tuzcu [43]	2020	Sentence Level	91k	Accuracy	MLP:89%, LR:84%,SVM:80%, NB:77%
Kilimci [24]	2020	Sentence Level	30k	Accuracy	CNN:74%, RNN:72%, LTSM:76%,MV:77%, STCK:78%
Kemaloğlu et al. [44]	2021	Sentence	25k	Accuracy	84.46%
Aktaş et al. [48]	2022	Sentence Level	1M	Accuracy	NN:86.37%, NB:83.55%, KNN:81.88%

the evolution of these approaches in the field of Turkish SA. We can conclude that deep learning tools such as transformers can provide immense results. Bert (Bidirectional Encoder Representations from Transformers) is a type of transformer model which is used extremely frequently for NLP tasks [45], including in language understanding, translation, and SA due to its ability for understanding the context and relationships between words in a sentence, which is essential for many NLP tasks [45].

TURKISH SENTIMENT ANALYSIS WITH HYBRID APPROACHES

In this category, Turkish SA studies that use hybrid-based approaches are presented. Hybrid-based approaches are highly preferred for Turkish SA over single approaches like dictionary-based approaches. The use of a hybrid approach in SA combines elements of both dictionary-based and machine learning-based methods to improve the performance and effectiveness of sentiment classification. This approach is used to improve the strength and efficiency of each method. As a result, in hybrid approaches, machine learning techniques are used as a baseline approach. In addition to the ML- based approach, dictionary-based approaches can be used. In the hybrid approach that uses dictionary-based methods, an external lexicon is used with ML techniques. Hybrid-based Turkish SA can be executed on different levels. These levels are aspect, document, word and sentence.

Aspect Level Sentiment Analysis

In this section, a study performed on an aspect level SA is reviewed based on methodology, dataset, and performance. The study reviewed below uses hybrid-based approaches to perform aspect-level SA.

In the study by Çetin and Eryiğit [46], SA was performed to analyze people's feelings to classify opinions as positive, negative, or neutral. As a data set, a set of Turkish restaurant reviews was used. The dataset consists of a total of 1415 sentences of restaurant reviews. Prior to the classification process, the data is prepared by using preprocessing techniques. A Word segmenter, Turkish character corrector, morphological analyzer, morphological disambiguation and dependency parser provided by ITU NLP tool were used to prepare the data. In this study, the CRF algorithm is used as the primary method in tagging the data. Logistic regression was used to separate the generated data. In order to determine the relationship between the words in a sentence, two different methods were used both separately and in combination. One of these is the relationship of neighboring words at a certain distance and the other is the detection of words associated with the loyalty tree. For the criteria aspect, the F1 criterion was used in the evaluation

of the system. According to results, it was observed that the system received a 76% F1-score. In the future, it is aimed to improve the methods that can overcome the difficulties created by the Turkish dataset and to use deep learning methods to improve the accuracy of the systems.

Sentence and Document Level Sentiment Analysis

In this section, studies performing sentence-level SA are reviewed based on methodology, dataset and performance. The studies reviewed below use hybrid based approaches to perform sentence-level and document-level SA. In hybrid approaches, different methodologies are combined and used to propose efficient models for the task of SA.

In this study by Onan [47], SA was performed on Turkish text documents to evaluate the predictive performance of 36 word embeddings based on representations obtained by three-word embedding methods (i.e., word-2vec, fastText and DOC2vec), two basic weighting functions (i.e., inverse document frequency and smooth inverse document frequency) and three vector pooling schemes (namely, weighted sum, center based approach and delta rule). For the dataset, a total of 21000 Twitter messages were used, of which 10500 were positive and 10500 were negative. Preprocessing steps such as string splitting, removing stop words, and finding roots were applied on the dataset. The area under the ROC curve (AUC) was used to evaluate the effectiveness of the methods. Experimental analysis shows that word2vec-based representation in conjunction with inverse document frequency-based weighting and center-based pooling yields promising results for SA in Turkish. (0.9131 AUC). The advantage of this study is that there is no study conducted on Turkish text documents that comprehensively analyzes the weighted word vectors, and this study constitutes a basis in this manner.

In the research by Gezici and Yankıoğlu [4], a model was built to estimate the sentiment value of movie reviews in Turkish. This model combines supervised learning and lexicon-based approaches. Firstly, it computes the average polarity of the words in the text and trains a classifier (Naive Bayes or SVM). Then, the effectiveness of more complex processing techniques and features are measured: handling negation; considering the effects of booster words; and using other features derived from the seed words. In this research, the basic approach obtained 67.49% accuracy with the Naive Bayes classifier and 67.61% with the SVM classifier; while the best results were obtained using all of the features, achieving 74.28% accuracy with the Naive Bayes and 75.52% with the SVM classifiers, respectively. The advantages of this work is that it considers the seed word occurrences and, in this way, classification accuracy significantly improves. Secondly, if a larger lexicon can be used, the system can achieve an improved classification performance.

In the study by Aktaş et al. [48], a SA was conducted on reviews of an online food delivery service using machine learning models in Turkish. The goal was to predict whether a given review was positive or negative. To create the dataset, 676,000 reviews were collected from "yemeksepeti.com", with 338,000 labeled as positive and 338,000 labeled as negative. The dataset was preprocessed via lemmatization and normalization. The study compared the performance of several machine learning algorithms and deep learning algorithms, including KNN, NB, and a neural network. The results showed that the neural network model had the highest accuracy rate, at 86.37%, followed by the NB model at 83.55% and the KNN model at 81.88%.

In the study conducted by Erşahin et al. [18], a SA approach is presented using a hybrid approach which consists of both dictionary-based and machine learning approaches. In the dictionary-based approach, a new lexicon was created by expanding the STN lexicon via the ASDICT model called eSTN. The STN lexicon was improved by adding synonyms of words to the dictionary using the "ASDICT" model. For the ML side, the classification problem is handled by using different ML models, which are, NB, SVM and "J48" classifiers. ASDICT, which was developed to identify synonyms and contains 70,000 words which were used to expand an external dictionary. In this study, different datasets containing hotel and movie reviews and tweets were utilized. The selected movie reviews, hotel reviews and tweets included a total of 1,345,726, 738,216 and 19,056 words respectively. The techniques that were used in experiments are: "NB", "NB+active learning", "Logistic regression+QER", "Lexicon" and "SVM". The results show that a hybrid approach to SA provides better results. The highest relative accuracy, at 83%, was achieved by the system. Future consideration is given to improving the overall performance of this system by using aspect-based SA and other subtasks of SA.

In this study conducted by Aydın et al. [51], it has been attempted to develop a hybrid SA method with a higher success rate than the dictionary-based and machine learning-based methods. In the study, tweets related to the Apple, Google and Microsoft companies were used as a dataset. These tweets were divided into 3 sections which were positive, negative and neutral. Specifically, three datasets were used. The first two datasets consisted of 479,988 tweets. These tweets were classified using the Pso and k-eyk algorithms. With the use of the multiple Pso classifier, it has been observed that the Pso population is oriented towards their own class from the 3 separated classes, and as a result, each particle can be classified according to its original class. The system was evaluated by comparing three methods on four parameters. These parameters are precision, sensitivity, f1-score and accuracy. The system achieved a 73% precision score.

In the study by Demir et al. [52], the authors tried to sentimentally analyze large amounts of Turkish text data in a short amount of time. In this study, three varying datasets were used. The first dataset contains about 5000 Turkish words, and the second and third dataset contains a total of 25000 words. In this study, mostly a dictionary-based

approach was used with machine-learning methods to conduct SA. In the dictionary-based approach, four different dictionaries were used. These dictionaries were by Afinn, Bing, NRC and SentiTurkNet. A word-level SA method was performed using a dictionary-based approach. Then, for classification of words in the dataset, the NB method was used. This hybrid system was evaluated. According to results, the system achieved a peak accuracy rate of 82%.

In the proposed paper [6], "The Turkish movie reviews" dataset (composed of 34990 positive and negative movie reviews in Turkish) had been used for the sentiment classification task. The model was fed with pre-processed and vectorized data. The pre-processing included the steps of tokenization, stopwords, special characters removal, fixing misspelled words, stemming, and detecting negation. Vectorization was performed with the VSM (Vector Space Model). This vectorization model uses features like TF, TF-IDF, and Word Embeddings (Word2Vec, GloVe). For the classification part, these extracted features were applied to the three different well-known ensemble algorithms which were AdaBoost Classifier (AdaBoost), Random Forest (RF), and GradientBoostingClassifier (GBC). The evaluation of the classifiers was made according to standard metrics such as precision, recall, and F1 score. The best result achieved among these models belongs to the RF classifier with 86% accuracy.

In [53], Saed Alqaraleh proposed a model based on Deep Convolutional Neural Networks (ConvNet). "Turkish movie reviews" dataset (composed of 34990 positive and negative movie reviews in Turkish) is used for the training and the evaluation of the model. The pre-processing of the model consisted of tokenization, cleaning, correcting any misspelled words, and negation handling steps. The Word2Vec word embedding model was used for the feature extraction stage. In the classification part, CNN based architecture model with an embedding layer, GlobalMaxPool1D layer, Dropout layer, and a Dense layer was built to obtain sentiment classification results. Standard metrics were used for the evaluation part. The average F1 result achieved was 82.36%. It was also found that 64 was the optimal number of filters for the model.

Harisu et. al. in [54] proposed a paper that compared TML algorithms (RSVM, RANF, MAXE, SVMs, and DECT) and deep learning (DL) models which are built using three main DL models called RNN, CNN, and hierarchical attention network (HAN). They used stemmed Turkish Twitter data to perform the SA task on. They also tried to increase the size of training data by applying a few data augmentation techniques called shift, shuffle, and hybrid (a combination of shifting and shuffling). To evaluate their proposed model, different evaluation metrics are used such as ACC, AUC, F1S, and RTM. While the TML algorithms are better in training time and runtime compared to the DL algorithms, in the case of AUC, ACC and F1S metrics, the DL algorithms outperformed the TML algorithms.

The authors in [55] proposed to achieve better results by combining different types of word embeddings (Word2Vec, fastText, character-level embedding) with different deep learning methods (LSTM, GRU, BiLSTM, CNN). The evaluation metrics were accuracy, precision, recall, and F1 score. The proposed model was evaluated on a dataset which was collected from Twitter regarding GSM operators in Turkey. It consists of 17,289 Turkish tweets and has 3 different labels for the sentiment (positive, negative, neutral). Features were extracted and used by word embeddings and DL algorithms both separately and in combination. The pre-trained BERT, ALBERT, ELECTRA, and DistilBERT models were used since they have infrequently been used in Turkish literature. The best model, ELECTRA, achieved 98.38% accuracy (k=10) for the hotel dataset and 92.21% accuracy (k=10) for the movie dataset. It was discovered that using TFM with a transformer would yield better results.

Another up-to-date study has been proposed in [57] including a new pretraining objective known as SSP. The main advantage of the proposed new pre-training task over NSP and SOP is that it makes better use of the dataset and generates more training input from it. As a result, models can be trained with more steps than in NSP and SOP. Models trained with SSP and SOP outperformed models trained with NSP. This demonstrated that using NSP alone in training is insufficient and that the model can achieve better results by performing other tasks. The SSP models performed similarly to the SOP models but outperformed the SOP models in masked word prediction.

Performance Comparisons of Hybrid Based Approaches

Table 3 presents the Hybrid-Based SA studies including their approaches and accuracy results. In this table, Hybrid-Based Turkish SA studies are listed based on their year of publication, applied approach, the level at which it is performed, the size of the data used, and highest performance achieved in the study.

In order to make a more meaningful comparison between these studies, more information must be known about the dataset. Without the information in the dataset, all these results present a range of performance in SA. Onan [47] stands out with an impressive accuracy of 91%, while Aydın et al. [51] and Demir et al. [52] also achieved high accuracy rates. On the other hand, Dehkharghani et al. [3] and Çetin and Eryiğit [46] achieves slightly lower accuracy results. However, it's worth noting that the techniques available during the year of the study's publication can also have a significant impact on the results. Additionally, the choice of the dataset size, analysis level, and specific approach significantly contributes to the variations in outcomes. Notably, Onan [47] demonstrates exceptionally accurate results in this category by utilizing different word-embedding models. Word embedding models are very effective in SA because they include contextual and semantic understanding of words. Their pre-trained state, along with their ability to handle context, contributes to their high performance in SA tasks [2].

Table 3. Comparisons of hybrid based sentiment analysis studies

Study	Year	Level	Data Size	Performance Metric	Result
Gezici and Yanıkoğlu [4]	2018	Sentence	10k	Accuracy	73.7%
Aydın et al. [51]	2018	Sentence	2k	Accuracy	77.1%
Çetin and Eryiğit [46]	2018	Sentence	1k	Accuracy	72%
Dehkharghani et al. [3]	2019	Document	60k	Accuracy	68%
Demir et al. [52]	2019	Word	91k	Accuracy	81%
Erşahin et al. [18]	2019	Sentence	220k	Accuracy	74.90%
Onan [47]	2020	Sentence	21k	Accuracy	91%
Kılıç et al. [50]	2020	Sentence	2k	Accuracy	78.3%
Köksal and Özgür [59]	2021	Sentence	5k	Accuracy	72.9%

TURKISH SENTIMENT ANALYSIS WITH HYBRID APPROACHES

In this section, it is explained how polarity lexicons are constructed and studies that work on constructing polarity lexicons are analyzed and reviewed. Polarity lexicons can be constructed using different methods. These methods are classified as statistical and dictionary based methods. In statistical-based lexicon construction, the statistical structure of the dataset that is used to construct the lexicon is used. In dictionary-based lexicon construction, an external lexicon is used as a baseline in the construction process. In the following two studies, polarity lexicons were constructed using statistical-based approaches.

In the study by Sağlam et al. [58], a Turkish sentiment lexicon was created for SA. This study aimed to develop an existing Turkish sentiment lexicon using data from online news media. In the aforementioned literature, there is not much useful research about developing an existing lexicon. In many studies about SA in Turkish, the online sources are collected and sentimentally analyzed but the data that is used has not been converted into a sentiment lexicon. This research primarily aimed to create an extended Turkish sentiment lexicon that can be used in various different studies. For the base Turkish sentiment lexicon, "SWNetTR" was used. "SWNetTR" has a capacity of 27,000 words. The data was collected using GDELT datasets. From the GDELT dataset, 100,000 online news documents were randomly chosen. After this process, raw text inputs were extracted from the selected documents. To create the dictionary that will be merged with "SWNetTR", first the collected raw text sentences were tokenized. As a result of this process, 14,000 words were received. Then, using the "Zemberek" morphological tool, the stems of all the words were received. Following all these processes, polarity scores of all the words were determined. All the words along with their morphological analysis and polarity scores were saved to a file. This file was named SWNetTR-GDELT. The "SWNetTR-GDELT" and base lexicon, which is "SWNetTR", was compared to prevent having duplicates. From this comparison, it was shown that the newly created dictionary "SWNetTR-GDELT had 10,000 unique words that did not exist in the base dictionary "SWNetTR. These found 10,000 unique words were added to the "SWNetTR" dictionary with their polarity scores. This way, the capacity of" SWNetTR" was increased by 10,000 words. Before "SWNetTR" had 27,000 words and the increased version had a capacity of 37,000 words. This new extended dictionary was now called "SWNetTR-PLUS". After this process, the accuracy of "SWNetTR" and "SWNetTR-PLUS" were calculated. According to the results, "SWNetTR" had a 60.6% accuracy rate of polarity classification performance. The extended dictionary "SWNetTR PLUS" had an accuracy rate of 72.2%.

In the following study, a polarity lexicon was constructed using dictionary-based approaches. In this research by Ayvaz et al. [10] it has been tried to extract significant information data using SA to create a Turkish sentiment lexicon for the purpose of contributing to the Turkish SA. In this research, existing SA lexicons are analyzed and a new sentiment lexicon is created by extending the content of these existing lexicons using the retrieved data from social platforms. In addition to the existing lexicons, basic emojis and scoring structure was added to this newly created sentiment lexicon. Furthermore, to evaluate the effectiveness of this newly formed sentiment lexicon, SA was performed on data retrieved from Twitter with specific tags. The SA studies were performed on two topics. The SA methodology for this study is presented in Figure 7.

In [60], Altınel, Buzlu and İpek created two sentiment polarity lexicons for Turkish and implemented statistical-based semantic algorithms in Turkish SA tasks. The first sentiment polarity dictionary, possessing a size of 159,876 Turkish words, is built with the use of a translator. The second sentiment polarity dictionary, with a size of 84,744 Turkish words, is built using GDELT (Global Data on Events, Languages, and Tone). They implemented baseline state-of-the-art models in order to compare the performance between the proposed system and state-of-the-art models. According to the experiment results, the algorithms developed in this study are beneficial because they can achieve higher classification performance than the baseline models on the Turkish

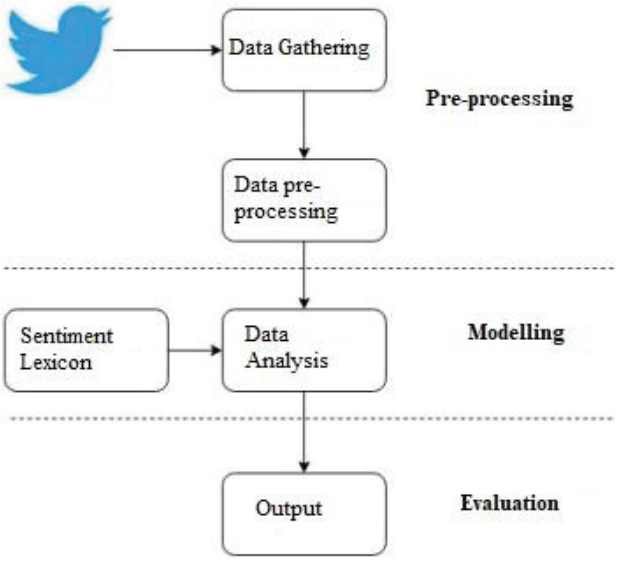


Figure 7. Sentiment analysis of the algorithm.

sentiment polarity detection task. In Figure 8. the representation of the GDELT website is presented.

POPULAR DATASETS AND LEXICONS

Public Datasets

In this section, the publicly available datasets used in the reviewed studies are collected and presented in Table 4. In this table, the information of the dataset type, the year of its creation, the language of the dataset, the size of the dataset, and the source link of the datasets can be found. To retrieve different types of data in multiple languages, the given sources and the table can be used.

Figure 8. The website of GDELT.

Most Used Lexicons in Turkish Sentiment Analysis

In Table 5, the most widely known lexicons in Turkish SA are presented. In this table, the names of the lexicons, their language, size, and information about the polarity state of the lexicons can be found. If a lexicon includes polarities of words, it is called a polarity lexicon. According to this table, the lexicon which has the highest size is the TDK lexicon which includes 616k words. But since the TDK lexicon does not include the information of word polarities, the lexicon has the highest size that also includes the information

Table 4. Public Datasets Used in Sentiment Analysis Studies

Dataset Type	Year	Language	Size	Source
Tweets	2021	Turkish	11k	https://www.kaggle.com/datasets/mrtbeyz/trke-sosyal-medya-paylam-veri-seti
Movie Reviews	2018	Turkish, English	348k	https://doi.org/10.1016/j.tele.2017.10.006
Lexicon	2018	Turkish	5k	GitHub - nazaan/Sentiment-Analysis
Movie Reviews	2017	Turkish	60k	http://myweb.sabanciuniv.edu/rdehkharghani/sentiment-analysis-in-turkish/
Text	2016	English	10k	https://github.com/dennybritz/cnn-text-classification-tf
Lexicon	2019	Turkish	616k	https://sozluk.gov.tr/
Text	2021	Multi-language	2Tb	https://www.gdeltproject.org/
Hotel, Movie Reviews	2016	Turkish	220k	http://humirapps.cs.hacettepe.edu.tr/tsad.aspx
Tweets	2019	Turkish	3k	http://www.kemik.yildiz.edu.tr/veri kumelerimiz.html
Tweets	2014	English	2k	https://drive.google.com/file/d/0BwPSGZHAP
Text	2020	Turkish	157	yoN2pZcVl1Qmp1OEU/view
Movie Reviews	2020	Turkish	105	http://www.kemik.yildiz.edu.tr/veri kumelerimiz.html
Survey(text)	2019	Turkish	27k	http://www.kemik.yildiz.edu.tr/veri kumelerimiz.html
Movie Reviews	2004	English	2k	http://demir.cs.deu.edu.tr/tremo-dataset/
Lexicon	2019	Turkish	13k	https://www.cs.cornell.edu/people/pabo/movie-review-data/
Lexicon	2019	Turkish	43k	https://github.com/sayvaz/turkish-lexicon/
Movie and Product Reviews	2013	Turkish	10k	http://demir.cs.deu.edu.tr/turkish-emotion-lexicon-tel-dataset/

Table 5. Most Used Lexicons in Turkish Sentiment Analysis

Name	Language	Size	Does it İnclude Polarity
SentiTurkNet	Turkish	14k	Yes
SenticNet	English	25k	Yes
Turkish WordNet	Turkish	77k	No
SentiWordNet	English	117k	Yes
TDK	Turkish	616k	No

of word polarities, which is the SentiWordNet lexicon, consisting of 117k words including polarities. Among the lexicons, which include the polarity information of words, the SentiWordNet lexicon has the maximum size which is 117k. Even though SentiWordNet has the largest size, SentiWordNet is an English dictionary. The second lexicon which has the highest size and includes polarity information of words is SentiTurkNet but compared to SentiWordNet, SentiTurkNet is a Turkish lexicon. The lexicon which has the largest size but does not include polarity information of words is TDK.

In Turkish SA, the most widely used lexicon is SentiTurkNet. Even though the TDK is the lexicon that has the largest size among other Turkish lexicons, the TDK library does not include polarities of words, which is why it is not commonly used in Turkish SA studies.

CONTROVERSIES, DISCUSSIONS AND COMPARISON OF THE APPROACHES FOR TURKISH SENTIMENT ANALYSIS

Researchers encounter numerous difficulties while implementing a sentiment text categorization methodology: The presence of base knowledge regarding a certain language: There are only a limited number of lexicon resources accessible for some languages (i.e., English, German, Chinese, etc.), and these languages are widely used languages internationally. In fact, the majority of languages lack their own lexical databases. This fact suggests that these knowledge-based systems can only be constructed for these specific languages. Additionally, knowledge-based systems for a specific language cannot be used effectively for another. Therefore, they are majorly language-dependent systems. The constantly expanding nature of language causes such resources to be typically expensive to be maintained and often unavailable in certain domains. Therefore, researchers should be encouraged to create knowledge bases for the other languages due to its effect on the improvement of the classification performance, or automatic translators should be made available.

The processing complexity of a knowledge base of this size: Knowledge-based automatic systems have substantial processing costs. These systems include the pre-processing part of a big corpus/data. This becomes an extra expense and radically grows according to the proportion of the corpus/data size. Researchers that implement these systems must optimize their algorithms/methodology to decrease the processing time/complexity.

Accessing excessive amounts of unlabeled/labeled data: DL algorithms are especially useful for supervised and unsupervised learning when there is a large amount of unlabeled data, and they typically learn data according to the weights in layers. The amount of data required for taking advantage of DL algorithms is sometimes difficult to access or collect. There are a lot of expenses associated with gathering, storing, and curating these data. Hardware computations: Implementing a machine learning-based system can be computationally expensive. In the training process of a machine learning-based system, the use of costly hardware resources may be a necessity. Prior to starting a project based on machine learning, this fact should be taken into account. Analysis: For a number of classification algorithms based on machine learning (e.g., decision trees, LR, and so on), it is possible to comprehend and explain the learned model, as well as the model's decision. This causes a number of issues in real-life applications. The privacy of the public must be considered by the researchers. To researchers that study SA, we recommend using various data mining methods to apply pre-processing and cleaning the data in their systems.

Furthermore, various ML algorithms can be used to detect and remove irrelevant data from large text corpora. However, in the methodology of the model, deciding to use a rule-based or lexicon-based methodology for sentiment classification is a difficult task that is dependent on certain factors such as the availability and size of the dataset and knowledge bases, as well as the problem itself. Different machine-learning methods have been used to produce a variety of text classification technologies. However, text classification offers more of a challenge since it contains semantic links between words, which are too complex for computers to model. For text categorization, algorithms to achieve high performance, it is essential to extract semantic relations in the correct forms. According to our analysis, selecting the best text sentiment classification method is dependent on the interconnected element size. Each category of approach has advantages over others as well as limitations as described above.

Current Challenges and Research Gaps

SA can be considered a difficult task to perform. SA is language dependent. The difficulty of SA is highly related to the structure of the language that is sentimentally analyzed. For languages similar to Turkish, performing SA can be more challenging compared to other languages due to the structure of the language. SA is a multiprocessing task, in order to perform SA many preprocessing stages must be applied to the data that will be used for SA. For preprocessing, many different techniques can be used to prepare the data for SA. These techniques are normalization, lemmatization, stemming and many more. All of these preprocessing stages are language-dependent as well, meaning that the process for these techniques vary between languages and, due to this reason, many challenges can occur in the process of SA for certain languages. If the preprocessing stages cannot be applied properly, the SA could not be performed effectively or accurately.

Turkish is a complex and rich language; it has an agglutinative structure. For languages that have an agglutinative structure, sentiment classification is a major problem. SA is more challenging to perform in these languages. Aside from the challenges resulting from the morphology and structure of languages, there are other difficulties that can cause major problems to the SA process. Even though SA is a very popular research topic in the Natural Language Analysis field, limited research is conducted on SA for the Turkish language. Due to this reality, there are also limited resources of data or other necessary resources to conduct Turkish SA. Additionally, privacy concerns may arise by the public that must be considered by researchers because the data extracted is often derived from people's posts and ideas from social media.

SA can be performed using various approaches such as lexicon-based, ML-based, and hybrid-based. For different approaches of SA, there are different requirements for resources. For dictionary-based approaches, an external lexicon is needed to perform SA and for ML-based and hybrid-based approaches, a sufficient amount of labeled data is necessary for performing SA. Thus, current resources may not be sufficient to source Turkish SA studies. Finding an adequately-sized Turkish polarity lexicon is difficult. The existing Turkish polarity lexicons are narrow in size. For languages like English, there are a wide variety of resources for SA. This widens the research area because with ready access to proper resources, conducting a SA study can be easier. Even though there are excessive amounts of data in online sources, processing and labeling these data requires sufficient hardware and this process is very time-consuming, contributing to the fact that there are limited resources for data as well. To summarize, the key points of Turkish SA are morphological complexity of Turkish, lack of resources and data privacy and ethics.

CONCLUSION

This paper presents an overview of Turkish SA. Many studies were compiled regarding Turkish SA. These studies were thoroughly reviewed and analyzed. In this paper, Turkish SA studies were categorized under four different categories according to their approaches. These studies are also categorized according to the level of SA performed. The SA studies were categorized into three different categories, these categories are dictionary-based SA, machine learning-based, and hybrid-based SA. Moreover, SA can be performed on four different sentiment levels which are word-level, sentence-level, document-level, and aspect-level. The main purpose of this survey was to collect different Turkish SA studies that use varying techniques and approaches and to analyze and compare the techniques of these studies to detect how different studies approach and propose solutions in the field of Turkish SA. Each Turkish sentiment study is reviewed and analyzed according to the approaches and techniques used to perform SA and the sources of datasets were presented. In each section, the studies are compared according to sentiment level, performance metric, and data size. SA is becoming a very popular research topic. In SA, three different approaches are used; these are lexicon-based, ml-based, and hybrid-based approaches. In summary; Lexicon-based methods offer simplicity and transparency, relying on predefined word lists. While these methods provide a base approach for the SA process, it is a manual process and it may struggle with the Turkish language's rich morphology, complex expressions, and the dynamic nature of language, which restrains their accuracy. Machine learning and deep learning techniques distinguish in capturing context and patterns in Turkish text. They have significantly improved the overall performance of Turkish SA, enabling a finer understanding of sentiments in various contexts, which is crucial for practical SA applications. Using ML techniques would be an improvement over using the lexicon approach alone. Lastly, with hybrid approaches, the aim is to combine the strengths of lexicon-based and ML-based techniques and it can enhance SA by incorporating domain-specific word lists while benefiting from machine learning models.

Using only lexicon-based methods is considered outdated due to all the reasons mentioned above. Instead, machine learning approaches, such as deep learning, are more effective as they have the ability to understand the context and can adapt well to certain situations. Hybrid methods combine ML's power with the benefits of sentiment lexicons and can offer a more balanced solution. Even though the lexicon approaches are outdated, we should not overlook its benefits to the SA process and, when combined with ML techniques, it can even outperform approaches that only employ ML techniques. This shift to the ML approaches is due to ML's increased ability and its future developments to handle the issues we encounter in languages such as idiomatic expressions, language structure etc.

Currently, interest in this field for non-English languages is still growing. In this study, we collected existing studies related strictly to Turkish SA to present a comprehensive study summarizing different approaches, and techniques used in Turkish SA. This survey can be a very helpful resource to gain detailed information about Turkish SA and to obtain general information about Turkish Sentiment analysis.

ACKNOWLEDGEMENT

This work was funded by the Scientific and Technological Research Council of Turkey (TÜBİTAK) with grant number 120E187. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of TÜBİTAK.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Girgin A, Gümüşçekiçci G. From past to present: Spam detection and identifying opinion leaders in social networks. Sigma J Eng Nat Sci 2022;40:441-463.

[2] Bilgin M. Classification of Turkish Tweets by Document Vectors and Investigation of the Effects of Parameter Changes On Classification Success. Sigma J Eng Nat Sci 2020;38:1581-1592.

[3] Dehkharghani R, Yanikoglu B, Saygin Y, Oflazer K. Sentiment analysis in Turkish at different granularity levels. Nat Lang Eng 2017;23:535-559. [CrossRef]

[4] Gezici G, Yanıkoğlu B. Sentiment analysis in Turkish. Turkish Nat Lang Process 2018;255-271. [CrossRef]

[5] Yıldırım E, Çetin FS, Eryiğit G, Temel T. The impact of NLP on Turkish sentiment analysis. Türkiye Bilisim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi. 2015;7:43-51.

[6] Alqaraleh S. Turkish sentiment analysis system via ensemble learning. Avrupa Bilim Teknol Derg 2020;122-129. [CrossRef]

[7] Can EF, Ezen-Can A, Can F. Multilingual sentiment analysis: An RNN-based framework for limited data. arXiv Prepr arXiv:1806.04511. 2018.

[8] Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: A survey. Ain Shams Eng J 2014;5:1093-1113. [CrossRef]

[9] Suat AT, Çınar Y. Borsa İstanbul'da finansal haberler ile piyasa değeri ilişkisinin metin madenciliği ve duygu (sentiment) analizi ile incelenmesi. Ankara Univ SBF Derg 2019;74:1-34. [CrossRef]

[10] Ayvaz S, Yıldırım S, Salman YB. Türkçe duygu kütüphanesi geliştirme: Sosyal medya verileriyle duygu analizi çalışması. Avrupa Bilim Teknol Derg 2019;16:51-60. [CrossRef]

[11] Vural AG, Cambazoglu B, Senkul P, Tokgoz ZO. A framework for sentiment analysis in Turkish: Application to polarity detection of movie reviews in Turkish. In: Computer and Information Sciences III. 2013; Springer; London; 437-445. [CrossRef]

[12] Çoban Ö, Özyer B, Özyer GT. Sentiment analysis for Turkish Twitter feeds. In: 2015 23nd Signal Processing and Communications Applications Conference (SIU). 2015; IEEE; 2388-2391. [CrossRef]

[13] Albayrak M, Topal K, Altıntaş V. Sosyal medya üzerinde veri analizi: Twitter. Suleyman Demirel Univ İktisadi ve İdari Bilimler Fakültesi Derg 2017;22(Kayfor 15 Özel Sayısı):1991-1998.

[14] Karaöz B, Gürsoy UT. Adaptif Öğrenme Sözlüğü Temelli Duygu Analiz Algoritması Önerisi. Bilisim Teknol Derg 2018;11:245-253. [CrossRef]

[15] Yüksel AS, Tan FG. Metin madenciliği teknikleri ile sosyal ağlarda bilgi keşfi. Mühendislik Bilimleri ve Tasarım Derg 2018;6:324-333. [CrossRef]

[16] Akgül ES, Ertano C, Diri B. Twitter verileri ile duygu analizi. Pamukkale Univ J Eng Sci 2016;22(2). [CrossRef]

[17] Aydın CR, Güngör T, Erkan A. Generating word and document embeddings for sentiment analysis. arXiv Prepr arXiv:2001.01269. 2020.

[18] Erşahin B, Aktaş Ö, Kilinc D, Erşahin M. A hybrid sentiment analysis method for Turkish. Turk J Electr Eng Comput Sci 2019;27:1780-1793. [CrossRef]

[19] Türkmenoglu C, Tantug AC. Sentiment analysis in Turkish media. In: International Conference on Machine Learning (ICML). 2014.

[20] Toçoğlu MA, Alpkocak A. Lexicon-based emotion analysis in Turkish. Turk J Electr Eng Comput Sci 2019;27:1213-1227. [CrossRef]

[21] Bayraktar K, Yavanoglu U, Ozbilen A. A Rule-Based Holistic Approach for Turkish Aspect-Based Sentiment Analysis. In: 2019 IEEE International Conference on Big Data (Big Data) 2019; IEEE; 2154-2158. [CrossRef]

[22] Ekinci E, Omurca Sİ. Ürün özelliklerinin konu modelleme yöntemi ile çıkartılması. Türkiye Bilisim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi 2017;9:51-58.
[23] Mutlu M, Özgür A. A Dataset and BERT-based models for targeted sentiment Analysis on Turkish Texts. arXiv Prepr arXiv:2205.04185. 2022. [CrossRef]
[24] Kilimci ZH. Borsa tahmini için Derin Topluluk Modelleri (DTM) ile finansal duygu analizi. Gazi Univ Mühendislik Mimarlık Fakültesi Derg 2020;35:635-650. [CrossRef]
[25] Onan A. Türkçe Twitter Mesajlarında Gizli Dirichlet Tahsisine Dayalı Duygu Analizi. 2017.
[26] Kaynar O, Yıldız M, Görmez Y, Albayrak A. Makine öğrenmesi yöntemleri ile Duygu Analizi. In: International Artificial Intelligence and Data Processing Symposium (IDAP'16). 2016;17-18.
[27] Aytekin YE, Keskin Ö. Türkiye'de faizsiz finans sisteminin duygu analizi bağlamında değerlendirilmesi. Uluslararası Islam Ekonomisi ve Finansı Araştırmaları Dergisi 2019;5:87-112.
[28] Kaya M, Fidan G, Toroslu IH. Sentiment analysis of Turkish political news. In: 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology. 2012; IEEE; Vol. 1; 174-180. [CrossRef]
[29] Ciftci B, Apaydin MS. A deep learning approach to sentiment analysis in Turkish. In: 2018 International Conference on Artificial Intelligence and Data Processing (IDAP). 2018; IEEE; 1-5. [CrossRef]
[30] Shehu HA, Tokat S. A hybrid approach for the sentiment analysis of Turkish Twitter data. In: The International Conference on Artificial Intelligence and Applied Mathematics in Engineering. Cham: Springer; 2019. p. 182-190. [CrossRef]
[31] Acikalin U, Bardak B, Kutlu M. Turkish Sentiment Analysis Using BERT. In: 2020 28th Signal Processing and Communications Applications Conference (SIU). 2020; IEEE; 1-4. [CrossRef]
[32] Aksu MÇ, Karaman E. FastText ve kelime çantası kelime temsil yöntemlerinin turistik mekanlar için yapılan Türkçe incelemeler kullanılarak karşılaştırılması. Avrupa Bilim Teknol Derg 2020;20:311-320. [CrossRef]
[33] Ballı C, Guzel MS, Bostanci E, Mishra A. Sentimental analysis of Twitter users from Turkish content with natural language processing. Comput Intell Neurosci 2022:2455160. [CrossRef]
[34] Baykara M, Gürtürk U. Sosyal Medya Paylaşımlarının Duygu Analizi Yöntemiyle Sınıflandırılması. In: 2nd International Conferance on Computer Science and Engineering 2017;911-916.
[35] Tuncer T, Çetintaş D. Bir sosyal ağdan alınan verilerin anlamsal kutuplandırılması. Bilgisayar Bilim 2019;4:1-6.
[36] Güran A, Uysal M, Doğrusöz Ö. Destek vektör makineleri parametre optimizasyonunun duygu analizi üzerindeki etkisi. Dokuz Eylül Univ Müh Fak Fen Müh Derg 2014;16:86-93.
[37] Aytekin Ç, Bayram MA. Türkçe metinler için duygu analizi yaklaşımı ile iletişimde bağlamdan bağımsız modellerin geliştirilmesi üzerine bir araştırma: karma veri modeli önerisi. Yeni Medya Elektron Derg 2021;5:12-25.
[38] Aytuğ O. Twitter mesajları üzerinde makine öğrenmesi yöntemlerine dayalı duygu analizi. Yönetim Bilisim Sistemleri Derg 2017;3(2):1-14.
[39] Sarıman G, Mutaf E. Covid-19 sürecinde Twitter mesajlarinin duygu analizi. Euroasia J Math Eng Nat Med Sci 2020;7:137-148. [CrossRef]
[40] Karabulut YE, Küçüksille EU. Twitter profesyonel izleme ve analiz aracı. Teknik Bilim Derg 2018;8:17-24.
[41] Kaynar O, Aydın Z, Görmez Y. Sentiment analizinde öznitelik düşürme yöntemlerinin oto kodlayıcılı derin öğrenme makinaları ile karşılaştırılması. Bilisim Teknol Derg 2017;10:319-326. [CrossRef]
[42] Alpkoçak A, Tocoglu MA, Çelikten A, Aygün İ. Türkçe Metinlerde Duygu Analizi için Farklı Makine Öğrenmesi Yöntemlerinin Karşılaştırılması. Dokuz Eylül Univ Müh Fak Fen Müh Derg 2019;21:719-725. [CrossRef]
[43] Tuzcu S. Çevrimiçi kullanıcı yorumlarının duygu analizi ile sınıflandırılması. Eskisehir Turk Dünyası Uygulama ve Araştırma Merkezi Bilisim Derg 2020;1:1-5.
[44] Kemaloğlu N, Küçüksille E, Özgünsever ME. Turkish Sentiment Analysis on Social Media. Sakarya Univ J Sci 2021;25:629-638. [CrossRef]
[45] Bharadiya J. A comprehensive survey of deep learning techniques natural language processing. Eur J Technol. 2023;7:58-66. [CrossRef]
[46] Çetin FS, Eryiğit G. Türkçe hedef tabanlı duygu analizi için alt görevlerin incelenmesi-hedef terim, hedef kategori ve duygu sınıfı belirleme. Bilisim Teknol Derg 2018;11:43-56. [CrossRef]
[47] Onan A. Sentiment Analysis in Turkish Based on Weighted Word Embeddings. In: 2020 28th Signal Processing and Communications Applications Conference (SIU). 2020; IEEE; 1-4. [CrossRef]
[48] Aktaş Ö, Coşkuner B, Soner İ. Turkish Sentiment Analysis Using Machine Learning Methods: Application on Online Food Order Site Reviews. arXiv Prepr arXiv:2201.03848. 2022.
[49] Seker SE. Sentimental analysis. YBS Ansiklopedi. 2016;3:21-36.
[50] Kılıç G, Budak I, Kılıç BS. Kara cuma etiketlerinin Tweet istatistikleri ve duygu analizi ile sıralanması. Selçuk Univ Sosyal Bilimler Meslek Yüksekokulu Derg 2020;23:131-140. [CrossRef]

[51] Aydın İ, Salur MU, Başkaya F. Duygu analizi için çoklu popülasyon tabanlı parçacık sürü optimizasyonu. Türkiye Bilisim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi. 2018;11:52-64.

[52] Demir Ö, Chawai AIB, Doğan B. Türkçe metinlerde sözlük tabanli yaklaşimla duygu analizi ve görselleştirme. Int Period Recent Technol Appl Eng 2019;1:58-66. [CrossRef]

[53] Alqaraleh S. Novel Turkish Sentiment Analysis System using ConvNet. 2021. [CrossRef]

[54] Shehu HA, Sharif MH, Sharif MHU, Datta R, Tokat S, Uyaver S, et al. Deep sentiment analysis: a case study on stemmed Turkish Twitter data. IEEE Access 2021;9:56836-56854. [CrossRef]

[55] Salur MU, Aydin I. A novel hybrid deep learning model for sentiment classification. IEEE Access 2020;8:58080-58093. [CrossRef]

[56] Guven ZA. The Comparison of Language Models with a Novel Text Filtering Approach for Turkish Sentiment Analysis. Trans Asian Low-Resource Lang Inf Process 2022;55:1-6. [CrossRef]

[57] Sonmezoz K, Amasyali MF. Same sentence prediction: A new pre-training task for BERT. In: 2021 Innovations in Intelligent Systems and Applications Conference (ASYU). 2021; IEEE; 1-6. [CrossRef]

[58] Sağlam F, Sever H, Genç B. Developing Turkish sentiment lexicon for sentiment analysis using online news media. In: 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA). 2016; IEEE; 1-5. [CrossRef]

[59] Köksal A, Özgür A. Twitter dataset and evaluation of transformers for Turkish sentiment analysis. In: 2021 29th Signal Processing and Communications Applications Conference (SIU). 2021; IEEE; 1-4. [CrossRef]

[60] Altınel AB, Buzlu K, İpek K. Performance analysis of different sentiment polarity dictionaries on Turkish sentiment detection. In: 2022 International Conference on Innovations in Intelligent Systems and Applications (INISTA). 2022;1-6. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2023.00100

Review Article

Evaluation of properties for synthetic polymers in medicine

Fatma İrem ŞAHİN[1], Nil ACARALI[1,*]

[1]*Department of Chemical Engineering, Yildiz Technical University, Esenler, İstanbul, 34220, Türkiye*

ARTICLE INFO

Article history
Received: 29 December 2022
Revised: 02 March 2023
Accepted: 28 March 2023

Keywords:
Biocompatible; Biodegradable; Health; Polymer; Synthetic

ABSTRACT

Today, the transformation of polymers into useful structures for the human body in medical field has been an interesting subject that affected everyone. Synthetic polymers have a wide range of uses in the health sector such as coating, cardiovascular, orthodontic surgery, tissue engineering, implant, and drug carrier with the development of technology. These polymers are known as polymers with various characteristics and applications artificially synthesized in accordance with chemical and thermodynamic laws. The polymers in health sector have a share of 41% in drug and release studies, 18% in treatment with therapy applications, 10% in vaccine production, 31% in studies on new approaches in this field. Synthetic polymers have ability to be produced cheaply and easily in large quantities. In this study, synthetic polymers such as polyethylene glycol, polyvinyl alcohol, polyurethane, polyolefin, polytetrafluoroethylene, silicone, polyvinyl chloride, poly-methyl methacrylate, polyester, polyamide and poly-lactic acid were investigated and details regarding the applications with explications of polymers were provided. It was seen that synthetic polymers could be evaluated in treatment of cancer and chronic diseases by determining the most appropriate methods and techniques with biocompatible, biodegradable, non-toxic materials.

Cite this article as: Şahin Fİ, Acaralı N. Evaluation of properties for synthetic polymers in medicine. Sigma J Eng Nat Sci 2024;42(4):1315–1324.

INTRODUCTION

The macromolecules referred to as polymers were synthesized by the chemical bonding of numerous smaller molecules, or repeating units known as monomers. In history, polymers were used to bind wooden parts together. Natural rubber used as one of the first polymers was a biopolymer. It was obtained as an adhesive from a tree species. In the 19th century, its properties were developed by Charles Goodyear and started to be evaluated for tire production. The first synthetic polymer was synthesized in 1907 as Bakelite. Later, it was turned into a commercial product and used in the production of molded parts such as handles, pipes, valve parts, knives, and buttons [1, 2].

Polymers were basically divided into two categories as natural and synthetic. Starch, lignin, cellulose, and rubber were just a few of the natural polymers that had been present since the existence of life. Synthetic polymers could be described as relatively modern materials as the polymers entered daily life after technological developments. The greatest progress in man-made synthetic polymers occurred in the 20th century. As components of many objects,

***Corresponding author.**
*E-mail address: nbaran@yildiz.edu.tr

This paper was recommended for publication in revised form by Editor in-Chief Ahmet Selim Dalkilic

polymers took place in many sectors such as textiles, paints, maintenance products, electrical tools, coatings and automotive sector in the daily life of modern society [3].

While engineering-based properties such as hardness, tensile stability and elasticity were often the primary properties in the selection of the polymer, toxicity and biocompatibility aspects were also considered especially in polymers for use in healthcare. Many natural and synthetic polymer types were encountered in the literature [4, 5] (Table 1).

In this study, general properties and usage areas of synthetic polymers such as polyethylene glycol, polyvinyl alcohol, polyurethane, polyolefin, polytetrafluoroethylene, silicone, polyvinyl chloride, poly-methyl methacrylate, polyester, polyamide and poly-lactic acid were described with an emphasis. In addition, it was aimed to form an idea about the application areas in the health sector and the latest developments in this field. Different from the literature, the studies conducted separately in the health sector were evaluated in drug and release studies, treatment, therapy, vaccine production, and new approaches in this field. It was thought that it would raise awareness about studies on cancer and chronic diseases in the field of medicine.

SYNTHETIC POLYMERS

Beverage bottles, cleaners, rugs, rubbers, adhesives, paints, toys, films, packaging materials, lubricants, food packaging, pharmaceuticals, and many common household items were produced from synthetic polymers. Polymers, some of which were in the form of composites to strengthen their structures, were being used as protective and structural materials by replacing wood and metals [6].

Staudinger realized that natural rubber had very high molecular weights in 1920. The modern era of synthetic polymers began with the discovery of nylon by Carothers in 1937. The fastest progress in the historical development of synthetic polymers occurred during World War II. With the invention of devices with improved technical features, an environment had been prepared for synthesizing more complex polymers. Recently, studies showed that advanced polymers as shape memory polymers could change control mechanism according to the magnetic field, light, temperature, and pH [7].

Synthetic polymers, unlike natural polymers, were not readily available and obtained under laboratory conditions through hydrocarbon building blocks. Synthetic polymers were an important component of biomaterials used as implant materials, cardiovascular therapeutic stents, orthodontic treatments, drug delivery devices, and tissue engineering scaffolds or hydrogels [8, 9].

The deterioration of polymers could be done in a controlled attitude. It was also physically sound, with features such as higher modulus strengths, sensitive ability to flex and flexibility, confidence in chemicals, and recyclability. According to their properties, molecular weight and structure, polymerization, connectivity, functionality, and easy synthesis could be distinguished. They could be classed in accordance with hydrophobicity, hydrophilicity and degradability. Polyethylene glycol, polyvinyl alcohol, polyurethane, polyolefin, polytetrafluoroethylene, silicone, polyvinyl chloride, poly-methyl methacrylate, polyester, polyamide and poly-lactic acid which had very advantageous due to the properties of synthetic polymers could be easily changed according to the type of application to be made [3].

Ethylene glycol monomer was a colourless, odourless, low-viscosity liquid. Polyethylene glycol was a linear polyether generally described by a molar mass of less than 20 000 g. Its usage included a broad range of industries, including cosmetics, pharmaceuticals, food producing, and inks. It was used as a thickening agent such as hydro alcoholic gels

Table 1. Overview of Natural and Synthetic Polymers

Type	Polymer	Properties & Applications	References
Natural	Natural Rubber	Biopolymer, adhesive, in tires	[1-5]
	Starch	Biodegradable, in food and pharmaceuticals	
	Lignin	Structural component in plants, in paper and biofuels	
	Cellulose	Found in plant cell walls, in textiles, paper, and as a food additive	
Synthetic	Polyethylene Glycol (PEG)	Water-soluble, in pharmaceuticals, cosmetics, and medical devices	[3-5]
	Polyvinyl Alcohol (PVA)	Water-soluble, in adhesives, coatings, and as a film former	
	Polyurethane	Versatile, in foams, coatings, and elastomers	
	Polytetrafluoroethylene (PTFE)	Non-stick, chemical resistant, in coatings and gaskets	
	Silicone	Flexible, heat-resistant, in medical implants and sealants	
	Polyvinyl Chloride (PVC)	Rigid or flexible, in pipes, cables, and medical devices	
	Poly-methyl Methacrylate (PMMA)	Transparent, in lenses and displays	
	Poly-lactic Acid (PLA)	Biodegradable, in packaging and medical implants	

and special lubricants in medical products. It was commonly found in medical treatments and vaccines. Polyethylene glycol had optimal biological and physicochemical properties. Especially in the field of biopharmaceuticals, it was used as a hydrophilic and anti-fouling polymer due to its good biocompatibility. It was preferred as a skeleton material due to the facility. It had a limited capacity for protein binding, cell adhesion, immunogenicity, and antigenicity. It offered a significant opportunity for usage in applications involving wound healing. Today, drugs with polyethylene glycol-modified were approved for clinical applications [10-12].

Polyvinyl alcohol was an alcohol compound with a significant number of hydroxyl groups along its molecular chain as hydrophilic polymer. Polyvinyl alcohol was non-toxic and had good biocompatibility. It was a great ingredient for film-forming [13]. It was widely used in the biomedical field as a component of hydrogels. Polyvinyl alcohol was biocompatible, inexpensive, and could be degraded by organisms [14]. It was utilized to create the nanofibers required for tissue engineering and wound healing when mixed with other polymers like chitosan and polyhydroxy butyrate. In addition, polyvinyl alcohol microneedles were preferred for single-step DNA polyplex vaccine encapsulation as it had high transfection efficiency [15].

A type of polymer substance called polyurethane was formed, when polyol and combined isocyanate in combined polyether reacted. Polyurethane had superior durability and im-permeability compared to conventional building materials. It had a high rate of foaming and expansion, environmentally friendly, highly adaptable, low density [16]. Various chemical and physical characteristics were used to create polyurethanes. Polycarbonate-based polyurethanes containing aromatic or aliphatic components were utilized in medicine when aliphatic compounds had less biologic stability, polyester and polyether. Thermoplastic polyurethanes were flexible without the need for plasticizers. Polyurethane was polymer that had the widest working area and it was used in various combinations because of its high degree of flexibility and capacity to change into biodegradable forms like polyester-urethane urea. Polyurethanes, an important member of polymers, were widely utilized in several applications from fibers, foams to electronic devices, coatings, adhesives, elastomers and biomedical technologies [16, 17].

Polyolefins were shaped by the polymerization of olefin monomer molecules such as ethylene, styrene and vinyl chloride [18]. Both polyethylene (PE) and polypropylene (PP) were extremely stable. Polyethylene was generated in dissimilar molecular weights and crystallinity. Polyolefin elastomers were widely used in packaging, adhesives, and automotive equipment due to low price, good machinability, and flexibility. High-density polyethylene with a similar molecular weight was used to create stable devices for implantation. Its primary applications were the sliding surfaces of artificial joints. Polypropylene had similar features with polyethylene, and it was used as surgical suture substances [19].

Polytetrafluoroethylene had high chemical resistance, high temperature stability and insulation properties. It also had many useful features such as a non-wetting surface, a low coefficient of friction and poor adhesion [20, 21]. Polytetrafluoroethylene had an ethylene structure as for 4 covalently bonded fluorine molecules. It was widely used in industries, pharmaceuticals and laboratories. It was a non-degradable, highly hydrophobic material. It showed some tissue growth and inflammation in body. It was used as a vascular graft in medical industry [22].

Silicon existed in various silicate structures in bonded form with oxygen. The mechanical properties of silicones influenced by a gel structure and varied chain lengths and crosslinks in their -Si-O structure were transferred to rubber elastomer and side chains were altered. The most typical ones were polydimethylsiloxane. Silicones were hydrophobic and bio-stable elastomers that they did not need plasticizers. Although the material structure varies according to the connected group, it basically had properties such as low surface tension, corrosion resistance and water repellency. Silicones were used for plastic surgery, intraocular lenses, dialysis membranes, and intraocular lenses [23].

Polyvinylchloride (PVC) was one of the thermoplastic polymers with the highest demand volume. PVC was a polymer with an ethylene structure containing a covalently bonded chlorine. It was thermally unstable at processing temperatures. It was one of the most widely used polymeric materials in the production of medical devices. It required an appropriate amount of stabilizer and plasticizer in order to achieve the desired flexibility and softness in its production and application. Stabilizers were needed to avoid the autocatalytic cleavage of HCl and polymer degradation after heat treatment. Strict PVC was transformed by plasticizers into a supple polymer which was employed for extracorporeal tubing [24, 25].

Polymethyl methacrylate (PMMA) was synthesized with methyl methacrylate monomers by using the exothermic polymerization technique between two glass plates. It was found application in dentistry and orthopedics as it was transformed into very hard polymers through radical polymerization. Since it was easy to maintenance and had low production costs, it also found use in the paper, paint, and automobile industries. It was also used as a coating on contact lenses. Nanotechnology was shown great importance in developments such as dental materials and biomaterials. PMMA could replace inorganic glass with its low-density optical transparency, high mechanical properties, and recyclable thermoplastic properties. PMMA was biocompatible with human cells as more appropriate for transplants and dental use [26, 27].

Polyester (PET) was the world's largest synthetic substances in manufacturing. It had low manufacturing costs and numerous applications. Polyesters were preferred for biomedical areas due to their simple degradability by non-enzymatic hydrolysis at ester bonds. Biologically stable and biodegradable polyesters were used in biomedicine,

and most of the biodegradable polymers belong to the polyester family. These polymers were used for a wide range of purposes from solid bodies for orthopedic areas to medication coatings on vascular stents [28].

Polyamides offered properties such as good stability, stiffness, high tensile strength, flexibility, and shock resistance with balanced chemical and mechanical properties. Nylon was the most prominent synthetic polyamide in medical field. The flexibility of polyurethanes and the strength of nylon were combined in polyamide block copolymers with soft segments for improved elasticity as catheter balloon materials for angioplasty. Polyamides could be preferred to form skin contact due to feature of biocompatibility. It has recently been employed to create porous scaffolds for bone regeneration which has been highly load bearing [29, 30].

Acetaldehyde and lactic acid polymerization could be used to create poly-lactic acids (PLA) synthetically. It contained features including the incorporation of several chemicals such metals, metal oxides, natural substances, antibiotics, and antibacterial activities. Additionally, attributes like mechanical performance and low immunogenicity made it a popular choice for film synthesis for biomedical devices and food packaging. PLA was a biocompatible and degradable polymer with thermal stability and favourable mechanical properties. It was preferred due to usage of in vivo as it degraded directly by hydrolysis without the use of catalysts or enzymes [2, 31, 32]. Popular synthetic polymers as poly(ethylene glycol), poly(vinyl alcohol), polyolefins, polyurethane, silicone, polyacrylates and polyamides were used in various applications [8, 33, 34] (Table 2).

Table 2. General Properties and Fields of Common Synthetic Polymers

Polymer	General Properties	Field	References
Polyethylene Glycol (PEG)	Colorless, odorless, low-viscosity liquid, hydrophilic, biocompatible, in medical treatments and vaccines	Cosmetics, pharmaceuticals, food producing, inks, medical products, biopharmaceuticals, wound healing	[10-12]
Polyvinyl Alcohol (PVA)	Non-toxic, hydrophilic, good biocompatibility, film-forming	Biomedical field, hydrogels, tissue engineering, wound healing, single-step DNA polyplex vaccine encapsulation	[13-15]
Polyurethane	Superior durability, impermeability, high rate of foaming and expansion, environmentally friendly, low density, flexible	Various chemical and physical characteristics, building materials, biomedical technologies, electronic devices, coatings, adhesives, elastomers, biomedical technologies	[16-17]
Polyolefins	Low price, good machinability, flexibility, stable	Packaging, adhesives, automotive equipment, artificial joints, surgical sutures	[18-19]
Polytetrafluoroethylene (PTFE)	High chemical resistance, high temperature stability, insulation properties, non-wetting surface, low coefficient of friction, poor adhesion	Industries, pharmaceuticals, laboratories, medical industry	[20-22]
Silicone	Hydrophobic, bio-stable elastomers, low surface tension, corrosion resistance, water repellency	Plastic surgery, intraocular lenses, dialysis membranes	[23]
Polyvinyl Chloride (PVC)	Thermally unstable at processing temperatures, requires stabilizers and plasticizers for flexibility and softness	Medical devices, extracorporeal tubing	[24-25]
Polymethyl Methacrylate (PMMA)	Low-density optical transparency, high mechanical properties, recyclable thermoplastic	Dentistry, orthopedics, paper industry, paint industry, automobile industry, contact lenses	[26-27]
Polyester (PET)	Low manufacturing costs, biodegradable	Biomedicine, orthopedic areas, medication coatings, vascular stents	[28]
Polyamides (Nylon)	Good stability, stiffness, high tensile strength, flexibility, shock resistance	Medical field, angioplasty catheter balloons, skin contact materials, bone regeneration	[29-30]
Poly-lactic Acid (PLA)	Biocompatible, degradable, thermal stability, favorable mechanical properties	Biomedical devices, food packaging, film synthesis	[2, 31-32]

USAGE OF POLYMERS IN HEALTH

Synthetic polymers were revolutionized the healthcare industry as well as known areas to community such as packaging, construction, textiles, and electronics. Polymer substances were used in tissue regeneration with the first plastic syringe developed. Polymeric materials had contributed to the development of regenerative medicine which envisioned the future of tissue engineering and organ transplants in the light of many years of work on stem cells and 3D printing. It was also available as nanocarrier, scaffold, fibers, lenses, antibacterial dressing materials for biomedical applications [32]. Polymers used in health sector could be divided into two broad categories. Pharmaceutical delivery systems and soluble and insoluble pharmaceuticals were part of the first category of synthetic polymers. Soluble blood plasticizers and insoluble prosthetics were part of the second category of synthetic polymers [35]. The use of synthetic polymers in various health disciplines and their dispersion percentages in polymer matrices [15] were given in Figure 1.

Basic criteria in polymers were processability, reproducibility, adjustability and cost-effectiveness. If material was to be used for therapeutic or regenerative purposes, the properties of the material such as biocompatible, non-immunogenic and non-toxic were considered [29]. One of the most important issues in polymers used in the field of health was the issue of biological compatibility. Since it was in contact with the body for a long time, it could be in harmony with tissues and organs. The material to be used required a small volume of high-quality materials. When the polymer was in contact with the body for an extended period in terms of sutures, vascular grafts, soft tissue replacements and implanted medication delivery, it became extremely important. It was used for the intended interaction between the therapeutic structure and the affected area. In the last 25 years, attempts were made to develop synthetic biomaterials, and many of these efforts were centred on preparing ready-to-use polymers which were biologically inert and steady in biological environments [1, 36]. The stimulus responsiveness, shape memory behaviour, and self-healing ability of polymeric substances were significant features for tissue engineering, medical devices, and cell therapy. Designing these polymeric materials also aided in the development of injectable hydrogels and controlled-release drug delivery systems [37].

Polymers used for therapeutic purposes in drug delivery systems exhibited an innovative and highly attractive approach by acting as a targeted active substance carrier. Polymers played significant role in the pharmaceutical industry because of ability in order to encapsulate or deliver therapeutic agents to several parts of the body. One of the main purposes of using polymers was to slow the rate at which drug molecules were exposed to the aqueous environment around the drug delivery system. For this purpose, stimulus-sensitive hydrogels obtained by cross-linking with greater water holding capacity were synthesized [38]. Polymers protected drugs from several external environments that influence the half-life. It was commonly used in the advancement of immediate or regulated drug delivery system using several mechanisms such as diffusion method, degradation and swelling method. It was adjusted according to its response to radiation, redox potential, pH, and temperature, magnetic and electric fields [39]. In a nano drug delivery system study, better retention of drug-loaded polymeric micelle in the tumour, prolonged circulation time, and reduced extra-site effects were achieved at the cancer site [40]. Models that could precisely forecast the rates at which integrated medicines would release from polymers and their degradation patterns were also necessary. In addition, future injectable systems based on liquid polymers that could be administered intravenously or in the form of injectable particles were the subjects of ongoing studies [41]. Several polymer applications in health [42] were shown in Figure 2.

Many polymer materials have been developed in last years, such as nanodrugs, micelles, vesicles, dendrimers, nanospheres, nanogels polymers and other structures. For example, in a study, single-chain polymer nanoparticles had properties such as multifunctional, heat-sensitive, light-sensitive, and catalytic activity. Polymer nanodrugs were used to arrange with aimed delivery and drug-controlled release functions that responded to stimuli. Conjugation of polymers and drugs could modify the pharmacological characteristics of

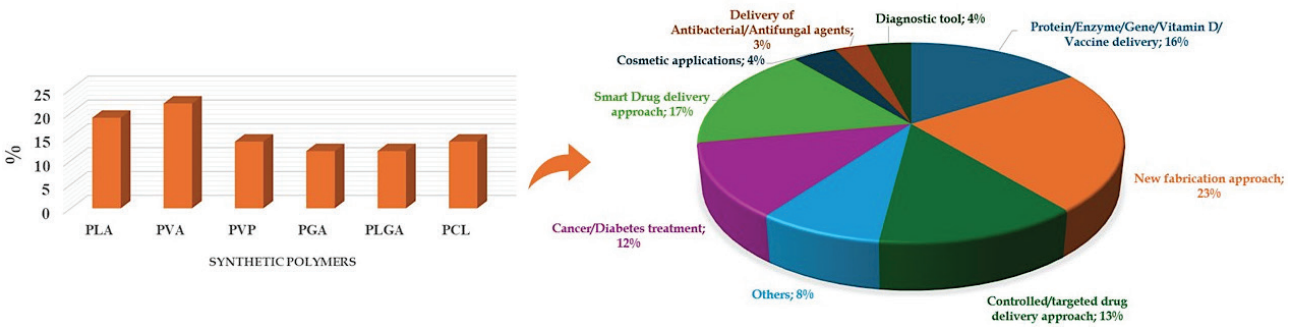


Figure 1. The use of synthetic polymers in various health disciplines and their dispersion percentages in polymer matrices.

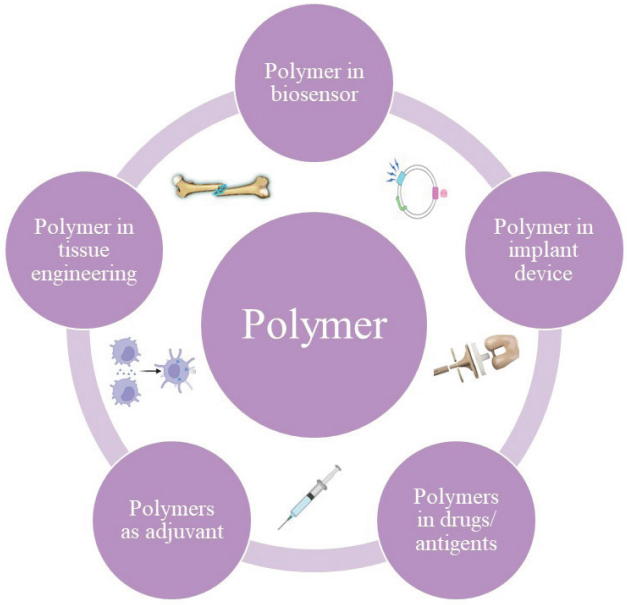


Figure 2. Several polymer applications in health.

drugs by increasing their solubility in drugs. It could improve drug delivery to the target area to be treated. Due to the design flexibility of synthetic polymers, features of polymer nanodrugs including size and porosity could be changed by varying the polymer's length and functionalization. Polymer nanodrugs provided properties such as responsiveness to stimuli, targeting ligands, and biological activity. Thus, it enabled intelligent and predictable transport in pathological areas. Polymer nanodrugs had obtained more attention due to structural diversity and simplicity of functionality. It was simple to functionalize polymer nanodrugs, and functional polymer nanodrugs were produced in a variety of shapes by modifying the structure and physicochemical characteristics of polymerized monomers. Polymeric vesicles among polymer nanodrugs were able to simultaneously load hydrophilic and hydrophobic drugs because of resemblance to liposome vesicles in structure [43]. Various tissue engineering techniques developed using polymer [44] were given in Figure 3.

Tissue engineering field took advantage of the structural and functional properties of polymers such as high surface area, good surface energy, controllable softness on surfaces and ease of manufacture. Studies on various polymer types conducted in tissue engineering to aid or mend cell adhesion, growth, proliferation, and adherence to surfaces. Development of cardiovascular tissue engineering devices was available in areas such as useful and biocompatible polymeric films, surface modification for interface with tough biological tissue. Biodegradable aliphatic polyesters were the synthetic polymers for scaffolds that were most frequently employed. These degradable polyesters were made from caprolactone, lactide, and glycolide. Polyurethanes were one of the most widely used polymers, especially in blood-contact biomedical applications, due to easy synthesis of different forms and the non-coagulating properties of polymer surfaces compared to other polymers. Some of its current applications included catheters, blood bags and artificial heart systems. Polyurethane fibers produced by electric field fiber spinning showed great improvements in the field of wound healing application. The first application of polycaprolactone and polylactic acid polymers was nanofiber scaffolds obtained by electric field fiber spinning in tissue engineering. PCL was very slow to degrade so it was used as a base material for the development of long-term implants due to its semi-crystalline nature and hydrophobicity. It was known that nano mats produced by electric field fiber drawing provided good support during the growth of vascular smooth muscle

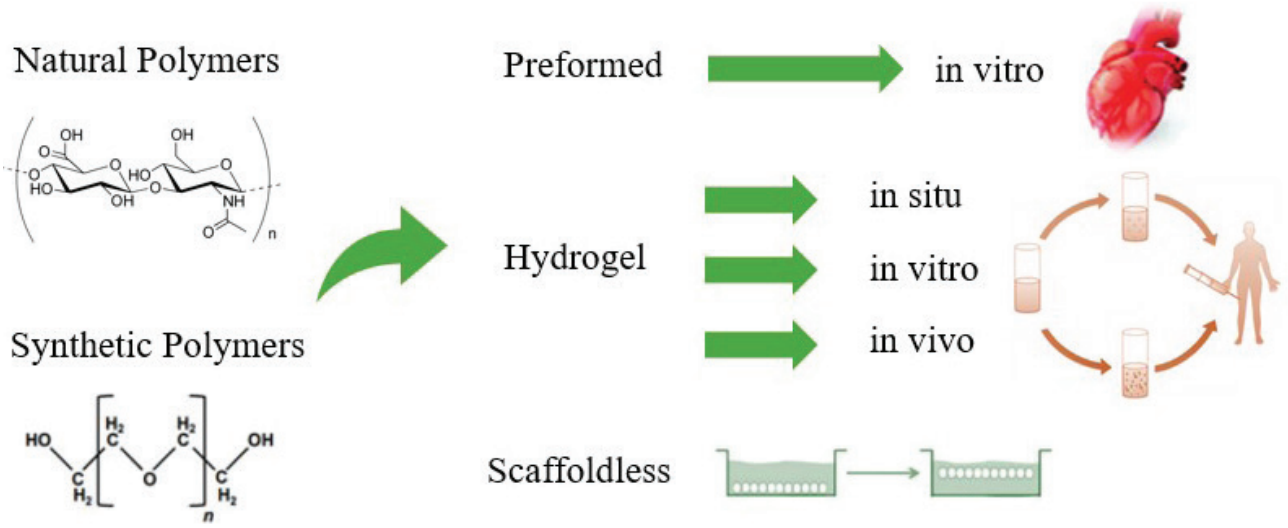


Figure 3. Various tissue engineering techniques developed using polymer.

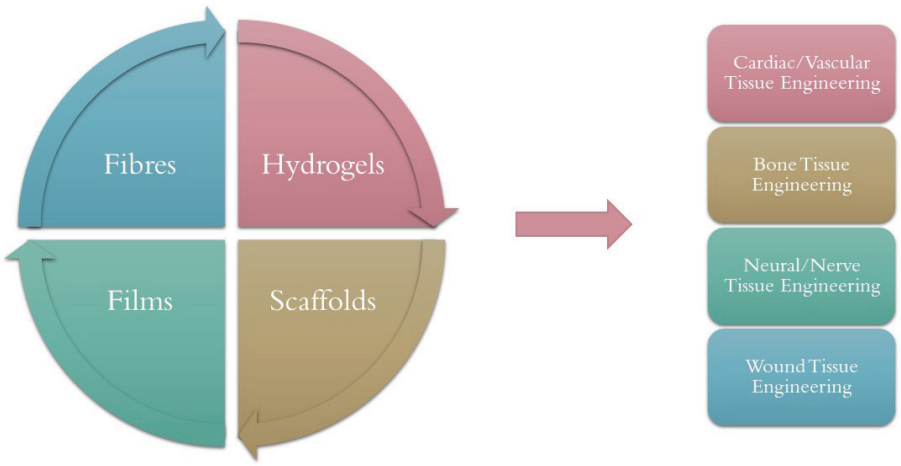


Figure 4. Major conductive polymers and in their tissue engineering usage

cells. The blood could be oxygenated using polymers and used in microporous shapes to function without harming the tissue [45-48]. Major conductive polymers and in tissue engineering usages [49] were shown in Figure 4.

In a research, 3D bioprinting was employed to create scaffolds from chitosan and PVA for liver tissue engineering. It showed higher cell viability in this structure compared to gels crosslinked by other methods [50]. In bone tissue engineering, the use of synthetically synthesized polymers started to replace traditional treatments as alternative techniques. Scaffolds obtained from biomaterials could be used in biomedical applications to extracellular similarities. Especially in this area, studies were continuing different metal or ceramic material combinations that could adapt to the skeleton system and structure to increase durability and improve properties [51]. In a study using PMMA and PEG, it was revealed that they could be used as a biomaterial for orthopedic applications in bone tissue engineering in the presence of gel phase appropriate mechanical and cellular properties [52]. Nanocomposites have emerged as an effective strategy to enhance the structural and functional properties of synthetic polymers in recent years. Studies on the preparation of nanocomposite polymers based on metal or carbon nanostructures in organic and inorganic nanofillers were focused to prepare new biomaterials with advanced properties. The key approach for treating nanocomposite materials was to increase the interfacial adhesion between polymers and nanostructures [53]. Synthetic polymers had received more and more interest as therapeutic agents, due to superior pharmacokinetic profile compared to small molecule medicines. These polymer therapies will probably result in the development of new medications with increased efficacy against a variety of ailments [54, 55]. In various health disciplines, the usage areas of polymers were given in Table 3.

Table 3. The Usage of Polymers in Various Health Disciplines

Health Discipline	Polymer	Usage	References
Pharmaceutical Delivery Systems	Synthetic Polymers	Drug delivery systems, encapsulation, controlled-release drug delivery	[35]
Soluble Pharmaceuticals	Blood Plasticizers	Soluble pharmaceutical applications	[35]
Insoluble Pharmaceuticals	Prosthetics	Insoluble pharmaceutical applications	[35]
Therapeutic/Regenerative	Biocompatible Polymers	Sutures, vascular grafts, soft tissue replacements, implanted medication delivery	[29]
Drug Delivery	Stimulus-sensitive Hydrogels	Targeted active substance carriers, immediate/ controlled drug delivery systems	[38]
Cancer Treatment	Polymeric Micelles	Drug-loaded polymeric micelles, prolonged circulation time, reduced extra-site effects	[40]
Injectable Systems	Liquid Polymers	Injectable hydrogels, injectable particles	[41]
Nanodrugs	Various Nanopolymers	Stimuli-responsive drug delivery, increased drug solubility, targeted drug delivery	[43]
Tissue Engineering	Polyurethane, PCL, PLA	Scaffold development, vascular smooth muscle cell growth, wound healing	[45-48]

Table 4. Applications and Advancements in Healthcare

Application	Advancements	References
Regenerative Medicine	Contributed to tissue engineering and organ transplants; stem cells and 3D printing	[32]
Stimulus Responsiveness	Significant for tissue engineering, medical devices, cell therapy; led to development of injectable hydrogels and controlled-release drug delivery systems	[37]
Nano Drug Delivery	Achieved better retention of drug-loaded polymeric micelles in tumors; prolonged circulation time; reduced extra-site effects; injectable systems	[40, 41]
Polymer Nanodrugs	Multifunctional, heat-sensitive, light-sensitive, catalytic activity; targeted delivery, drug-controlled release, increased solubility, improved drug delivery	[43]
Cardiovascular Devices	Development of polymeric films, surface modification for tough biological tissue; aliphatic polyesters like caprolactone, lactide, glycolide	[44, 45-48]
Wound Healing	Polyurethane fibers produced by electric field fiber spinning; used in catheters, blood bags, artificial heart systems	[45-48]
Bone Tissue Engineering	Use of biomaterial scaffolds for extracellular similarities; metal/ceramic combinations for skeleton adaptation	[50, 51]
Orthopedic Applications	PMMA and PEG as biomaterials with appropriate mechanical and cellular properties	[52]
Nanocomposites	Enhanced properties of synthetic polymers using metal/carbon nanostructures to increase interfacial adhesion between polymers and nanostructures	[53]

RESULTS AND DISCUSSION

In summary, synthetic polymers entered lives rapidly in 20th century with the latest technological developments and many types of polymers were used in almost all areas of health. In this study, it was aimed to give an idea about synthetic polymers and usage in the field. Although problems such as toxicity were encountered in synthetic polymers, it could be controlled with new synthesis techniques. Since the purification of natural polymers was difficult and costly, their commercial production and use was also limited. Synthetic polymers were needed to meet the biochemical and biomechanical requirements of emerging technologies such as tissue engineering, medical regeneration, gene therapy, new drug delivery systems and implantable devices. Currently, research areas were developing where nanoparticles, nano drugs, imaging in the biological environment, and the synthesis of biomedically applicable polymers were combined to facilitate targeted delivery of agents. Especially in the treatment of cancer and chronic diseases with polymeric encapsulation techniques, the best methods and techniques will be developed by reacting and interacting with the host cells, tissues, and organs. Studies continue for the development of smart polymer materials focused on the cell or tissue that were desired to be directly improved and repaired. In near future, more benefits will be provided in the field of health from these polymers developed in line with the needs (Table 4).

CONCLUSION

The advancements facilitated the development of polymers with enhanced biocompatibility and functionality for a wide range of medical applications. The studies into nanocarriers, nanodrugs, and other innovative polymer-based systems highlighted the potential for further advancements in targeted drug delivery and therapeutic interventions. The integration of smart polymers and polymeric encapsulation techniques offered promising areas for improving treatment efficacy, particularly in the management of cancer and chronic diseases. Future research and development would likely continue to focus on optimizing polymer properties for specific medical applications, enhancing the precision and effectiveness of treatment modalities. Considering technological advancements and clinical needs, synthetic polymers would play an increasingly significant role in advancing healthcare solutions and addressing complex medical challenges. As a result, it was expected that this research study on synthetic polymers would shed light on future studies and research in the field of health.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Rodriguez F, Cohen C, Ober CK, Archer LA. Principles of Polymer Systems. 6th ed. New York: CRC Press; 2015. [CrossRef]

[2] Mehmood A, Raina N, Phakeenuya V, Wonganu B, Cheenkachorn K. The current status and market trend of polylactic acid as biopolymer: Awareness and needs for sustainable development. Mater Today Chem 2022;72:3049–3055. [CrossRef]

[3] Saldívar-Guerra E, Vivaldo-Lima E. Introduction to Polymers and Polymer Types. In: Saldívar-Guerra E, Vivaldo-Lima E, editors. Handbook of polymer synthesis, characterization, and processing. New Jersey: Wiley; 2013. p. 1–44. [CrossRef]

[4] Suamte L, Tirkey A, Jayasekhar Babu P. Design of 3D smart scaffolds using natural, synthetic and hybrid derived polymers for skin regenerative applications. Smart Mater Med 2022;4:243–256. [CrossRef]

[5] Sharma S, Aiswarya T, Mirza I, Saha S. Biocompatible polymers and their applications. In: Hashmi MSJ, editor. Encyclopedia of Materials: Plastics and Polymers. Amsterdam: Elsevier; 2022. p. 796–819. [CrossRef]

[6] Orchin M, Macomber RS, Pinhas AR, Wilson RM. The vocabulary and concepts of organic chemistry. New Jersey: John Wiley & Sons Inc; 2005. p. 291–342. [CrossRef]

[7] Xanthos M. Polymers and polymer composites. In: Xanthos M, editor. Functional Fillers for Plastics. New Jersey: Wiley; 2005. p. 1–16. [CrossRef]

[8] Maitz M. Applications of synthetic polymers in clinical medicine. Biosurf Biotribol 2015;1:161–176. [CrossRef]

[9] Okamoto M, John B. Synthetic biopolymer nanocomposites for tissue engineering scaffolds. Prog Polym Sci 2013;38:1487–1503. [CrossRef]

[10] Noel J, Jannot Y, Métivier C, Sgreva NR. Thermal characterization of polyethylene glycol 600 in liquid and solid phase and across the phase transition. Thermochim Acta 2022;716:179326. [CrossRef]

[11] Hu J, Liu S. Emerging trends of discrete Poly (ethylene glycol) in biomedical applications. Curr Opin Biomed Eng 2022;24:100419. [CrossRef]

[12] Yue H, Zhao Y, Ma X, Gong J. Ethylene glycol: Properties, synthesis, and applications. Chem Soc Rev 2012;41:4218–4244. [CrossRef]

[13] Zhu L, Feng L, Luo H, Dong RS, Wang MY, Yao G, et al. Characterization of polyvinyl alcohol-nanocellulose composite film and its release effect on tetracycline hydrochloride. Ind Crops Prod 2022;188:115723. [CrossRef]

[14] Zhang R, Zhang D, Sun X, Song X, Yan KC, Liang H. Polyvinyl alcohol/gelatin hydrogels regulate cell adhesion and chromatin accessibility. Int J Biol Macromol 2022;219:672–684. [CrossRef]

[15] Koyani RD. Synthetic polymers for microneedle synthesis: From then to now. J Drug Deliv Sci Technol 2020;60:102071. [CrossRef]

[16] Li M, Fang H, Zhang C, Du M, Wang F. Study on the new polyurethane material suitable for foaming in water. Constr Build Mater 2022;354:129163. [CrossRef]

[17] Yuan L, Zhou W, Shen Y, Li Z. Chemically recyclable polyurethanes based on bio-renewable γ-butyrolactone: From thermoplastics to elastomers. Polym Degrad Stab 2022;204:110116. [CrossRef]

[18] Elgharbawy AS, Ali RM. A comprehensive review of the polyolefin composites and their properties. Heliyon 2022;8:e09932. [CrossRef]

[19] Cai Y, Zheng J, Hu Y, Wei J, Fan H. The preparation of polyolefin elastomer functionalized with polysiloxane and its effect in ethylene-propylene-diene monomer/silicon rubber blends. Eur Polym J 2022;177:111468. [CrossRef]

[20] Zhu C, Su Q, Wei Z, Wang X, Long S, Zhang G, et al. Poly (arylene sulfide sulfone)/polytetrafluoroethylene composite film: Thermal, mechanical, hydrophobic and low dielectric constant properties. Mater Lett 2023;330:133306. [CrossRef]

[21] Fitriani SW, Ikeda S, Tani M, Yajima H, Furuta H, Hatta A. Hydrophilization of polytetrafluoroethylene using an atmospheric-pressure plasma of argon gas with water-ethanol vapor. Mater Chem Phys 2022;282:125974. [CrossRef]

[22] Yang H, Wang Y, Wang Z, Yuan S, Niu C, Liu Y, et al. Effect of polytetrafluoroethylene nanoplastics on combined inhibition of ciprofloxacin and bivalent copper on nitrogen removal, sludge activity and microbial community in sequencing batch reactor. Bioresour Technol 2022;360:127627. [CrossRef]

[23] Chi H, Wang S, Li T, Li Z. Recent progress in using hybrid silicon polymer composites for wastewater treatment. Chemosphere 2021;263:128380. [CrossRef]

[24] Putrawan IDGA, Indarto A, Octavia Y. Thermal stabilization of polyvinyl chloride by calcium and zinc carboxylates derived from byproduct of palm oil refining. Heliyon 2022;8:e10079. [CrossRef]

[25] Polaskova M, Sedlacek T, Kasparkova V, Filip P. Substantial drop of plasticizer migration from polyvinyl chloride catheters using co-extruded thermoplastic polyurethane layers. Mater Today Commun 2022;32:103895. [CrossRef]

[26] Kaur H, Thakur A. Applications of poly (methyl methacrylate) polymer in dentistry: A review. Mater Today Proc 2022;50:1619–1625. [CrossRef]

[27] Thirumala Patil M, Lakshminarasimhan S, Santhosh G. Optical and thermal studies of host Poly (methyl methacrylate) (PMMA) based nanocomposites: A review. Mater Today Proc 2021;46:2564–2571. [CrossRef]

[28] Chen J, Jia K, Zhao Q, You H, Chen Z, Shi L, et al. Intelligent polyester metafabric for scalable personal hydrothermal self-adaptive adjustment. Chem Eng J 2023;451:138875. [CrossRef]
[29] Pugliese R, Beltrami B, Regondi S, Lunetta C. Polymeric biomaterials for 3D printing in medicine: An overview. Ann 3D Print Med 2021;2:100011. [CrossRef]
[30] Kumar Maurya A, Kumar S, Singh M, Manik G. Polyamide fiber reinforced polymeric composites: A short review. Mater Today Proc 2023;80:98–103. [CrossRef]
[31] Rao SH, Harini B, Shadamarshan RPK, Balagangadharan K, Selvamurugan N. Natural and synthetic polymers/bioceramics/bioactive compounds-mediated cell signalling in bone tissue engineering. Int J Biol Macromol 2018;110:88–96. [CrossRef]
[32] Terzopoulou Z, Zamboulis A, Koumentakou I, Michailidou G, Noordam MJ, Bikiaris DN. Biocompatible synthetic polymers for tissue engineering purposes. Biomacromolecules 2022;23:1841–1863. [CrossRef]
[33] Gaaz TS, Sulong AB, Akhtar MN, Kadhum AA, Mohamad AB, Al-Amiery AA. Properties and applications of polyvinyl alcohol, halloysite nanotubes and their nanocomposites. Molecules 2015;20:22833–22847. [CrossRef]
[34] Foster LJ. PEGylation and BioPEGylation of Polyhydroxyalkanoates: Synthesis, Characterisation and Applications. London, UK: IntechOpen; 2010. p. 243–249.
[35] Pierre TS, Chiellini E. Review: Biodegradability of synthetic polymers used for medical and pharmaceutical applications. J Bioact Compat Polym 1986;1:467–497. [CrossRef]
[36] Gunatillake P, Mayadunne R, Adhikari R. Recent developments in biodegradable synthetic polymers. Biotechnol Annu Rev 2006;12:301–347. [CrossRef]
[37] Huang HJ, Tsai YL, Lin SH, Hsu SH. Smart polymers for cell therapy and precision medicine. J Biomed Sci 2019;26:73. [CrossRef]
[38] Suhail M, Shih CM, Liu JY, Hsieh WC, Lin YW, Lin IL, et al. Synthesis of glutamic acid/polyvinyl alcohol-based hydrogels for controlled drug release: In-vitro characterization and in-vivo evaluation. J Drug Deliv Sci Technol 2022;75:103715. [CrossRef]
[39] Bhatt S, Pathak A, Grover P, Bharadwaj A, Bhatia D, Tomar R, et al. Different aspects of polymers - A review article. Mater Today Proc 2022;64:1490–1495. [CrossRef]
[40] Chaudhuri A, Ramesh K, Kumar DN, Dehari D, Singh S, Kumar D, et al. Polymeric micelles: A novel drug delivery system for the treatment of breast cancer. J Drug Deliv Sci Technol. 2022;77:103886. [CrossRef]
[41] Anju S, Prajitha N, Sukanya V, Mohanan P. Complicity of degradable polymers in health-care applications. Mater Today Chem 2020;16:100236. [CrossRef]
[42] Shakya AK, Nandakumar KS. Polymers as immunological adjuvants: An update on recent developments. J BioSci Biotech 2012;1:199–210.
[43] Zhou J, Zhang Y, Wang R. Controllable loading and release of nanodrugs in polymeric vesicles. Giant 2022;12:100126. [CrossRef]
[44] Ye KY, Black LD 3rd. Strategies for tissue engineering cardiac constructs to affect functional repair following myocardial infarction. J Cardiovasc Transl Res 2011;4:575–591. [CrossRef]
[45] Açarı İK, Sel E, Özcan İ, Ateş B, Köytepe S, Thakur VK. Chemistry and engineering of brush type polymers: Perspective towards tissue engineering. Adv Colloid Interface Sci 2022;305:102694. [CrossRef]
[46] Arif ZU, Khalid MY, Noroozi R, Sadeghianmaryan A, Jalalvand M, Hossain M. Recent advances in 3D-printed polylactide and polycaprolactone-based biomaterials for tissue engineering applications. Int J Biol Macromol 2022;218:930–968. [CrossRef]
[47] Joseph FPJ, Arun KJ, Navas A, Irene J. Biomedical applications of polymers -An overview. Curr Trends Biomed Eng Biosci 2018;15:555909. [CrossRef]
[48] Can N, Ersoy M. Nanofiber structured polymeric tissue scaffolds. J Text Eng 2014;21:37–50. [Turkish] [CrossRef]
[49] Bolívar-Monsalve J, Alvarez MM, Hosseini S, Espinosa-Hernandez MA, Ceballos-González CF, Sanchez-Dominguez M, et al. Engineering bioactive synthetic polymers for biomedical applications: A review with emphasis on tissue engineering and controlled release. Mater Adv 2021;2:4447–4478. [CrossRef]
[50] Bhatt SS, Thakur G, Nune M. Preparation and characterization of PVA/Chitosan cross-linked 3D scaffolds for liver tissue engineering. Mater Today Proc 2023 Mar 1. doi: 10.1016/j.matpr.2023.02.251. [Epub ahead of print]. [CrossRef]
[51] Narmatha Christy P, Khaleel Basha S, Sugantha Kumari V. Multifunctional organic and inorganic hybrid bionanocomposite of chitosan/poly(vinyl alcohol)/nanobioactive glass/nanocellulose for bone tissue engineering. J Mech Behav Biomed Mater 2022;135:105427. [CrossRef]
[52] Faruq O, Sarkar K, Lee B. Physicochemical property and cytocompatibility of HyA-PEG loaded PMMA based bone cement. Mater Chem Phys 2022;295:127142. [CrossRef]
[53] Armentano I, Dottori M, Chiralt A, Mattioli S, Kenny JM. Biodegradable polymer matrix nanocomposites for tissue engineering: A review. Polym Degrad Stab 2010;95(11):2126–2146. [CrossRef]
[54] Twaites B, de las Heras Alarcón C, Alexander C. Synthetic polymers as drugs and therapeutics. J Mater Chem 2005;15(4):441–455. [CrossRef]
[55] Dirauf M, Muljajew I, Weber C, Schubert US. Recent advances in degradable synthetic polymers for biomedical applications-Beyond polyesters. Prog Polym Sci 2022;129:101547. [CrossRef]