e-ISSN: 2602-3563

ACTA INFOLOGICA (ACIN)

DECEMBER, 2024
Volume: 8 | Issue: 2

ISTANBUL UNIVERSITY, INFORMATICS DEPARTMENT

acin.istanbul.edu.tr
https://dergipark.org.tr/tr/pub/acin

İSTANBUL UNIVERSITY PRESS

Acta Infologica
DECEMBER 2024,
VOLUME 8 ISSUE 2

İSTANBUL UNIVERSITY PRESS

Indexing and Abstracting

Web of Science – Emerging Sources Citation Index
TÜBİTAK-ULAKBİM TR Dizin
EBSCO Applied Sciences Source Ultimate
Erih Plus
DOAJ
Bielefeld Academic Search Engine (BASE)
OpenAIRE
ResearchBib
ASOS Index
Gale Cengage

Owner
Prof. Dr. Sevinç GÜLSEÇEN
Istanbul University, Department of Informatics, Istanbul, Turkiye

Responsible Manager
Prof. Dr. Sevinç GÜLSEÇEN
Istanbul University, Department of Informatics, Istanbul, Turkiye

Correspondence Address
Istanbul University, Department of Informatics
Kalenderhane Mahallesi, 16 Mart Şehitleri Caddesi, No: 8
Vezneciler, Fatih, İstanbul, Türkiye
Phone: +90 212 440 00 00/10037
E-mail: acin@istanbul.edu.tr
http://iupress.istanbul.edu.tr/tr/journal/acin/home
https://dergipark.org.tr/tr/pub/acin

Publisher
Istanbul University Press
İstanbul Üniversitesi Merkez Kampüsü, 34452 Beyazıt,
Fatih, İstanbul, Türkiye
Phone: +90 212 440 00 00

The publication language of the journal is English

This is a scholarly, international, peer-reviewed and open-access journal published biannually in June and December.

Publication Type: Periodical

İSTANBUL UNIVERSITY PRESS

EDITORIAL MANAGEMENT BOARD

İSTANBUL UNIVERSITY PRESS

EDITORIAL BOARD

Acta Infologica
DECEMBER 2024,
VOLUME 8 ISSUE 2

İSTANBUL UNIVERSITY PRESS

CONTENTS

İSTANBUL UNIVERSITY PRESS

PREFACE TO DECEMBER 2024 ISSUE

As editorial team of the journal we have started our mission with the idea to create a peer-reviewed journal to focus upon to rapidly moving and exciting interdisciplinary area of informatics where science, technology and mathematics combines and interacts. Nowadays digital twin and transfer learning, for example, are two powerful concepts in technology, often intersecting in applications related to machine learning, artificial intelligence, and real-world systems modeling. A digital twin is a virtual representation of a physical object, system, or process. It uses real-time data from sensors, IoT devices, and other sources to simulate, predict, and optimize performance. Digital twins are widely used in industries like manufacturing, healthcare, smart cities, and logistics. Transfer learning, on the other hand, can be used to create the AI models powering digital twins. For example, a pre-trained neural network can be fine-tuned to simulate a specific industrial machine.

As it is in this issue, the diversity and high standard work published in the journal is reflected in every future issue. As chief-editor I can say that the growth of the journal in size and quality is the greatest happiness of our team.

For Acta INFOLOGICA to continue to thrive with that standard and diversity we look forward to the submission and publication of novel and leading articles in the future.

I wish to thank all authors, readers and referees for their invaluable efforts.

Sincerely
Sevinç GÜLSEÇEN
Chief-Editor

İSTANBUL UNIVERSITY PRESS

Acta Infologica
acin 2024, 8 (2), 88–99
DOI: 10.26650/acin.1451569

Research Article

Redesigning Municipality Logos in the Context of Visual Reading Using Artificial Intelligence

Faruk Dursun[1]

[1]Sakarya University, Faculty of Management, Management Information Systems, Sakarya, Türkiye

Corresponding author : Faruk Dursun
E-mail : farukdursun@sakarya.edu.tr

ABSTRACT

The intensive use of technology in the way of business brings about different applications and usage areas. Artificial intelligence, on the other hand, reveals the last point that technology has reached and expands its usage area daily. Every example of artificial intelligence, which we can define as computer software that emerged to realise human abilities such as thinking, evaluation, problem solving by taking people out of the game or minimising their impact in jobs that technically require human elements, is surprising, and the point it will reach cannot be fully predicted. In this study, it is aimed to analyse the logos of institutions/organisations as their windows to the outside world and the first element that welcomes the audience by using visual reading and redrawing them using three artificial intelligence applications (Midjourney, Stable Diffusion and DALL-E 2). For this purpose, the logos of the 11 municipalities in the Marmara region were included in the scope of this study. When the results were evaluated, it was observed that the new visuals were quite successful. The visual meanings of the original logo were largely included in the new logo. Colours were also used in the new logos in a very close fashion to the original. The Midjourney AI application produced more effective drawings than the other two applications. On the other hand, it was observed that the elements of local culture were excluded from the drawings of the three AI tools.

Keywords: Artificial intelligence, visual reading, logo, visual design, artificial intelligence-design

Submitted : 12.03.2024
Revision Requested : 14.05.2024
Last Revision Received : 15.05.2024
Accepted : 27.06.2024
Published Online : 19.07.2024

1. INTRODUCTION

Artificial intelligence, which is the study of calculations that make perception, movement, and reasoning possible, has become one of the most popular topics in recent times. The use of artificial intelligence applications, which we encounter examples in many sectors, from medicine to industrial design, is becoming inevitable for many sectors. For example, chat robots integrated into websites and mobile applications facilitate the work and actions of users by creating a help menu that lists customer preferences and/or problems. New devices and software facilitate the work of users, aim to reduce the margin of error caused by human error, and provide cost advantages by reducing time and processes. Artificial intelligence, which has begun to be used effectively in the industrial dimension, has reached a dimension that can create visual design outputs. Users obtain drawings and visual representations from artificial intelligence, depending on the texts they provide. Tanughara (2023) used Midjourney, an artificial intelligence drawing module, in the process of developing architectural designs and argued that the drawings generated by artificial intelligence produce concept ideas. Çeken and Şen (2023), in their study on the use of artificial intelligence in the graphic design industry, argued that the integration of artificial intelligence and graphic design is an exciting process that has the potential to transform the creation and consumption patterns of visual media. Al-Qatry and Rady (2023) examined the use of artificial intelligence in the design of women's clothing and applied three artificial intelligence tools, Lexica, Dreamlikve, and Midjourney, and found that the Midjourney artificial intelligence tool was superior to the other two tools. Accordingly, in terms of validating the basics and elements of design and achieving the innovative and functional values of new designs, in addition to the impact of the AI tool used on the features of new designs, the Midjourney tool is the best tool in terms of its use in creating various designs for women's wear, followed by the Lexica tool and finally the Dreamlike tool. Borji (2022) studied the reproduction of faces in nature using three AI tools, Stable Diffusion, Midjourney, and DALL-E 2, and found that Stable Diffusion produced better faces than the other tools. Alawadhi and Yousef (2023) examined journey, an artificial intelligence tool, in the creation of plastic furniture in the context of aesthetics, functionality, and production technology and revealed the effectiveness of artificial intelligence in plastic furniture design. Göring et al. (2023a) compared the images produced through Midjourney and DALL-E 2 and found that the images produced through Midjourney were more realistic. Göring et al. (2023b) found that out of 135 images generated using artificial intelligence tools, the images generated using Midjourney and DALL-E 2 tools demonstrated higher performance than real images. Images generated using Glide and Craiyon had lower performance. Tsidylo and Sendra (2023), in their study on the use of artificial intelligence in the education of future designers, found that elements such as environment, lighting, colour, mood, and composition were more effective in Midjourney drawings than Stable Diffusion and DALL E 2 and argued that Midjourney is a suitable tool to use as a methodological innovation. Rozdolska (2022) examined the use of artificial intelligence tools in architecture and found that the drawing created using journey, an artificial intelligence tool, within the framework of a scenario was highly aesthetic and remarkable and therefore had high potential in the context of application. Fahim (2023) examined the use of artificial intelligence tools in packaging and packaging designs suitable for the different cultural textures of the market to be addressed in international product marketing and determined the potential of Midjourney in terms of designs suitable for the cultural characteristics of the target market. Hariffadzillah et al. (2023) examined the use of artificial intelligence tools in the creation of illustrated storybooks for children and found that the visuals produced by Midjourney and DALL-E tools encourage reading and learning. Bao and Xiang (2023) investigated the perspectives of architects and architecture students on the use of AI tools and found that AI can help architects to some extent achieve satisfactory performance. He also argued that the effective application of AI generators will significantly optimise the design process, allowing architects to explore more creative and aesthetic aspects. Çelik (2023) argued that AI tools such as journey, DALL-E 2, Stable Diffusion, Crayon, and Nightcafe, with their high data processing potential, can lead to a paradigm shift in architectural processes. Chen and Kao (2022), in their study of the midjourney-generated visuals of Chinese poet Tang Yin's poem on the painting "Lonely Fisherman on the River in Autumn", found that the generated visuals were not as detailed as the original paintings, but the result of the combination of painting styles and ambient lighting settings produced a visual that was as good as the original in terms of overall mood. Wasielewski (2023) used DALL-E and Midjourney AI tools in his study on hand and counting images and found that the views represented the data in plural form. Chen and Chen (2023) generated landscape expressions in the poem The Hard Road to Shu using the Midjourney AI tool and transferred artworks such as classical Chinese landscape paintings to AI models and found that they were successful in terms of stylistic features, composition, and cultural symbolism. Radhakrishnan (2023) argued that the use of AI tools such as Midjourney in architecture can play a critical role in ensuring that the future of architectural creativity is not interrupted. Hanna (2022) found that when creating artistic and creative advertisements, Midjourney offers different and impressive designs in the context of the words entered. Sánchez and Heras (2023) argued that the versatility of journey, an artificial intelligence

tool for teaching classical antiquity, is an effective and dynamic parameter for students' learning time. Shuhan et al. (2023) conducted a study on the perceptions of designers and design students from Zhejiang and Shanghai Jio Tong universities on the use of Midjourney in design processes and found a strong positive correlation between the actual use of Midjourney and perceived ease of use, perceived usefulness, attitude and behavioural intention. The results also suggest that Midjourney can be used in design education and has a positive impact on creativity. Hakimshafaei (2023) compared generative AI tools in the fields of architecture and design and found that Midjourney was able to recognise all words in the prompt and use them in the generated images according to DALLE-2 and Stable Diffusion. Midjourney can also generate new designs that follow the main idea of the prompt. In the DALLE-2 case, not all words in the prompt could be successfully incorporated into the design. In Stable Diffusion, he found that it did not include all words in the prompt and did not produce high-quality and detailed images. Lu (2023), in his study on drawing emotions using artificial intelligence tools, found that the outputs of Midjourney were more impressive than those of stable diffusion. King (2022), in his study to determine the perception of schizophrenia by journey, an artificial intelligence tool, found that the outputs obtained by using the word "schizophrenia" contain blood and fear expressions with unnatural facial features such as abnormally opened mouths with too many teeth and large red eyes. He argued that this does not match the schizophrenia patients he encountered in the hospital and that artificial intelligence has a structure open to prejudice. Abduljawad and Alsalmani (2022) used DALL-E 2, Stable Diffusion, and Midjourney AI tools to create datasets that are difficult to collect and found that DALL-E 2, Stable Diffusion, and Midjourney demonstrated the best performance, respectively. According to Turgay et al. (2023), Midjourney is used to increase designers' creativity, even though it has the limitation of selecting and intervening in a certain area using an interface. Nistler and Pojeta (2023) found that the outputs of journey, an artificial intelligence tool, for Jezerka stream restoration, Water tower, water animals, and the TGM WRI building contained successful images. Taecharungroj and Kompaniets (2023) used Midjourney to explore place experiences and AI-generated logos for Scandinavian city branding and found that its ability to adapt inputs such as place experiences produced successful outputs in terms of its ability to produce different materials that accurately summarise the unique characteristics of the city.

Graphic design is a perfect combination of aesthetics and communication. However, given the complexity and speed of the digital age, designers have focused on further developing this field by seeking new tools and methods. At this point, artificial intelligence tools have come into play and have expanded the boundaries of design through innovations that they bring to the world of graphic design. The impact of graphic design on user experience cannot be ignored. AI tools can analyse user habits, personalise content and optimise user-friendly interfaces. This helps designers create designs that are not only aesthetically pleasing but also in line with user needs. In the future, the combination of graphic design with AI is likely to further accelerate the transformation of the industry. AI tools continue to offer designers various possibilities to improve creative processes, enrich user experiences, and make the design world more accessible.

In conclusion, the use of AI tools in graphic design can make the design world more effective, efficient, and authentic. This not only offers designers the opportunity to accelerate and improve their work but also opens the door for them to explore new and innovative design approaches. The convergence of AI and graphic design has opened up a new and exciting dimension of creativity, transforming the world of design into the future.

Visual Reading

Visual reading is a type of reading that involves understanding and mentally constructing visuals (Göçer & Tabak, 2012) and includes reading, understanding, and interpreting visuals such as shapes, symbols, pictures, graphics, tables, body language, and natural and social events (Kaya, 2012; Baş & Örs, 2015). In addition, visual reading is the ability to read, understand, and interpret information supported by pictures and graphic images (Güven & Aktaş, 2013). Reading, understanding, and structuring shapes, symbols, pictures, graphics, tables, body language, natural events, and social events in the mind constitute visual reading (Özdemir & Erdem, 2011). At this point, visual reading helps the interlocutors in the formation of mental images, improve understanding of concepts, and solve comprehension problems (Onan, 2012). According to Stokes (2005), visual reading is the ability of the interlocutors to analyse what they see in a message. Büyüktopbaş and Uçar (2021) defined visual reading as the thought that emerges as a result of the connection established with the information in the mind after the message is obtained by visual senses and transferred to the mind. Within the framework of the definitions made and in the light of Sarıkaya's (2017) study, it is possible to define visual reading as a concept that includes the elements of shape, picture, line, rhythm-movement, texture, colour, diagrams and diagrams; it has many benefits in information exchange through its concise and short structures that improve thinking, understanding, learning, and integrating information.

Logo

A logo is a typographic whole containing the name of the institution and organisation to which it belongs (Bayırlı & Kılıç, 2022). Logos are the face of an institution or organisation (Demir & Tür, 2019); While it stands out with its simplicity, memorable, permanent, versatile, and suitable for the target audience, it is more than a business description; it is a flag, a signature (Sevildi, 2014). Logos that are recognised when they are seen, that cannot be defined in words, and that are based on visual perception are striking elements of visual identity, and the stronger the logos, the less they need a typeface (Karsak, 2009). City-specific logos are the first element that matches between the audience and the city in terms of defining the identity of the city and conveying visual impressions (Görgülü, 2019).

Artificial Intelligence

Artificial intelligence, which prioritises the effective use of factors such as precision, performance, and efficient use of time and cost by minimising the human element, is a software designed to create an intelligent behaviour model (Erdem et al., 2023). According to Huang et al. (2021), artificial intelligence has evolved into a unified structure with four sub-fields: computer vision, natural language processing, robotic process automation, and expert systems. In addition, according to Efe and Tunçbilek (2023), artificial intelligence encompasses several interconnected technologies, such as data mining, machine learning, speech analysis, image recognition, sensing, and sentiment analysis. According to Ba et al. (2022), advanced models such as DALL-E 2 and Mdjourney can produce realistic text-to-image images.

Models

Stable Diffusion

Launched in 2022, Stable Diffusion uses a deep learning technique called latent diffusion to create images based on text descriptions. Stable Diffusion employs CLIP to project a text prompt into a common text-image embedding space and select a coarse noisy image that is semantically close to the input prompt. The image is then subjected to a denoising method based on the latent diffusion model to produce the final image. In addition to a text prompt, the text-to-image rendering script in Stable Diffusion allows users to input various parameters, such as sampling type, image dimensions, and seed value (Dehouche and Dehouche, 2023).

Midjourney

Midjourney (Alawadhi and Yousef, 2023), a text-to-image AI tool developed by an independent research lab that allows users to create high-quality, well-structured, and detailed images based on textual descriptions and create a wide range of art forms from realistic to abstract styles, applies the principles of low floor, high ceiling, and wide walls. Users have a very low barrier to entry, as they can start getting the results they want with almost no training (low floor), with training the results scale up to commercial and award-winning results (high ceiling), and the results are extended to a wide range of applications (wide walls) (Vartiainen and Tedre, 2023).

DALL-E 2

Based on the unClip method developed by Ramesh et al., DALL-E 2 employs a single image encoder to transform text and images into a diffusion-based joint representation space (prior). Image generation is performed by a similarly trained decoder that translates the encoding of the prior data back into an image (Ploennings and Berger, 2023).

2. METHOD

The question posed in this study is "how deep can AI applications provide insights into logo design and how can these insights support graphic design?". Based on this question, this study addresses the challenges associated with graphic design and present potential AI tools as a means to achieve the following goals:

1. To explore state-of-the-art AI techniques and methods available to support creativity and ideation in the early stages of graphic design.
2. Evaluate the performance and effectiveness of the AI platforms journey, DALL-E 2, and Stabil Diffusion in the context of logo design to create different designs, styles, and future considerations.
3. Assess the challenges of AI text-image generators and provide practical guidelines and recommendations for designers to adopt AI technology and take advantage of its benefits.

2.1. Data

In this study, the municipal logos of 11 provinces in the Marmara region were included in the scope of the evaluation. These cities are Balıkesir, Bilecik, Bursa, Çanakkale, Çanakkale, Edirne, İstanbul, Kırklareli, Kocaeli, Sakarya, Tekirdağ and Yalova. The logos of these cities were analysed by visual reading and redrawn using the Midjourney, DALL-E 2, and Stable Diffusion artificial intelligence tools. Because the artificial intelligence tools are English artificial intelligence tools, the texts analysed by the visual reading method were translated into English, and the drawing was performed based on the English texts.

2.2. Visual Reading Outcomes

Balıkesir Metropolitan Municipality: While an Efe figure is used in the logo to draw attention to the cultural elements of the city, this figure also emphasises the bird paradise feature of the city. Again, the olive branch symbolises the peace and tranquillity of the city, and the agricultural role of the region is also emphasised in the logo.

Bilecik Municipality: The 16 gold-coloured 8-pointed stars in the Bilecik Municipality logo symbolise the 16 Turkish States in the history of the world. The image in the centre of the logo, consisting of two arrows and a bow, represents the Kayı Obası.

Bursa Metropolitan Municipality: The mountains on the logo represent Uludağ, a symbol of the city, while the minarets around it symbolise another symbol, the Ulu Cami. In addition, the dome in the middle of the logo symbolises the Green Tomb, while the sword and shield team and game, a cultural element, are visualised at the bottom. The stars around the logo symbolise the rising value and potential of the city.

Municipality of Çanakkale: The blue colour in the logo of Çanakkale Municipality represents the Marmara Sea, while the castle figure on the logo is separated in the form of a ceramic jug, symbolising the Dardanelles strait. The fish under the logo represent fishing, one of the city's main sources of livelihood.

Edirne Municipality: In the logo of Edirne Municipality, the figure represented by 4 minarets symbolises the Selimiye Mosque and the complex of the mosque, while the green colour used represents the nature of the city.

Istanbul Metropolitan Municipality: For Istanbul, which is defined as Seven Hills, 7 triangle point to each hill. Seven Hills, which is actually a kind of nickname for Istanbul, stems from the fact that the city is built on 7 different hills. The city walls in two separate parts symbolise the two sides of the city, while the blue colour used indicates that the city is a sea city.

Kırklareli Municipality: The logo features a bunch of grapes, as Kırklareli was a city of grapes in line with the word "Lozengrad", the name of the city during the Bulgarian period. Below the bunch of grapes lies the white-tailed eagle, which is endangered in Demirköy and represents wealth and power. Above the bunch of grapes is the Karagöz figure. The leaf on Karagöz represents both the grape leaf and the importance given to the environment.

Kocaeli Metropolitan Municipality: Three symbols representing the city are used in the logo. Among these symbols, the industrial wheel is used dominantly and effectively in the logo and emphasises the city's industrial characteristics. In addition, the Kocaeli clock tower and Atatürk monument are two other elements in the logo. Sakarya Metropolitan Municipality: The blue and green colours used in the logo, which consists of the Sakarya Bridge and 3 cogwheel symbols, represent the nature of the city.

Tekirdağ Metropolitan Municipality: In the logo consisting of the first letters of Tekirdağ Metropolitan Municipality, cherry is used because it symbolises Tekirdağ. The cherry is positioned with its leaves in the logo. While Tekirdağ's naturalness and the sea are symbolised with the colour blue, forests and agricultural culture are emphasised with the colour green.

Yalova Municipality: The image symbolised by the carnation on the Yalova Municipality logo expresses the city's active flower growing place, while the image just below it symbolises the richness of underground water and hot springs. While the waves express that Yalova is a sea city, the green and blue colours used symbolise the city's motto, which combines blue and green.

2.3. Limitations

This study was conducted on provincial municipalities in the Marmara region, and the scope be expanded to include regional and district municipalities.

3. DISCUSSION AND CONCLUSION

Official Logo	Midjourney AI Tool	Stable-diffusion AI Tool	DALL-E AI Tool
			
			
			
			


T.C.
EDİRNE BELEDİYESİ
Mussqe
Meduucaer Mriige
İSTANBUL
BÜYÜKŞEHİR
BELEDİYESİ
T.C. KIRKLARELİ 1870 BELEDİYESİ
Kirkide Kdara
KOCAELİ
BÜYÜKŞEHİR
BELEDİYESİ
ÇALIŞINCA OLUYOR
BYKIABAY CTAKLY


SAKARYA
BÜYÜKŞEHİR BELEDİYESİ
SYBIA SKGBA
T.C.
TEKİRDAĞ
BÜYÜKŞEHİR BELEDİYESİ
TEBBAB
TRIFRIGE
TERBI
T EKEKIAAD-I
YALOVA BELEDİYESİ
1927

When we look at the logo outputs created using artificial intelligence tools, it can be seen that many elements in the original logo are successfully used by all three artificial intelligence tools in the outputs produced for Balıkesir Metropolitan Municipality. The element symbolising the city's feature of being a bird sanctuary is visualised. In addition, while the olive branch in the original logo symbolises both the agricultural role of the city and its peaceful nature, this element is also emphasised in the generated logos. The Efe figure is not included in the new logo. It can be seen that Midjourney produced a very successful visuals project for Bilecik Municipality. Almost all elements in the input text were included in the outputs produced by the AI tool in question. For Stable Diffusion and DALL-E, it is not easy to say whether they produce successful output. Looking at the output produced for Bursa Metropolitan Municipality, it can be said that Midjourney produced a successful output in line with the text entered. Here too, local elements (the sword and shield team and the game) are not included in the logo. The outputs produced by Stable Diffusion and DALL-E are not as successful as journey. As an exceptional case, DALL-E created a logo using the input "green tomb" even though it was a cultural element. When the new logo produced for Çanakkale Municipality is analysed, it is seen that all inputs are used effectively in all three AI tools. Considering the placement and alignment of the elements, it is possible to conclude that the outputs produced by Midjourney are more successful than the others. In the original logo, the fish figure symbolising fishing, one of the important livelihoods of the city, is used under the logo. This usage is also effectively used in the Midjourney illustrations. Again, the ceramic pitcher figure and the use of sea elements are considered more successful in Midjourney drawings. The logos produced for Edirne Municipality represent the Selimiye Mosque and its complex, symbolising Edirne, and the colour green, representing the city. It is possible to say that the outputs of the three artificial intelligence tools successfully represent these features, whereas the output produced with Stable Diffusion creates the impression of a painting rather than a logo. When we look at the logos produced for Istanbul Metropolitan Municipality, it is seen that the elements in the text analysed by visual reading are more dominant in the logo produced by journey. The visual produced by Stabil Diffusion resembled a cartoon scene more than a logo, while the visual produced by DALL-E resembled a camera angle from a movie frame. In the logo produced for Kırklareli Municipality, elements such as grape clusters, white-tailed eagles, and grape leaves in the text given to the artificial intelligence tools were successfully used in all three tools. However, Karagöz's figure was not included in the three AI tools. It is assumed that this figure is excluded from the generated outputs because it is a local element. The industrial wheel in the logo of Kocaeli Municipality was used dominantly in the logos produced by the three artificial intelligence tools. In addition, the clock tower figure also found a place in the new logo. On the other hand, the Atatürk monument is excluded from the logos produced by the three artificial intelligence agencies. When we look at the outputs produced for Sakarya Metropolitan Municipality, it can be seen that the elements in the input text, which are analysed with visual reading, are effectively placed, and the colours are used as they are in the original logo. While the visuals are used together and in transition with the logic of the logo in Midjourney and the DALL-E artificial intelligence tools, they are symbolised in a disorganised structure in Stable Diffusion. Looking at the logos produced by the three artificial intelligence tools for Tekirdağ Metropolitan Municipality, it can be seen that the output produced by the Midjourney AI tool meets most of the elements in the text analysed by visual reading. While the logo design is more organised with the use of elements together, the elements are scattered in the stable diffusion. When we look at the output produced with DALL-E, it is seen that most of the elements in the input text are excluded. When we look at the visuals created for Yalova Municipality, it is seen that the carnation, which is one of the symbols of the city, and the richness of the underground waters find place in the logos. In addition, the colours blue and green, which also characterise the city, are also included in the logos. On the other hand, while the visual created with the Stable Diffusion application creates an impression that is far from a logo design, it can be said that the visual created with the DALL-E application resembles an oil painting. Midjourney's outputs, on the other hand, are in line with the logo form.

As a result, when we look at the designs created by artificial intelligence applications, it can be said that elements obtained by visual reading and given to artificial intelligence are included in the new logos. In some logos, elements reflecting more intense, general Turkish culture were successfully represented. In particular, general elements such as mosques, minarets, and the Turkish flag can be evaluated in this category. More localised cultural elements, such as Karagöz in the logo of Kırklareli Municipality were excluded from the logos. This is the same as the image representing the sword and shield game in the logo of Bursa Municipality. Here, the sword-shield figure, as a local cultural element, was not included in the new artificial intelligence drawing. In addition, logo designs can be said to carry different cultural influences depending on the textual content obtained from the visual reading method. While the dominant elements of Ottoman culture in the logo of Bilecik Municipality are at the forefront, the logo of Kocaeli Municipality has a more European design. Again, the colours used in the original designs of the logos were used effectively and efficiently in the logo redrawn by AI. It is noteworthy that artificial intelligence fails to add text to logos. This can be attributed to the fact that the artificial intelligence application works with the visualisation model.

Artificial intelligence offers several advantages and innovations in terms of graphic design. In this context;

Increased Speed and Efficiency: AI can increase speed and efficiency by automating graphic design processes. For example, algorithms that can handle large datasets perform repetitive tasks in the design process more quickly and effectively.

Personalisation and Segmentation: AI can create personalised graphic designs based on user behaviours and preferences. User segmentation and analysis can be used to customise designs for specific target audiences.

Style Transfer and Creativity: AI can transfer styles between different artistic styles and create creative designs. This gives designers the opportunity to bring different aesthetics together and create unique designs.

Colour Selection and Harmony: AI algorithms can help with colour choices based on colour theory and psychology. With colour harmony analysis, the visual impact of designs can be increased.

Visual Recognition and Content Analysis: AI can recognise objects and content in images via visual recognition and content analysis. This can help designers create meaningful graphics that align with the content.

User Experience and Interaction: By analysing user behaviours, AI can provide information to design better user experiences for websites and applications. Recommender systems can suggest personalised content and designs to users.

Education and Skills Development: Graphic designers should be trained to understand and utilise AI technologies. Aptitude towards AI-based design tools can increase designers' competitive advantage.

Human-Machine Collaboration: Artificial intelligence can help designers, but human creativity and aesthetic sense are important. Human-machine collaboration makes the most of the power of AI.

Ethics and Transparency: Ethics and transparency principles should be considered when using artificial intelligence. Transparency should be ensured in terms of the criteria on which designs are created and how personal data is used.

The use of AI in graphic design can provide designers with the opportunity to spend more time and energy focusing on creative areas. However, technology must be used correctly and ethically.

Peer Review: Externally peer-reviewed.

Conflict of Interest: The author has no conflict of interest to declare.

Grant Support: The author declared that this study has received no financial support.

ORCID ID of the author

Faruk Dursun 0000-0003-1571-1107

REFERENCES

Abduljawad, M., & Alsalmani, A. (2022). Towards Creating Exotic Remote Sensing Datasets using Image Generating AI. *International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, 84-88. https://doi.org/ 10.1109/ICECTA57148.2022.9990245

Alawadhi, A. M. & Yousef, D. M. (2023). Revolutionizing Plastic Furniture Design: Harnessing the Power of Artificial Intelligence. *Arab International Journal of Information Technology & Data*, 3(4), 7-46. https://doi.org/10.21608/AIJTID.2023.220627.1063

Al-Qatry, D., A. Q., & Rady, A. J. A. (2023). A comparative analytical study of the use of artificial intelligence (AI) tools in generating various designs for women's clothing. *International Design Journal*, 13(2), 363-380. https://doi.org/10.21608/IDJ.2023.288429

Ba, Z., Zhong, J., Lei, J., Cheng, P., Wang, Q., Qin, Z. Wang, Z., & Ren, K. (2022). SurrogatePrompt: Bypassing the Safety Filter of Text-To-Image Models via Substitution. https://doi.org/10.48550/arXiv.2309.14122

Bao, Y., & Xiang, C. (2023). Exploration of Conceptual Design Generation based on the Deep Learning Model- Discussing the Application of AI Generator to the Preliminary Architectural Design Process. *xArch symposium-Creativity in the age of digital reproduction*, 1-7.

Baş, B., & Örs, E. (2015). A Study on Visual Reading Skills of First Grade Students at Primary School. *Journal of Theory and Practice in Education*, 11(1), 225-244. https://doi.org/10.17244/eku.36093

Bayırlı, Ü., & Kılıç, E. C. (2022). Türkiye'deki Üniversitelerin Amblem ve Logolarının Şekil, Renk ve Bağlam Açısından İncelenmesi. *Sanat&Tasarım Dergisi*, 12(2), 425-446. https://doi.org/10.20488/sanattasarim.1221731

Borji, A. (2022). Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2. arXiv, Oct. 02, 2022. https://doi.org/ 10.48550/arXiv.2210.00586.

Büyüktopbaş, H., & Uçar, T. F. (2021). Görsel Okuma ve Tipografik Açılımları. *Sanat & Tasarım Dergisi*, 11(1), 158-175. https://doi.org/10.20488/sanattasarim.971055

Çeken, B., & Şen, O. (2023). Grafik Tasarım Sektöründe Yapay Zekanın Kullanılması (Midjourney). *1st International Conference on Contemporary Academic Research*, 138-145. https://doi.org/10.59287/iccar.767

Çelik, T. (2023). Generative design experiments with artificial intelligence: reinterpretation of shape grammar. *Open House International*, 30(1), 1-21. https://doi.org/10.1108/OHI-04-2023-0079

Chen, H. C., & Chen, Z. (2023). Using ChatGPT and Midjourney to Generate Chinese Landscape Painting of Tang Poem 'The Difficult Road to Shu'. *International Journal of Social Sciences and Artistic Innovations*, 2(3), 1-10. https://doi.org/10.35745/ijssai2023v03.02.0001

Chen, H. C., & Kao, Y. F. (2022). Mirroring Reality: Surreal Tourist Gaze of Chinese Landscape Painting by Artificial Neural Networks. *5th International Conference on Knowledge Innovation and Invention (ICKII)*, 155-159. https://doi.org/10.1109/ICKII55100.2022.9983554

Dehouche, N., & Dehouche, K. (2023). What's in a text-to-image prompt? The potential of stable diffusion in visual arts education. *Heliyon*, 9, 1-12. https://doi.org/10.1016/j.heliyon.2023.e16757

Demir, Ş. Ş., & Tür, E. (2019). Otel logosu ne söyle? Tüketiciler ne algılar? *Journal of Tourism Theory and Research*, 5(1), 51-64. https://dx.doi.org/10.24288/jttr.510437

Efe, A., & Tunçbilek, M. (2023). Yapay Zeka Algoritmaları ile Dönüşen Denetim Araçları Üzerine Bir Değerlendirme, *Denetişim*, 14(27), 72-102. https://doi.org/10.58348/denetisim.1195294

Erdem, G., Ermiş, Y., & Özkan, D. (2023). Yapay Zekâ ile Entegre Ultrasonografi Cihazı Kullanımının Periferik Sinir Bloğu Uygulamalarında Enjeksiyon Alanını Görüntülemeye Olan Etkisi. *JARSS*, 31(1), 21-27. https://doi.org/10.54875/jarss.2023.24381

Fahim, J. Y. (2023). Cross-Cultural Differences in Product Marketing. *IPHS 300 AI for Humanity*.

Göçer, A., & Tabak, G. (2012). Analysis of the 5th Grade Students' Turkish Workbooks in terms of Visual Reading Activities. *Elementary Education Online*, 11(3), 790-799. Retrieved from https://dergipark.org.tr/en/download/article-file/90568

Görgülü, Y. (2019). Kentlerin Markalaşması Sürecinde Logoların Kimlik İnşası: (Antalya Örneği) Göstergebilimsel Bir Çözümleme. *Elektronik Cumhuriyet İletişim Dergisi*, 1(3), 65-83. Retrieved from http://ecider.cumhuriyet.edu.tr/tr/pub/issue/51555/645037

Göring, S., Rao, R. R. R., Merten, R., & Raake, A. (2023a). Analysis of Appeal for Realistic AI-Generated Photos. *IEEE Access*, 11, 38999-39012. https://doi.org/10.1109/ACCESS.2023.3267968

Göring, S., Rao, R. R. R., Merten, R., & Raake, A. (2023b). Appeal and quality assessment for AI-generated images. *15th International Conference on Quality of Multimedia Experience (QoMEX)*,115-118. https://doi.org/10.1109/QOMEX58391.2023.10178486

Güven, M., & Aktaş, B. (2013). Eleştirel okuma ve görsel okuma arasındaki ilişki. *International Journal of Curriculum and Instructional Studies*, 3(6), 31-45. Retrieved from https://ijocis.com/index.php/ijocis/article/download/89/77

Hakimshafaei, M. (2023). Survey of Generative AI in Architecture and Design. University of California Santa Cruz Master Theisis. Retrieved from https://escholarship.org/uc/item/47x6k9j8

Hanna, D. M. (2022). The Use of Artificial Intelligence Art Generator "Midjourney" in Artistic and Advertising Creativity. *Journal of Design Sciences and Applied Arts*, 4(2), 42-58. https://doi.org/10.21608/jdsaa.2023.169144.1231

Hariffadzillah, T. S. N. T., Fadil, M. Z. M., Harun, A., Ayob, M. A., & Razak, M. R. A. (2023). Exploring Ai Technology To Create Children's Illustrated Storybook. *4th International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, 224-228. https://doi.org/10.1109/AIDAS60501.2023.10284665

Huang, A., Chao, Y., de la Mora Velasco, E., Bilgihan, A., & Wei, W. (2022). When Artificial Intelligence Meets the Hospitality and Tourism Industry: An Assessment Framework to Inform Theory and Management. *Journal of Hospitality and Tourism Insights*, 5(5), 1080-1100. https://dx.doi.org/10.1108/JHTI-01-2021-0021

Karsak, B. B. (2009). Logo Değişiminin Kurumsal Kimlik ve Kurumsal İmajla Bağlantısı: Unilever Örneği. *Marmara İletişim Dergisi*, 15, 113-120. Retrieved from https://dergipark.org.tr/tr/pub/maruid/issue/25255/266984

Kaya, M. F. (2012). The determination of the in-class applications of visual literacy and problems faced during these applications with regard to the views of Turkish classroom teachers. *Procedia Social and Behavioral Sciencesi*, 46, 2205-2209. https://doi.org/10.1016/j.sbspro.2012.05.455

King, M. (2022). Harmful biases in artificial intelligence. *Lancet Psychiatry*, 9(11), 48-48. https://doi.org/10.1016/S2215-0366(22)00312-1

Lu, R. T. E. (2023). Sentiments Re-Imaged by AI. New York University Tandon Scholl of Engineering Master Theisis. Retrieved from https://www.proquest.com/dissertations-theses/sentiments-re-imaged-ai/docview/2819976188/se-2

Nistler, J., & Pojeta, T. (2023). Graphical use of AI. *VTEI*, 4, 1-3. Retrieved from https://www.vtei.cz/en/2023/08/graphical-use-of-ai/

Onan, B. (2012). Türkçenin Ana Dili Olarak Öğretiminde Bilgi İşleme Süreci, *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 8(1), 96-113. https://doi.org/10.17860/efd.44365

Özdemir, N., & Erdem, İ. (2011). Türkçe Öğretmeni Adaylarının Yazma Alışkanlıklarının Bazı Değişkenler Açısından İncelenmesi. *Uşak Üniversitesi Sosyal Bilimler Dergisi*, 4(2), 101-125. https://doi.org/10.12780/UUSBD103

Ploennings, J., & Berger, M. (2023). AI art in architecture. Ploennigs and Berger AI in Civil Engineering, 2(8), 1-11. https://doi.org/10.1007/s43503-023-00018-y

Radhakrishnan, M. (2023). Is Midjourney-Ai the New Anti-Hero of Architectural Imagery & Creativity?, *Global Scientificy Journals*, 11(1), 94-104. Retrieved from https://www.globalscientificjournal.com/journal_volume11_issue1_January_2023_edition.html

Rozdolska, A. J. (2022). Artificial intelligence as part of future practices in the architect's work: MidJourney generative tool as part of a process of creating an architectural form. *Architectus*, 3(71), 95-104. https://doi.org/10.37190/arc220310

Sánchez, C. D., & Heras, D. C. (2023). Use of Open Access AI in teaching classical antiquity. A methodological proposal. *The Journal of Classics Teaching*, 1-5. https://doi.org/10.1017/S2058631023000739

Sarıkaya, B. (2017). Türkçe Öğretiminde Görsel Okuma. *Muş Alparslan Üniversitesi Sosyal Bilimler Dergisi*, 5(3), 779-796. https://doi.org/10.18506/anemon.298912

Sevildi, İ. (2014). *Tipografi ve Logo Tasarımındaki Önemi*. [Yayımlanmamış yüksek lisans tezi]. İstanbul Arel Üniversitesi. https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp

Shuhan, S., Yeutong, C., & Mai, Y. (2023). Measuring designers use of Midjourney on the technology acceptance model. *International Association of Societies of Design Research Congress*, 1-8. https://doi.org/10.21606/iasdr.2023.794

Stokes, S. (2005). Visual Literacy in Teaching and Learning: A Literature Perspective. *Electronic Journal for the Integration of Technology in Education*, 1(1),10-18. Retrieved from https://www.researchgate.net/publication/252798531_Visual_Literacy_in_Teaching_and_Learning_A_Literature_Perspective

Taecharungroj, V., & Kompaniets, R. O. (2023). Exploring place experiences and AI-generated logos for Nordic city branding. 7th Annual International Place Branding Association (IPBA) Conference, 62-62.

Tanughara, S. (2023). Review Using Artificial Intelligence-Generating Images: Exploring Material Ideas from MidJourney to Improve Vernacular Designs, *Journal of Artificial Intelligence in Architecture*, 2(1), 48-57. https://doi.org/10.24002/jarina.v2i2.7537

Tsidylo, I. M., & Sendra, C. E. (2023). Artificial Intelligence as a Methodological Innovation in the Training Of Future Designers: Midjourney Tools. *Information Technologies and Learning Tools*, 97(5), 203-214. https://doi.org/10.33407/itlt.v97i5.5338

Turgay, O., Kaptan, M., Erkartal, P. Ö., Demir, T., & Doğukanlı, Ç. (2023). Interdisciplinary Trajectories: Regenerating Barcelona Pavilion With Midjourney Günçe, K. (ed). *Interior Architectural Issues Design, Theory & Philosophy*, (pp.299-320).Livre de Lyon.

Vartiainen, H., & Tedre, M. (2023). Using artificial intelligence in craft education: crafting withtext-to-image generative models. *Digital Creativity*, 34(1), 1-21. https://doi.org/10.1080/14626268.2023.2174557

Wasielewski, A. (2023). "Midjourney Can't Count": Questions of Representation and Meaning for Text-to-Image Generators. *The Interdisciplinary Journal of Image Sciences*, 37(1), 71-82. https://doi.org/10.1453/1614-0885-1-2023-15454

How cite this article

Dursun, F. (2024). Redesigning municipality logos in the context of visual reading using artificial ıntelligence. *Acta Infologica, 8*(2), 88-99. https://doi.org/10.26650/acin.1451569

İSTANBUL UNIVERSITY PRESS

Acta Infologica
acin 2024, 8 (2), 100–117
DOI: 10.26650/acin.1463879

Research Article

Detection and Prevention of Medical Fraud using Machine Learning

Ceyda ÜNAL[1], Gökçe Sinem ERBUĞA[2]

[1]Dokuz Eylül University, Faculty of Economics and Administrative Sciences, Department of Management Information Systems, İzmir, Türkiye
[2]Dokuz Eylül University, Faculty of Economics and Administrative Sciences, Department of Business Administration, İzmir, Türkiye

Corresponding author : Gökçe Sinem ERBUĞA
E-mail : gokce.erbuga@deu.edu.tr, gokceerbuga@gmail.com

* This study is a revised and expanded version of the paper presented at the 22nd International Business Congress organized by Nişantaşı University on 07-09 September 2023.

ABSTRACT

Presently, there is an upward trend in the mean life expectancy of individuals due to reductions in maternal and infant mortality, as well as deaths caused by non-communicable diseases like cardiovascular disease. A decline in life expectancy results in a corresponding increase in health expenditures sustained by both public and private entities, including insurance providers. The healthcare sector has become an extremely comprehensive and critical industry due to the following factors: the increase in healthcare expenditures, particularly during the pandemic; the cost of each component in the healthcare sector; the increasingly chaotic healthcare technology ecosystem; the growing expectations of numerous and diverse stakeholders; and the presence of numerous and new actors in the sector. Nevertheless, this circumstance exposes the health sector to many hazards, thereby increasing its susceptibility to fraudulent activities. The sector's substantial volume will inevitably lead to expensive fraudulent activities. For this reason, prospective medical frauds should be prevented and detected immediately. Machine learning is considered one of the most powerful and optimal approaches to prevent medical fraud. An example application is used to assess the efficacy of machine learning in the medical fraud detection context as part of the research. The objective of the proposed application is to classify provider-side medical fraud by applying various machine learning techniques and medical claims.

Keywords: Machine Learning, Artificial Intelligence, Healthcare Sector, Medical Fraud.

Submitted	: 02.04.2024
Revision Requested	: 17.07.2024
Last Revision Received	: 21.08.2024
Accepted	: 08.09.2024
Published Online	: 16.09.2024

1. INTRODUCTION

Noncommunicable diseases (NCDs) cause the highest disease burden worldwide, and their effects have worsened over the last 20 years. NCDs, which are a significant threat to people from all geographical backgrounds and age groups, affect elderly individuals and genetically and psychologically disadvantaged individuals to a greater extent. In addition, such diseases can cause greater losses in countries where lower-income groups are the majority, as stated by the World Health Organization.

Non-communicable diseases (NCDs) have grown to be a contributory factor in more than 40 million deaths worldwide, representing an increase of more than 30% since 2000. In other words, approximately 75% of deaths worldwide are caused by non-communicable diseases (NCDs) (WHO, 2023a). Although these diseases are quite prevalent, the lengthy and expensive procedures required to treat them force nations to set aside enormous sums of money for medical expenses. Therefore, countries must allocate huge budgets for health expenditures.

Considering 2020 data, the country that allocated the highest budget for health expenditures in the world was the United States, which constituted approximately 16.8% of the US GDP. The United States is followed by England (12.8%), Germany (12.5%), France (12.4%), and Canada (11.6%). Considering OECD countries, the ratio of health expenditures to GDP is determined as 8.8%, while in Türkiye this rate is equal to 4.7% of GDP (Euronews, 2022).

When evaluated on a global basis, the importance of health expenditures for all countries increases because of high health expenditures and rigid supply elasticity in health services. For this reason, any act of corruption or fraud in the process of realizing health expenditures will cause the costs that both the state, private sector organizations, and individuals have to bear to increase many times.

The healthcare sector is further made more complex by the high level of uncertainty, the significant number of players involved, the presence of asymmetric knowledge among these actors, and numerous other considerations. This complex structure makes the sector more open to corruption and fraud and paves the way for more illegal tasks to be undertaken in the industry (Avcı & Teyyare, 2012:199). Making the best use of the resources already available is crucial for ensuring the greatest level of effectiveness from health facilities offered in the health sector, where expenses are fairly high and unavoidable (tight flexibility) (Aydın & Yenimahalleli Yaşar, 2020:64). In this way, more people will be able to access health care services, and these people will be able to receive them inexpensively and efficiently. As a result, it is crucial to identify and prevent corruption and fraud that will disrupt the balance of optimal resource use in healthcare.

In this regard, being able to detect and prevent fraud that has occurred or will occur in the healthcare sector is an indispensable part of healthcare management and medical accounting. However, the large number of participants in the health sector and the existence of a large amount of data make it challenging to analyze and interpret such data using classical data analytics methods. In addition, it is not possible to store the obtained data and preserve them or use them effectively for diagnostic purposes during the treatment process using traditional methods. For this reason, developments in information and communication technologies are frequently used to collect, store, analyze, and interpret existing large volumes of data (Kurşun, 2021:921; Altındiş & Kıran Morkoç, 2018:257). Likewise, technology has begun to be frequently used to detect medical frauds occurring in the healthcare sector. One of the methods used to prevent medical fraud is to detect fraud using machine learning. With machine learning, medical fraud can be prevented more effectively and at lower cost, and data in healthcare services can be interpreted more effectively and efficiently.

Within the scope of the study conducted in this regard, the concept of medical fraud was first explained in detail. In the following section, the detection of medical fraud is examined, followed by research on the topic published in the literature. After discussing how and in what way machine learning is used to detect medical fraud, the practical application of machine learning to detecting medical fraud is examined in an application case. Finally, the inferences drawn from the application were discussed, and necessary suggestions were made in the Conclusion.

2. CONCEPTUAL FRAMEWORK

Spending on healthcare is rising significantly because of population longevity and intensified access to healthcare globally. Due to the nature of the health sector, resources allocated to this sector have a large share in global and domestic economies. For example, healthcare expenditures in the United States reached 3.65 trillion USD or 11,172 USD per capita in 2018 (Ekin et al., 2021:1183). The uncertainties and risks in this sector are excessive, and there are many actors and imbalances between their knowledge about the sector (Avcı & Teyyare, 2012:199). This situation makes supervision of the health sector quite difficult. Therefore, the sector offers a more suitable playground for corruption than many other sectors.

Medical fraud creates a huge burden on both the health systems and economies of countries. Although an exact

amount cannot be determined, expenses due to healthcare fraud, forgery, and abuse in the medical sector can reach one-tenth of all healthcare expenditures. This rate is approximately $100 billion in US health expenditures, which exceed $1 trillion each year (US Department of Justice, 2024). In addition to the financial burden it imposes, fraud increases the cost of healthcare services and causes a huge decrease in their quality.

Because of medical fraud cases, $500 billion the $7.5 trillion annually spent on health care worldwide is wasted because of systemic corruption. According to another source, the global health system has reduced 6.2% of its average annual revenue because of corruption and fraudulent activities (Bozhenko, 2022:32). Although losses incurred because of medical fraud activities averaged $1,297,560, 1-5% of these activities resulted in losses of $150,000 or less, according to the report of the US Sentencing Commission (USSC). The average loss caused by criminal activities was $9,500,000 (USSC, 2022).

To evaluate the situation within European borders, the damage caused by fraud and corruption cases in the healthcare sector is more than €50 billion. As shown in the report titled "The Financial Cost of Healthcare Fraud" prepared by Gee, Button, and Brooks from the Center for Counter Fraud Studies at the University of Portsmouth, EU countries spend more than €1 trillion on the health sector each year. €56 billions of this expenditure is wasted each year as losses due to corruption. The waste generated in the health sector worldwide can reach up to €180 billion.

This amount corresponds to more than 5% of the budget allocated by these countries to health expenditures (Vincke & Cylus, 2011:14). On the other hand, the annual amount of health funds withheld from the use of the health system each year is equal to the amount required to achieve Universal Health Coverage (UHC), the core commitment of the UN's Sustainable Development Goals, which aim to provide affordable, accessible, and quality health care for all (Transparency International, 2023).

The National Health Care Anti-Fraud Association (NHCAA) stated that economic losses from medical fraud are in the tens of billions of USD each year. Another optimistic estimate by the NHCAA states that the cost of medical fraud is 3% of the total amount of money spent on medical care in the US. However, the cost of medical fraud can reach up to 10% of annual healthcare expenditures, according to some government and law enforcement agencies, representing approximately more than 300 billion USD (NHCAA, 2021). In addition, it is estimated by Transparency International (2023) that fraud within the field of medical care causes an estimated 140,000 pediatric fatalities each year.

When we look at China, one of the largest economies in the world, it is seen that it ranks 5th in the health and health systems ranking, and the economic value of the health sector exceeds 2 trillion yuan. In this sector, which is a huge market, China's National Health Care Security Administration (NHSA) randomly inspected approximately 200 000 medical institutions in 2018. Because of the inspections conducted, approximately 1/3 of these institutions committed medical fraud (Zhang et al., 2020:2). According to the 2020 data of the National Healthcare Security Administration (NHSA), approximately 10% of health expenditures in China were wasted with these abuses.

Medical fraud affects all healthcare professionals, institutions, local and national governments, development banks, aid organizations, etc. actively operating in the healthcare sector. It directly affects many actors. To achieve the goal of "ensuring healthy lives and promote well-being for all at all ages" (SDG-3), one of the Sustainable Development Goals set by the United Nations, it is imperative for all nations and health institutions to devise strategies that tackle corruption (Vian, 2020:114).

Healthcare organizations are exceptionally susceptible to medical fraud. This is due to the following factors: unpredictability regarding service demand (including who will become unwell and when, as well as what they will require), the complex interactions of numerous different scattered parties, such as payers, suppliers, customers, and regulators, asymmetric information between different actors, and difficulty in identifying and controlling different interests, etc. (Vian, 2007:84). It is extremely important to determine to what extent private providers should be entrusted with important public roles in the provision of health (medical) services. In addition, as Savedoff and Kussman (2006) declare, the large public budget allocated for medical expenses in many countries requires that medical expenses be made transparently and meticulously.

Medical fraud observed in healthcare companies finds its place in the literature as a particular form of white-collar crime used to express the dishonest provision of medical services for the purpose of economic gain, that is, profit (Ogunbanjo et al., 2014:10). White-collar crimes, which include acts of "deception, concealment, or abuse of trust" and are not related to actions involving physical force or violence/threats (FBI, 1989:3), are committed by individuals or organizations established for the purpose of committing crimes with the motivation of individuals or economic gain.

The European Healthcare Fraud & Corruption Network (EHFCN) defines fraudulent acts carried out in the healthcare sector as "obtaining any gain/benefit improperly by deliberately violating a certain rule". Corruption is explained as "the abuse of authority by involving a third party in a criminal act and, as a result, obtaining a certain gain illegally" (Küçük, 2022: 588).

Patients/taxpayers, suppliers (companies from which the government purchases drugs and equipment), regulators (ministry of health, pharmaceutical regulatory agency), and suppliers (medical facilities such as hospitals, health clinics, and individual or group physician practices) are the five key actors in health systems as identified by the Organization for Economic Co-operation and Development (OECD) (Couffinhal & Frankowski, 2017). Crimes resulting from breaches of integrity, such as fraud and corruption, manifest due to engagement among these diverse stakeholders (Vian, 2020:116).

The planning of medical fraud crime offenses can vary in complexity. These offenses can also be committed by patients, healthcare professionals, or other entities who purposefully mislead the healthcare system to obtain fraudulent benefits or compensation (Thomson Reuters, 2021). The people who commit the crime of medical fraud include doctors, nurses, pharmaceutical companies, pharmacists, health technicians, medical officers, physiotherapists, and other healthcare professionals. In addition, individuals who demand healthcare (patients, consumers), healthcare companies, insurance companies, or actors who play an intermediary role in service delivery (medical service and equipment suppliers) can also be perpetrators of medical fraud (Price and Norris, 2009: 286). These people perform different types of medical criminal acts in illegal and unethical ways.

Medical fraud is frequently performed in various ways, such as upcoding (charging for a more expensive diagnosis or procedure), providing needless treatments or screenings, paying for healthcare services that were never rendered, unbundling, different criminal acts include falsifying the seriousness of a medical disease, exaggerating, manipulating, and paying illegal kickbacks in exchange for preferential treatment (Vian, 2007). There are many types of fraud and corruption in the healthcare industry. Even though the literature contains an extensive amount of research that subjected these criminal acts to different classifications, it is not possible to discuss a common typological classification. In the report titled "Corruption in the Healthcare Sector" prepared by the European Commission (EC), acts of fraud that may occur in the healthcare sector are grouped under the following six headings:

- Bribery in providing medical services, and
- Medical equipment supply fraud,
- Improper marketing relationships,
- Abuse of (high-level) professional positions,
- Unnecessary refund requests,
- Abuses and corruption related to medicines and medical devices (EC, 2017:9).

Transparency International, on the other hand, classifies fraud in the health sector into five categories:

- Embezzlement, and theft from the health budget or user fees,
- Corrupt procurement practices,
- Corrupt practices in payment systems,
- Corrupt practices in the pharmaceutical supply chain and
- Corrupt practices in the provision and delivery of health services (Transparency International, 2006:18).

Based on the known identity of the offender, Küçük (2022) classified fraud and corruption offenses in the healthcare industry into three categories: actions committed by patients, actions carried out by healthcare providers, and actions committed by patient or supplier actors, in other words, by third parties. A similar classification was observed in Ekin et al. (2018). The study by Ekin et al. in question classifies the concept of medical fraud through the perpetrators of the action and discusses it in three basic categories: crimes committed by the service provider (hospital, doctor, etc.), crimes committed by the consumer (patient), and crimes committed by insurers.

It is extremely difficult to identify frauds that occur with the help of globalization in the health sector and developments in information technologies using traditional auditing methods. It is possible to say that new crime types are emerging, and current crime detection methods are insufficient to counter fraud. Therefore, to carry out an effective fraud detection process, more advanced and comprehensive crime prevention systems that incorporate different statistical approaches are required (Li et al., 2008:275).

In order to identify the abuses encountered in the healthcare sector, the existence of a beneficial audit and control system and its effective functioning must be ensured. In the process where the rapid transformation in information technologies has not yet clearly demonstrated their effects, audits and controls are performed by audit personnel, such as auditors, controllers, and inspectors, and require longer time and more effort. This classical audit and control process, in other words, a process performed manually, mostly involves the examination of physical documents and files and long-term field work (Turğay et al., 2020:5).

Traditional healthcare fraud detection methods, which are often limited to efforts to detect fraud rather than prevent

it, have not yet been sufficiently efficient and effective. Detecting receivables before payment has been shown by the International Social Security Association (ISSA) as a more effective way to prevent fraud and corruption in the healthcare sector. Therefore, the paradigm of inappropriate healthcare expenditure management is shifting from surveillance management to prevention (ISSA, 2022).

The widespread use of digitalization and the resulting social transformation in society have led to the need for audit activities on electronic platforms. In addition, the large volume of data used in audits and accumulated over the years makes it worthwhile for audit staff to examine these data. These difficulties encountered during the audit process reveal that audit activities should not be satisfied by classical methods and that data should be examined in a more reliable way using new analysis and modeling techniques. In this context, in line with current developments, the concept of continuous auditing, which refers to "a form of auditing that enables the auditing of real-time accounting information (data) without the need for physical documents and is carried out by using a number of special computer-based audit programs", has emerged (Orhan, 2015: 85).

Because medical fraud causes losses between 3% and 10% of countries' health expenditures (Shin et al., 2012:7441), it is extremely important to audit these transactions carefully. The intensity and complexity of criminal allegations that may lead to medical fraud and corruption necessitate frequent use of information technology and data analysis methods in audit processes (Ekin, 2019:4107; Turgay et al., 2020:6). In addition, data analysis and modeling techniques have gained importance in this field, resulting in the emergence of various data analysis techniques.

Although classical (traditional) methods, such as linear discriminant analysis or logistic regression analysis, which are used to reveal fraud and corruption, are effective in solving many cases, there are more powerful and effective analysis methods, such as artificial neural networks (Bolton and Hand, 2002:237). He, Wang, Graco, and Hawkins (1997) and He, Graco, and Yao (1999) used neural networks, genetic algorithms, and nearest neighbor methods in their studies. In their study, Major and Riedinger (1992) used statistical information-based medical fraud detection methods that compared observations with those that were most similar. Different methods, such as artificial intelligence, machine learning, distributed and parallel computing, econometrics, expert systems, fuzzy logic, outlier detection, pattern recognition, and visualization, have also been used in the literature (Bolton and Hand, 2002:245).

By means of data-driven inventiveness, fraud and corruption detection and prevention technological advances, including in the cognitive field of computing, data mining, analytics, machine learning, and various other kinds of artificial intelligence (AI), have greatly advanced. Some of these methods are (ISSA, 2022):

- Biometric recognition using a fingerprint scanner, iris recognition, or facial recognition
- Predictive modeling methods, such as data mining, predictive analytics, and quantitative analysis techniques, to detect patterns in supplier fraud and behavior
- Artificial intelligence-based pattern recognition techniques used to identify coding and billing errors
- Blockchain applications make it impossible for fraudulent practices to delete or modify data and allow detailed asset tracking.

The Centers for Medicare and Medicaid Services (CMS), an American health insurance provider, publishes the healthcare data used by most researchers to detect healthcare fraud. Raw data for detecting fraud in healthcare businesses often come from insurance claims, including government healthcare data, physician data, clinical data, and private insurance company data.

As the rate of digital transformation in the healthcare industry accelerates, healthcare organizations' digitalization processes have begun to take on a completely novel form. In healthcare organizations, the emergence of electronic health data of various sizes and types has created new opportunities for automated fraud detection. Specifically, in the context of automated processes, machine learning and data mining techniques are crucial for detecting abuse in such data (Joudaki et al., 2015). Currently, machine learning techniques are regarded as the most essential components for identifying inappropriate use and fraud. Data mining and machine learning, as a collective, encompass methodologies that leverage artificial intelligence, statistics, and mathematics to extract and discover valuable insights from databases (Aydoğan Duman & Sagiroglu, 2017). According to Alpaydın (2020), machine learning refers to the capacity of computer algorithms to gain decision-making skill using data and statistical theory when constructing models. In essence, there are three classifications for machine learning methodologies: supervised, unsupervised, and semi-supervised learning.

Supervised learning methods are employed to identify fraudulent or dishonesty activities in healthcare organizations. These methods utilize data samples that have been previously identified as fraudulent or non-fraudulent or labeled as a result. Models developed using these data are crucial for automated identification of previously identified fraud and abuse patterns. Methods of supervised learning employ a variety of techniques, including classification and regression algorithms.

Support Vector Machines (Francis, Pepper, & Strong, 2011; Kirlidog & Asuk, 2012), Neural Networks (Liou, Tang, & Chen, 2008), and decision trees (Branting et al., 2016) are examples of supervised machine learning techniques used to classify fraud detection in healthcare organizations. In addition, regression analysis methods are also used in fraud detection (Francis, Pepper, & Strong, 2011). At the same time, some studies have utilized naive Bayes and decision trees for big data analysis on the Hadoop platform (Dora & Sekheran, 2015). In this context, the most critical drawback of supervised machine learning methods is the need for human input and the required output, i.e., labeled data. In particular, for fraud detection, the acquisition and interpretation of labeled data are laborious and time-consuming (Saravanan and Sujatha, 2018). Unsupervised learning techniques have been suggested to address these disadvantages of supervised learning.

Unsupervised learning approaches identify fraudulent activities within an unannotated dataset, operating under the general assumption that a substantial proportion of the data comprises legitimate activities (Abdallah, Maarof, & Zainal, 2016). In contrast to supervised learning, the proposed model is constructed without the use of labeled data. One of the primary benefits of unsupervised learning is that it enables the precise detection of fraudulent activity, even in the absence or presence of inadequate labeled data (Bolton and Hand, 2001). Basic unsupervised learning methods include clustering, association rules, and outlier detection. Different machine learning methods such as association rule analysis (Shan et al., 2008), k-means (Shan et al., 2009), probabilistic programing (Bauder & Khoshgoftaar, 2016), are used in fraud detection in healthcare.

The benefits of both supervised and unsupervised learning are combined in semi-supervised learning. It can be considered a hybrid method that uses the features of supervised and unsupervised learning to achieve more accurate results. When there is a relatively small proportion of labeled data with a large amount of unlabeled data, semi-supervised machine learning techniques are frequently used. Building models that consider both labeled and unlabeled input is, in essence, the primary objective of semi-supervised learning (Zu, Wang, & Wu, 2011). With the assistance of domain experts, semi-supervised learning is also used in unsupervised learning techniques such as clustering and outlier detection (van Capelleveen et al., 2016).

3. LITERATURE REVIEW

Acts of fraud and corruption in the healthcare sector deepen social inequality, and in this context, poor people and disadvantaged groups are affected the most. In addition, fraud and corruption prevent the fight against important diseases because of the diversion of resources and funds for financing health services through different channels. For this reason, fighting corruption and fraud in healthcare is extremely important across the world to supply outstanding medical care and effectively address both current and potential risks to global healthcare by making healthcare accessible to all (Transparency International, 2023).

The U.S. The Department of Justice reported in 2023 that a hospice medical director was sentenced to four years and two months in prison. This sentence was imposed because the director submitted over $150 million in false and fraudulent Medicare claims for hospice and other medical services. According to court filings, the healthcare organization's initiatives enroll people with incurable diseases, including dementia and Alzheimer's disease, and those with low mental capacity in retirement, nursing, and public housing. Patients were informed that their prognosis was less than six months by the corporation. Over $18 million in unnecessary services were approved (U.S. Department of Justice, 2023).

The Georgetown University Memory Disorders Program's recent paper claims that medical fraud has been moved to academic research. Recent Alzheimer's disease (AD) research fraud raises serious issues. Several figures in frequently cited 2006 Nature research on animal models of AD may have been modified, leading to problematic results (GU University Memory Disorders Program, 2023). People are prosecuted for Alzheimer's misdiagnosis.

Because of the problems it creates in the pharmaceutical supply chain, medical fraud also causes disruptions to individual treatment processes. Corruption also results in the waste of existing economic and human resources by limiting countries' capacity to manage national and global health risks, according to Transparency International (2023).

Fraud and corruption crimes committed in the healthcare sector, in addition to their huge financial losses, also threaten the quality and safe delivery of medical services that the healthcare system can offer to individuals (Li et al., 2008:275). In this regard, acts of medical fraud and corruption should be detected as soon as possible and even prevented before they are revealed to enhance the standards of the amenities provided and concurrently cut back on service expenses.

The World Health Organization (WHO) attach great importance to the anti-fraud and anti-corruption, transparency, and accountability (ACTA) steps to be taken in this direction. Within the framework of UN Sustainable Development

Goal (SDGs) 16.5, reducing and even preventing corruption and bribery is necessary to improve health services, to prevent inequalities in health services, and to improve lives (WHO, 2023b).

In conformity with the 2020 report by the Organization for Economic Co-operation and Development (OECD), over 45% of the global population holds the belief that the healthcare industry is highly corrupt. From this vantage point, one of the primary concerns regarding steady growth for both highly wealthy countries and developing nations globally is the fight against fraud and corruption in the health sector.

The literature highlights the US Health Care Financing Administration (HCFA) among prominent government health departments. A pair of healthcare programs known as "Medicare" and "Medicaid" exist in the United States. Medicare is a government-administered social insurance plan for those with end-stage renal illness, as well as those aged 65 years or above, or those with specific disabilities who are younger than 65 years. This program offers prescription drug coverage, hospital insurance, and health insurance. Medicaid is administered by individual states, with each state establishing its own eligibility and service requirements. Medicaid is restricted to low-income families and individuals who satisfy the eligibility requirements established by state and federal legislation (Liu & Vasarhelyi, 2013). Several data types are subjected to distinct analyses because health-system data vary by country.

Liou et al. (2008) examined claims submitted to the Taiwan National Health Insurance for outpatient services for diabetes using supervised methods. The average drug cost, average diagnostic fee, average amount claimed, average dispensing days, average medical expenditure per day, average counseling and treatment fees, average drug cost per day, average dispensing service fees, and average drug cost per day were the expenditure-related characteristics that they compared two groups of fraudulent and non-fraudulent (with/without fraud) claims to develop detection models. To detect fraud, three machine learning techniques—Classification Trees, Logistic Regression, and Neural Networks—were compared. All three approaches were successful in achieving accuracy although the Classification Tree model performed better overall, obtaining a 99% correct recognition rate.

Lin et al. (2008) used clustering techniques to analyze data about medical practitioners covered by the Taiwan National Health Insurance. A total of 10 attributes were used to categorize the physician data. The critical clusters were identified and ranked according to expert opinion regarding the influence of the clusters on health expenditures.

Aral et al. (2012) used a drug's commercial name, market price, prescription number, prescriber age, gender, and indication to detect prescription forgery. The model identified fraudulent prescriptions with a true positive rate of 77.4% and false positive rate of 6% in an adult cardiac surgery database. Similarly, Shin et al. (2012) examined 38 claims criteria to identify fraudulent claims in 3,705 outpatient internal medicine clinics. Using these features, a risk score was calculated, and a decision tree model was used to classify the providers. These studies demonstrate how prescription-specific data and broader claim characteristics can be used to detect and prevent medical fraud.

A model for detecting prescription forgeries was proposed by Aral et al. (2012). Six features were identified for this purpose: the prescribed drug's commercial name, the prescribed drug's market price, prescription number, the prescriber's age, gender, and the prescribed drug's indication. When evaluating the effectiveness of the model in identifying fraudulent medical prescriptions (with a true positive rate of 77.4% and a false positive rate of 6%) in the adult cardiac surgery database, the model demonstrated satisfactory performance in distinguishing between fraudulent and non-fraudulent prescriptions.

Shin et al. (2012) attempted to identify fraudulent claims in 3,705 outpatient internal medicine clinics. A total of 38 characteristics were identified from outpatient claims submitted to a health insurance organization as part of the research. Based on these attributes, a risk score was computed to represent the probability of fraudulent activity. Providers are classified using a decision tree model.

Srinivasan et al. (2013) used rule-based data mining to detect health insurance claim fraud, misuse, waste, and errors using Medicare data. This big data technique helps private health insurers find hidden cost overruns that standard information systems miss. In addition, Branting et al. (2016) developed a decision tree and graph analysis method to automatically detect and predict fraud. Through anomaly detection and predictive analysis, these methods provide a comprehensive framework to detect and prevent health insurance claim fraud.

A machine learning model was proposed by Bauder et al. (2016) to identify anomalous physician conduct in health insurance claims. The model attempts to identify instances in which physicians deviate from the established standards of their specialty, thereby notifying decision makers of billing procedure abuse or fraud. Through the use of five-fold cross-validation, the model sensitivity, specificity, and F1 score were computed. By employing the Naive Bayes algorithm, the model accurately forecasts multiple classifications of physicians with F1 scores exceeding 0.90.

Seven essential stages were outlined by Joudaki et al. (2015) as a framework for examining healthcare and insurance claims to detect fraud and abuse (following the data pre-processing phase). 1) Domain experts (Sokol et al., 2001; Li et al.) identify the most significant characteristics of the data; 2) Automated algorithms or expert opinion, such as

association analysis, identify new features that are indicative of fraudulent or malicious behavior (Li et al.); 3) Outlier detection methods identify unusual records for further investigation (Shan et al., 2009). 4) Process of extracting outliers from the dataset and clustering or clustering records according to the extracted features (Lin et al., 5) Determining outlier clusters and conducting additional analysis on the records contained within these clusters to detect fraudulent or malicious records 6) Constructing supervised models by selecting the most discriminatory features from the labeled records from the previous step (Liou et al., 2008). 7) Implementing supervised methods during regular online processing tasks and unsupervised methods (such as outlier detection and clustering) during designated time intervals in order to enhance preceding procedures and identify novel instances of fraudulent activity.

4. METHODOLOGY

4.1. Dataset[1]

The data used in this study were obtained from the "*Healthcare Provider Fraud Detection Analysis*" dataset on the Kaggle platform. This dataset is also a combination of the four datasets. The dataset contains retrospective data. These datasets are:

- Inpatient Data: This data contains information about the hospital claims submitted on behalf of inpatients. In addition, the admission and discharge dates, and the admission diagnostic code, are included.
- Outpatient Data: This data includes claim details for patients who visited the hospitals but were not admitted.
- Beneficiary Details Data: This data consists of information about the beneficiaries, including their ages, health conditions, region, and more.
- Target Class (PotentialFraud): This class includes fraud classes. (Yes/No).

This dataset was acquired from the Kaggle platform's "Healthcare Fraud Detection Analysis" dataset, is also composed of four datasets. The final dataset was acquired via data preprocessing and feature engineering. The final dataset and its descriptions are presented in Table 1.

4.2. Data Preprocessing/Feature Engineering

In the first version of the dataset, before conducting a comprehensive analysis of the dataset and prior to conducting the exploratory data analysis phases, 4904 non-fraudulent and 506 fraudulent activities were observed (Table 1). The "BeneID" feature, which is shared by all four datasets, was used to merge the datasets. This feature comprises individualized patient identities. There were 212796 fraudulent records (fraudulent-Yes) and 345415 normal records (non-fraudulent-No) in the final stage. During the phase of exploratory data analysis, "pandas-profiling" module of the Python library pandas was implemented. By using the extremely high-level application programing interface (API) provided by pandas profiling, a data scientist can generate an exhaustive profile report. Five major sections comprise the output report: Introduction, Variables, Interactions, Correlation, and Missing Values. Pandas-profiling is widely regarded as the most powerful Python library for exploratory data analysis (Brugman, 2019).

It was detected in the output report that certain attributes (DOD and CLIMPROCEDURECode) were missing values. The dataset was eliminated of these characteristics. Consequently, 133980 outpatient and 31289 inpatient records were acquired. At this point in the dataset, three distinct physician types (specialist, operator, and other) were merged into a single physician, for 100737 distinct physicians. The most crucial stage in comprehending the data and, consequently, detecting fraud consists of answering inquiries such as the number of beneficiaries in the dataset, the quantity and variety of physicians, the standing of physicians affiliated with various providers/hospitals, the correlation between patient age and medical claims, the correlation between patient age and chronic conditions, and so forth.

The goal of feature engineering is to select effective feature sets for different fraud models. After feature engineering, the dataset shown in Table 1 was obtained. Table 1 provides a description of each feature in the dataset, a description of the feature, and statistical information (range- max/min values-distribution-0% missing values) for each feature.

[1] https://www.kaggle.com/datasets/rohitrox/healthcare-provider-fraud-detection-analysis

Table 1. Dataset used for modeling

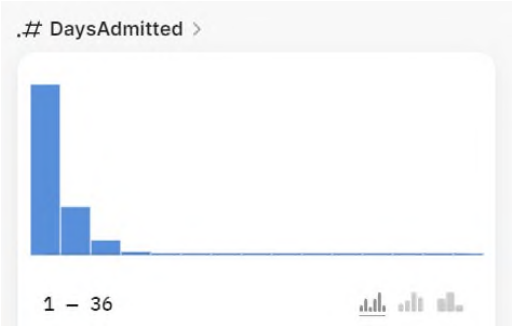
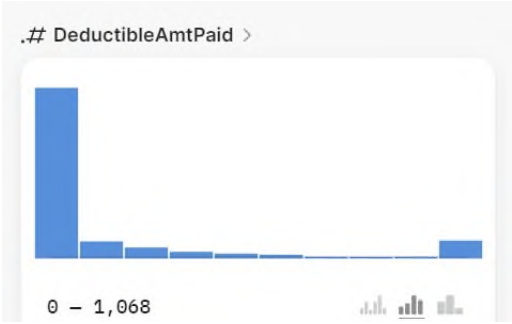
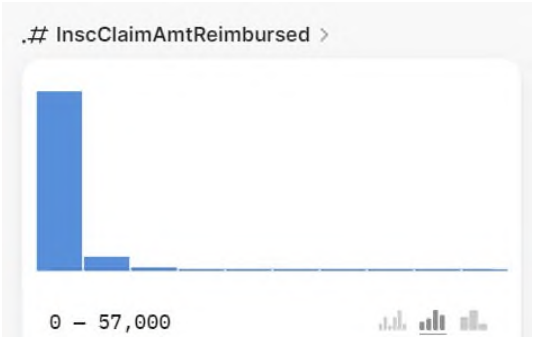
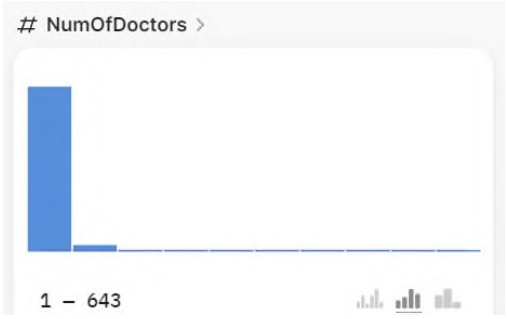
Feature	Description	Statistical Information
DaysAdmitted	Duration of the inpatient -hospital stay (days)	
DeductibleAmtPaid	Amount paid by the patient. (total amount requested-amount reimbursed)	
Provider	Provider's ID	
InscClaimAmtReimbursed	Amount reimbursed for the claim	
NumOfDoctors	Number of physicians	

Table 1. Continued

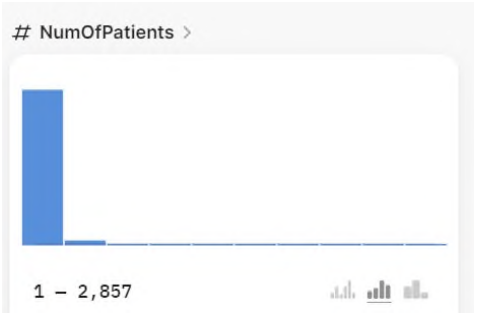
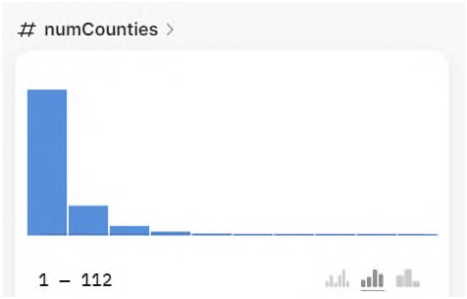
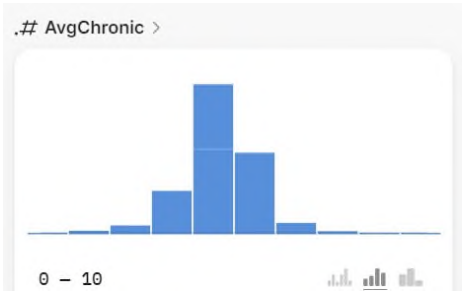
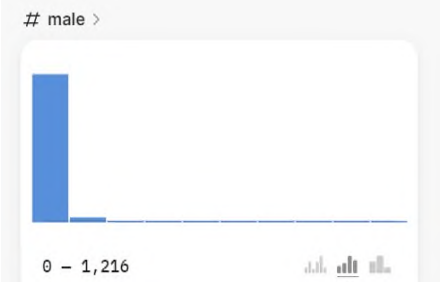
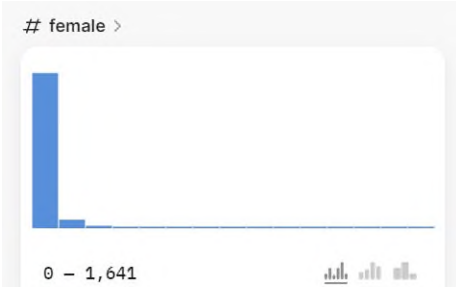
NumOfPatients	Number of patients	
numCountries	Number of countries (for each provider)	
AvgChronic	Average chronic diseases	
AvgClaim	Average claim	
Age	Patient's age	

Table 1. Continued

male	Number of men	
female	Number of women	
PotentialFraud	Whether the recording is fraudulent **(target class)**	

4.3. Modeling

Because the dataset is an imbalanced dataset; in other words, non-fraudulent records are more than fraudulent records, XGBoost, LGBM, and Random Forest (RF), which are ensemble learning-based models that work well on imbalanced datasets, were preferred. In addition, Logistic Regression (LR) was used because the binary classification problem was addressed after the data were balanced using the SMOTE method. The hyperparameters were first trained using the default parameters provided by the model library in Python. For the RF model, hyperparameter optimization was also performed (max depth: 4, number of trees: 500). In the modeling phase, the data were divided into a 70% training set and a 30% test set. The machine learning models employed are described in Table 2.

5. FINDINGS

In this section, the results of the LR, NB, SVM, RF, XGBoost, and LightGBM models applied to the preprocessed dataset are evaluated according to different performance metrics. The applied model parameters are first trained on the default parameters. As the accuracy metric may be misleading due to imbalances in the dataset, sensitivity/recall, and precision metrics are included.

The ROC/AUC metric for the LR model was also included in the findings. The confusion matrix, a matrix summarizing the performance of a machine learning model on a set of test data, is also discussed. Table 3 lists the performance metrics of the models. When evaluating classification models that attempt to predict a categorical label for each input sample, a confusion matrix is frequently used.

The number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) generated by the model on the test data is displayed in the matrix. Figure 1 shows the confusion matrix resulting from the RF model. The ROC Curve for the LR model is shown in Figure 2.

Table 2. Machine Learning Models

Model	Description
Logistic Regression (LR)	When using logistic regression, it is important to consider the connections between discrete variables. Logistic regression differs significantly from linear regression in that it uses binary or multiple outcome variables rather than numeric variables. As Bircan (2010) pointed out, this technology has widespread application in healthcare.
Naive Bayes (NB)	The Naive Bayes algorithm is a supervised machine learning algorithm that relies on conditional probability and applies the Bayes' theorem (Vangara et al., 2020). The proposed approach relies on the assumption that the properties of the input data are conditionally independent given class, which allows the algorithm to make quick and precise predictions.
Random Forest (RF)	An ensemble learning classifier called Random Forest (RF) generates multiple decision trees by selecting utilizing a randomly selected subset of training samples and variables. The RF classifier uses a CART collection to generate predictions (Breiman, 2001). Trees are generated by selecting a subset of training examples using a replacement method known as the bagging approach. It is possible for certain samples to be selected multiple times, whereas others may not be selected at all (Belgiu and Dra˘gut, 2016: 25).
Support Vector Machine (SVM)	Classification was performed by SVM using either a linear or nonlinear function. This involves estimating the most suitable function to separate the data, as Özkan (2016) explained. The algorithm places all feature vectors in a virtual space and divides the samples by a line known as a hyperplane. The hyperplane was designed to effectively separate classes by maximizing the margin (Burkov, 2019).
XGBoost	The development of this method was performed by Chen et al. (2016). This implementation of gradient boosting machines is highly advanced and can greatly enhance the computational power of boosting tree algorithms. It was created with a strong focus on optimizing model performance and computational speed. Boosting is a powerful technique in ensemble learning that involves adding new models to correct errors caused by existing models. Models are added to the model recursive until any further improvement is no longer detected (Ogunleye and Wang, 2020: 2133).
LightGBM	Ever since its introduction by Ke (2017), LightGBM has attracted significant research attention. LightGBM is a highly efficient implementation of gradient-boosting trees, and it is known for its adaptability and effectiveness. LightGBM primarily uses histogram algorithms and other algorithms to enhance the computational power and prediction accuracy of the algorithm. First, continuous feature values are expressed in M integers, and then a histogram of width M is decomposed. Using the decomposed values of the histogram, the data is analyzed to determine the decision tree. By leveraging the histogram algorithm, significant improvements can be made to the time complexity. Additionally, the fuzzy partitioning method outperforms the decision tree, making it a valuable approach (Wang et al., 2022: 261).

Table 3. Model Performances

Model	Accuracy	Precision	Sensitivity/ Recall	F1-Score
LR	0.820	0.796	0.823	0.809
NB	0.896	0.94	0.95	0.945
SVM	0.918	0.92	1.00	0.958
RF	0.823	0.808	0.914	0.858
XGBoost	0.925	0.95	0.97	0.96
LightGBM	0.930	0.95	0.98	0.965

There is a trade-off between Precision and Recall metrics, especially in healthcare. For example, high precision requires low FP; thus, a classifier that maximizes precision only returns strongly positive predictions, which may result in missing positive occurrences. In this context, which of the precision and recall metrics are maximized depends on the application. Recall is a statistic used when the expense of estimating FN is significant. It should be maximized. For example, if a Fraud Detection model misclassifies a fraudulent transaction as non-fraudulent, it may have adverse implications for the bank. Therefore, the Recall value is of critical importance because of the performed analyses.

As shown in Table 3, the SVM model performed the best. This is not surprising in the context of the SVM model structure. Fraud detection frequently involves complicated sets of data with numerous features. SVM is particularly suitable for datasets with many dimensions because it can effectively manage a large number of input parameters without observing a noticeable decline in performance. This is especially advantageous when handling sophisticated transactional data that involve multiple features.

The SVM model is followed by ensemble learning models (LightGBM, XGBoost, RF) when the Recall value is considered. When the precision value was considered, it was observed that the model with the best performance was LightGBM. The reasons for the high performance of ensemble learning models are that they require minimal data preprocessing in classification problems, are not sensitive to outliers, and can work with imbalanced and missing data, as well as the acquired dataset. As mentioned previously, the LightGBM and SVM models performed best on the recall metric. Figures 1 and 2 were chosen to illustrate the different models and their evaluation methods. Therefore, confusion matrix and ROC curve, which are the most commonly used modeling methods to provide visual support, were preferred.

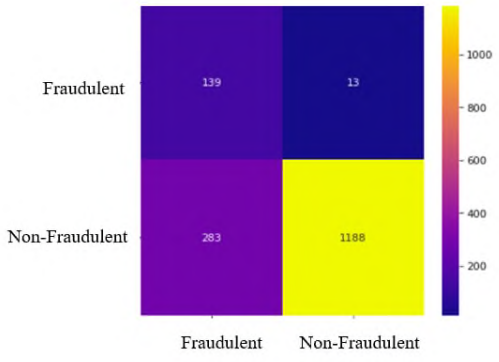


Figure 1. Confusion Matrix (RF)

In Figure 1, the confusion matrix for RF can be observed in terms of fraudulent and non-fraudulent activities. As can

be seen, The Random Forest model correctly predicted 139 of 152 fraudulent activities and 1188 of 1471 non-fraudulent activities.

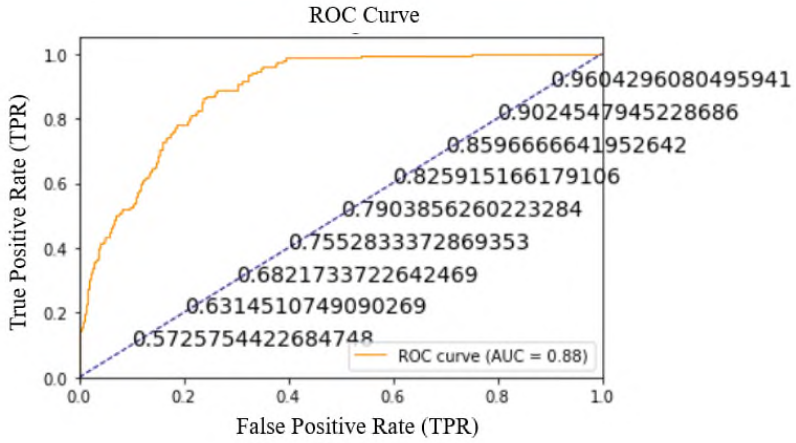


Figure 2. ROC Curve (LR)

Figure 2 shows the ROC curve for logistic regression. This method, which is frequently used, especially for classification problems in the medical domain, graphically presents the relationship between sensitivity (TP) and 1-specificity (FP) values to determine an ideal "cut-off" value. The area under the curve (AUC) provides a good measure of discrimination (Fan et al., 2006: 19). Figure 2 shows the ROC space. Accordingly, a well-performing classifier is expected to be near the upper left corner of the graph. As can be seen from the ROC curve for the LR model, the classifier performance reached approximately 96% because of the analysis. This demonstrates the success of the proposed LR model in binary classification.

6. DISCUSSION AND CONCLUSION

Medical fraud is a significant concern for numerous health systems. It occurs when an individual deliberately (willingly) submits false or deceitful statements, orchestrates the dissemination of such statements, or manipulates the facts to secure payment for healthcare services to which they are not legally entitled. Medical fraud may manifest itself through various means, including yet not limited to referrals for government-prohibited health care services, bribery, or kickbacks for publicly funded health care services (Centers for Medicare & Medicaid Services, 2021:6). With greater precision, the occurrences of medical fraud in the healthcare industry vary extensively and encompass a wide array of techniques and procedures. Schemes to commit medical fraud may involve the work of a single individual or the collaboration of an organization. Infiltrating healthcare systems' programs and functioning as healthcare providers or suppliers is not beyond the capability of organized criminal syndicates (Centers for Medicare & Medicaid Services, 2021:6).

The exact cost of medical fraud in society is unknown. According to estimates from the National Health Care Anti-Fraud Association, Medicare and Medicaid fraud cost taxpayers more than 100 billion a year (Zamost and Brewer, 2023). According to another estimate, the Medicare program in the US provides healthcare to more than 60 million US citizens, whereas Medicare loses $20 to $70 billion annually due to fraud, waste, and abuse (Johnson and Khoshgoftaar, 2019:31). Over the projected period, the worldwide healthcare fraud market is anticipated to increase at an annual growth rate of 20.45%, from US$1.65 billion in 2022 to US$5.03 billion in 2028 (Arizton, 2022).

Medical fraud can put the health and well-being of healthcare beneficiaries at risk and cost billions of dollars to many sectors of society, including beneficiaries of healthcare programs, healthcare institutions and organizations, and taxpayers. As countries' health services continue to serve increasing numbers of beneficiaries, the impact of these losses and risks grows. Today, medical fraud can frequently be detected and avoided using machine learning technologies that have an immense impact on health programs' capability to offer affordable, high-quality medical services.

This paper presents a comprehensive performance analysis of an example of the implementation of machine learning algorithms for medical fraud detection. This study uses machine learning methods to binary classify cases into cases where fraud is present and cases where fraud is absent to detect and correct fraud with higher accuracy. More precisely,

automatic classification of medical fraud can be achieved using machine learning. In the analysis phase, LR, Naïve Bayes, SVM, RF, XGBoost, and LightGBM models, which are frequently used for binary classification, were used. In evaluating the results of the analysis, not only the accuracy, precision and recall metrics were used to the imbalanced nature of the dataset. When focusing on accuracy and precision, the LightGBM model appears to be the best performing model (90.3%); however, in terms of precision, the SVM model yielded the highest results. In this case, the preferred model will vary depending on whether only the recall ratio is taken into account or the F1 score, which is the harmonic mean of the precision and recall ratio, is calculated and included in the analysis.

The potential for medical fraud to be detected by AI systems may raise the anxiety of healthcare professionals, including medical secretaries, billing specialists, and auditors, regarding their employment prospects. The potential outcome of AI taking over these tasks is that employees may experience anxiety regarding their employment status. Öztırak (2023) and Özbek (2024) examined the AI concerns of accounting professionals and innovation-oriented behaviors of employees, respectively. Nevertheless, this study primarily focused on the collaboration between humans and artificial intelligence. Simply put, when AI identifies fraudulent activities in extensive data collections, employees can assist the AI by examining and confirming these findings. This collaboration can help ensure the continued significance of employee responsibilities and foster a positive perception of AI as a supportive tool rather than a source of concern.

The demonstration of the significant performance of machine learning over other approaches, especially manual detection methods, proves that machine learning algorithms will be used much more frequently in future fraud detection applications.

Peer Review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- C.Ü., G.S.E.; Data Acquisition- C.Ü., G.S.E.; Data Analysis/Interpretation- C.Ü., G.S.E.; Drafting Manuscript- C.Ü., G.S.E.; Critical Revision of Manuscript- C.Ü., G.S.E.; Final Approval and Accountability- C.Ü., G.S.E.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

ORCID IDs of the authors

Ceyda ÜNAL 0000-0002-5503-8124
Gökçe Sinem ERBUĞA 0000-0003-1604-4668

REFERENCES

Abdallah, A., Maarof, M.A. and Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications, 68*, 90-113.,

Alpaydin, E. (2020). *Introduction to machine learning.* MIT press.

Altındiş, S., & Morkoç, İ. K. (2018). Sağlık hizmetlerinde büyük veri. Nigde Omer Halisdemir University *Academic Review of Economics and Administrative Sciences. 11*(2), 257-271.

Aral, K. D., Güvenir, H. A., Sabuncuoğlu, İ., & Akar, A. R. (2012). A prescription fraud detection model. *Computer Methods and Programs in Biomedicine, 106*(1),37-46. http://dx.doi.org/10.1016/j.cmpb.2011.09.003

Arizton. (2022). Healthcare Fraud Analytics Market- Global Outlook & Forecast 2023-2028. https://www.arizton.com/market-reports/healthcare-fraud-analytics-market

Avcı, M., and Teyyare, E. (2012). Sağlık sektöründe yolsuzluk: Teorik bir değerlendirme. TheInternationalJournalofEconomicandSocialResearch(IJESR) 8(2). 199-221.

Aydın, J. C., & Yaşar, G. (2020). Sağlık Harcamalarının Gelir Esnekliği Açısından Değerlendirilmesi: Sistematik Bir Derleme. JournalofAnkaraHealthSciences,9(1), 63-80.

Aydoğan Duman & Sağıroğlu, Ş. (2017, October). Heath care fraud detection methods and new approaches. *2017 International Conference on Computer Science and Engineering (UBMK)* (pp. 839-844). IEEE.

Bauder, R. A., and Khoshgoftaar, T. M. (2016, December). A probabilistic programing approach for outlier detection in healthcare claims. *2016 15th IEEE international conference on machine learning and applications (ICMLA)* (pp. 347-354). IEEE.

Bauder, R. A., Khoshgoftaar, T. M., Richter, A., & Herland, M. (2016, Kasım). Predicting medical provider specialties to detect anomalous insurance claims. *2016 IEEE 28th international conference on tools with artificial intelligence (ICTAI)* (ss. 784-790). IEEE.

Belgiu, M., & Drăguţ, L. (2016). Random Forest in Remote Sensing: A review of Applications and Future Directions. *ISPRS Journal of Photogrammetry and Remote Sensing, 114*, 24-31.

Bircan, H. (2004). Lojistik regresyon analizi: Tıp verileri üzerine bir uygulama. *Kocaeli University Journal of Social Sciences*, (8), 185-208.

Branting, L. K., Reeder, F., Gold, J., & Champney, T. (2016, Ağustos). Graph analytics for healthcare fraud risk estimation. *2016 IEEE/ACM Intl Conf. on Advances in Social Networks Analysis and Mining (ASONAM)* (ss. 845-851). IEEE.

Bolton, R. J., & Hand, D. J. (2001). Unsupervised profiling methods for fraud detection. *Credit scoring and credit control VII*, 235-255.

Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical science*, 17(3), 235-255.

Bozhenko, V. (2022). Tackling corruption in the health sector. *Health Economics and Management Review*, 3, 32-39.

Brugman, S. (2019). pandas-profiling: exploratory data analysis for Python. https://github.com/pandas-profiling/pandas-profiling.

Burkov, A. (2019). Hundred-page Machine Learning Book. Quebec City, QC, Canada: Andriy Burkov.

Centers for Medicare and Medicaid Services (CMS). (2021). Medicare Fraud & Abuse: Prevent, Detect, Report. https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/Fraud-Abuse-MLN4649244.pdf

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACCM signed international conference on knowledge discovery and data mining* (ss. 785-794).

Couffinhal, A., and Frankowski, A. (2017). Wasting with intention: Fraud, abuse, corruption, and other integrity violations in the health sector.

Dora, P. and Sekharan, G. H. (2015). Healthcare insurance fraud detection leveraging big data analytics. IJSR, 4(4), 2073-2076.

Ekin, T. (2019). An Integrated Decision-Making Framework for Medical Audit Sampling.

Ekin, T., Frigau, L., & Conversano, C. (2021). Health care fraud classifiers in practice. *Applied Stochastic Models in Business and Industry, 37*(6), 1182-1199.

European Commission, Directorate-General for Migration and Home Affairs (EC). (2017). Weistra, K., Swart, L., Oortwijn, W., et al., *Updated study on corruption in the healthcare sector: final report*, Publications Office. https://op.europa.eu/en/publication-detail/-/publication/9537ddb7-a41e-11e7-9ca9-01aa75ed71a1/language-en

Euronews. (2022). Sağlık harcamalarının milli gelire oranı: OECD ve AB'de sağlığa en çok pay ayıran ülkeler hangileri? Retrieved from https://tr.euronews.com/next/2022/04/05/sagl-k-harcamalar-n-n-milli-gelire-oran-oecd-ve-ab-de-sagl-ga-en-az-pay-ay-ran-ulke-turkiy

J. Fan, S. Upadhye. Ve Worster, A. (2006). Understanding Receiver Operating Characteristic (ROC) Curves. *Canadian Journal of Emergency Medicine, 8*(1), 19-20.

Federal Bureau of Investigation (FBI). (1989). White collar crime: a report to the public. Washington, DC: Government Printing Office.

Francis, C., Pepper, N., & Strong, H. (2011, Ağustos). Using support vector machines to detect medical fraud and abuse. *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (ss. 8291-8294). IEEE.

Gee, J., Button, M., & Brooks, G. (2010). The financial cost of healthcare fraud: data from around the world. (*London: MacIntyre Hudson/CCFS*.

Georgetown University Memory Disorders Program (2023). Allegations of fraud in Alzheimer's disease research: Death of the amyloid hypothesis?, https://memory.georgetown.edu/news/allegations-of-fraud-in-alzheimers-disease-research-death-of-the-amyloid-hypothesis%EF%BF%BC/

He, H., Graco, W., & Yao, X. (1999). Application Of Genetic Algorithm And K-Nearest Neighbor Method İn Medical Fraud Detection. Lecture Notes in Comput. Sci. 1585 74–81. Springer, Berlin.

He, H. X., Wang, J. C., Graco, W., & Hawkins, S. (1997). Application of Neural Networks detect Medical Fraud. Expert Systems with Applications 13 329–336.

ISSA. (2022). Detecting fraud in healthcare through emerging Technologies. https://ww1.issa.int/analysis/detecting-fraud-health-care-through-emerging-technologies

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Medicare fraud detection using neural networks. *Journal of Big Data, 6*(1), 1-35.

Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., and Arab, M. (2015). Using data mining to detect healthcare fraud and abuse: a review of literature. *Global Journal of Health Sciences 7*(1), 194.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma W, et al. & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems, 30*.

Kirlidog, M., and Asuk, C. (2012). A fraud detection approach with data mining in health insurance. *Procedia-Social and Behavioral Sciences, 62*, 989-994.

Kurşun, A. (2021). Büyük Veri ve Sağlık Hizmetlerinde Büyük Veri İşleme Araçları. *Hacettepe Sağlık İdaresi Dergisi, 24*(4), 921-940.

Küçük, A. (2022). Sağlık Hizmet Ödemelerinde Usulsüzlük Türleri ve Mücadele Stratejileri. *Sayıştay Dergisi, 33*(127), 585-607.

Li, J., Huang, K. Y., Jin, J., & Shi, J. (2008). Survey on statistical methods for healthcare fraud detection. *Health Care Management Science*, 11(3), 275-287. http://dx.doi.org/ 10.1007/s10729-007-9045-4.

C. Lin, C. M. Lin, S. T. Li, and S. C. (2008). Intelligent physician segmentation and management based on KDD approach. Expert Systems with Applications, 34(3), 1963-1973. http://dx.doi.org/10.1016/j.eswa.2007.02.038

Liou, F. M., Tang, Y. C., & Chen, J. Y. (2008). Detecting hospital fraud and claim abuse through diabetic outpatient services. *Health Care Manag Sci. 11*, 353-358.

Liu, Q., |& Vasarhelyi, M. (2013, Kasım). Healthcare fraud detection: A survey and clustering model incorporating geo-location information. *Brisbane, 29th World Continuous Auditing and Reporting Symposium (29WCARS)*.

Major, J. A., & Riedinger, D. R. (1992). Efd: A Hybrid Knowledge/Statistical-Based System for Detecting Fraud. *International Journal Of Intelligent Systems*. 687–703.

National Health Care Anti-Fraud Association (NHCAA). (2021). The Challenge of Health Care Fraud. https://www.nhcaa.org/tools-insights/

about-health-care-fraud/the-challenge-of-health-care-fraud/

OECD (2020). Public Integrity for an Effective COVID-19 Response and Recovery. https://read.oecd-ilibrary.org/view/?ref=129_129931-ygq2xb8qax&title=PublicIntegrityforanEffectiveCOVID-19ResponseandRecovery

Ogunbanjo, G. A., and D. K. (2014). Ethics in health care: healthcare fraud. *South African Family Practice, 56*(1), S10-S13.

Ogunleye, A., & Wang, Q. G. (2019). XGBoost Model for Chronic Kidney Disease Diagnosis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 17*(6), 2131-2140.

Orhan, M. S., & Serçemeli, M. (2015). İç denetim stratejisinde sürekli denetimin uygulanabilirliğine ilişkin bir araştırma. *Giresun Üniversitesi İktisadi ve İdari Bilimler Dergisi, 1*(2), 83-110.

Özbek, A. (2024). Muhasebe Meslek Mensuplarının Yapay Zekâ Kaygılarının Gelecekte İstihdam Edilebilirlik Algıları Üzerine Bir Çalışma. *Alanya Akademik Bakış, 8*(1), 254-267.

Özkan, Y. (2016). Veri Madenciliği Yöntemleri. İstanbul: Papatya Yayıncılık.

Öztırak, M. (2023). A study on the impact of artificial intelligence anxiety on innovation-oriented behaviors of employees. *Optimum Ekonomi ve Yönetim Bilimleri Dergisi, 10*(2), 267-286.

Price, M., & Norris, D. M. (2009). Health care fraud: physicians as white-collar criminals? *Journal of the American Academy of Psychiatry and the Law Online, 37*(3), 286-289.

Saravanan, R., and Sujatha, P. (2018, Haziran). State-of-the-art machine learning algorithms: a perspective of supervised learning approaches in data classification. In *2018 Second international conference on intelligent computing and control systems (ICICCS)* (ss. 945-949). IEEE.

Savedoff, W. D., & Hussmann, K. (2006). The causes of corruption in the healthcare sector: a focus on health care systems. *Transparency International. Global Corruption Report*, 4-16.

Shan, Y., Jeacocke, D., Murray, D. W., & Sutinen, A. (2008, Kasım). Medical specialist billing patterns for health service management. *Proceedings of the 7th Australasian Data Mining Conference, 87* (ss. 105-110).

Shan, Y., Murray, D. W., & Sutinen, A. (2009, Aralık). Discover inappropriate billings using a local density-based outlier detection method. *Proceedings of the Eighth Australasian Data Mining Conference, 101* (ss. 93-98).

Shin, H., Park, H., Lee, J., & Jhee, W. C. (2012). A scoring model to detect abusive billing patterns in health insurance claims. *Expert Systems with Applications*, 39(8), 7441-7450. http://dx.doi.org/10.1016/j.eswa.2012.01.105.

Sokol, L., Garcia, B., Rodriguez, J., West, M., & Johnson, K. (2001). Using data mining to find fraudulent HCFA healthcare claims. *Topics in Health Information Management*, 22(1), 1-13.

Srinivasan, U., & Arunasalam, B. (2013). Leveraging big data analytics to reduce healthcare costs. *IT professional, 15*(6), 21-28.

Thomson Reuters. (2021). Organized crime using sophisticated technology in next wave of government healthcare fraud schemes. https://www.thomsonreuters.com/en-us/posts/investigation-fraud-and-risk/healthcare-fraud-webinar/

Transparency International. (2006). Global Corruption Report 2006. https://images.transparencycdn.org/images/2006_GCR_HealthSector_EN.pdf

Transparency International. (2023). Health. https://www.transparency.org/en/our-priorities/health-and-corruption

Turgay, İ., Doğan, S., & Mengi, B. T. (2020). İç Denetim Faaliyetlerinde Sürekli Denetim: Analitik İnceleme Prosedürlerinin Kullanımı. *Denetişim*, (21), 5-26.

U.S. Department of Justice (2023). Retrieved from https://www.justice.gov/opa/pr/hospice-medical-director-sentenced-150m-hospice-fraud-scheme.

U.S. Department of Justice (2024). Criminal Resource Manual CRM 500-999, 976- Health Care Fraud—Generally. https://www.justice.gov/archives/jm/criminal-resource-manual-976-health-care-fraud-generally#

USSC (2022). Quick Facts—Health Care Fraud Offenses. https://www.ussc.gov/sites/default/files/pdf/research-and-publications/quick-facts/Health_Care_Fraud_FY22.pdf

Van Capelleveen, G., Poel, M., Mueller, R. M., Thornton, D., and van Hillegersberg, J. (2016). Outlier detection in healthcare fraud: A case study in the Medicaid dental domain. *International Journal of Accounting Information Systems 21*, 18-31.

Vangara, V., Vangara, S. P., & Thirupathur, K. (2020). Opinion mining classification using naive bayes algorithm. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 9(5), 495-498.

Vian, T. (2020). Corruption and administration in healthcare. In *Handbook on corruption, ethics, and integrity in public administration* (pp. 115-128). Edward Elgar Publishing.

Vincke, P. ve Cylus, J. (2011). Healthcare fraud and corruption in Europe: An overview. Eurohealth, 17(4), 14-18.

Wang, D. N., Li, L., & Zhao, D. (2022). Corporate finance risk prediction based on LightGBM. *Information Sciences*, 602, 259-268.

World Health Organization (WHO). (2023a). *World health statistics 2023: monitoring health for the SDGs, sustainable development goals.* World Health Organization. Retrieved from https://www.who.int/publications-detail-redirect/9789240074323.

World Health Organization (WHO). (2023b). Reducing Health System Corruption. https://www.who.int/activities/reducing-health-system-corruption

Zamost, S. & Brewer, C. (2023). Inside the mind of criminals: How to brazenly steal $100 billion from Medicare and Medicaid. https://www.cnbc.com/2023/03/09/how-medicare-and-medicaid-fraud-became-a-100b-problem-for-the-us.html#:~:text=Fraud%20flourishes&text=Taxpayers%20are%20losing%20more%20than,Health%20Care%20Anti%2DFraud%20Association.

Zhang, C., Xiao, X., & Wu, C. (2020). Medical fraud and abuse detection system based on machine learning. *International journal of environmental research and public health, 17*(19), 7265.

Zhu, S., Wang, Y., & Wu, Y. (2011, Ağustos). Health care fraud detection using nonnegative matrix factorization. 2011 6th International Conference on Computer Science & Education (ICCSE) (ss. 499-503). IEEE.

How cite this article

Ünal, C., & Erbuğa, G. S. (2024). Detection and Prevention of Medical Fraud using Machine Learning. *Acta Infologica, 8*(2), 100-117. https://doi.org/10.26650/acin.1463879

İSTANBUL UNIVERSITY PRESS

Acta Infologica
acin 2024, 8 (2), 118–132
DOI: 10.26650/acin.1422463

Research Article

A Qualitative Study on Job Satisfaction of Public Information Workers

Vildan ATEŞ[1], Hatice GÜNAY[2]

[1]Ankara Yıldırım Beyazıt University, Business School Department of Management Information Systems, Ankara, Türkiye
[2]Ankara Yıldırım Beyazıt University, Social Sciences Institute, Ankara, Türkiye

Corresponding author : Vildan ATEŞ
E-mail : vates@aybu.edu.tr

ABSTRACT

Organizations worldwide, in both the private and public sectors, depend on their workforce to achieve optimal productivity, resulting in enhanced organizational effectiveness. Therefore, ensuring employee job satisfaction is a necessity for every organization. This study investigates the job satisfaction of public information technology (IT) employees in Türkiye. In this study, the qualitative research method was preferred, and interviews were conducted with 10 participants using a semi-structured interview form. The data obtained from the interviews was analyzed using descriptive and content analysis techniques. The study revealed that the participants' qualifications were compatible with their job requirements. However, it was found that their overall job satisfaction was insufficient. Furthermore, the study revealed varying perspectives on topics such as employee satisfaction with workload, salary, corporate support, commitment, work-life balance, teamwork, and attitude. The results of this study also have practical implications for managers in public institutions because they provide them with a more comprehensive understanding of the factors that contribute to job satisfaction among IT public employees. The conclusion section presents recommendations for public institution managers and decision makers to enhance the job satisfaction of IT employees in public institutions.

Keywords: Public Institution, Information Technology Employees, Job Satisfaction, Qualitative Research Method.

Submitted : 19.01.2024
Revision Requested : 08.08.2024
Last Revision Received : 22.08.2024
Accepted : 02.09.2024
Published Online : 24.09.2024

1. INTRODUCTION

In today's competitive business environment, the success and sustainability of organizations depend on employee performance and satisfaction. Job satisfaction refers to employees' general feelings about their jobs. The state of well-being and happiness related to performance in and around the workplace. The concept of job satisfaction is so broad that it cannot be defined in a single, comprehensive manner. The positive affect that results from an evaluation of one's job and work experience is also referred to as job satisfaction (Riyanto, Endri, & Herlisha, 2021). The term "job satisfaction" describes the emotions that arise from an employee's assessment of how much their job meets their needs. It concerns an employee's emotional state about how much they enjoy their role at work (Kong, Jiang, Chan & Zhou, 2018). It is a multifaceted construct comprising various emotional states and conditions. According to Miner (1992), job satisfaction is an important factor in the functioning of an organization and is one of the main indicators of an organization's health. More employee productivity translates into better service and value for your customers. Job satisfaction increases productivity from employees. Increased customer pleasure and loyalty because of this value supports an organization's profitability and long-term success. Therefore, organizations attach great importance to job satisfaction. Satisfied workers are eager to complete their responsibilities meaningfully, whereas dissatisfied workers are less motivated. Consequently, satisfied workers frequently show greater interest in the business itself. Because satisfied employees contribute to the effectiveness and long-term success of organizations, employee satisfaction levels are important. The effectiveness and productivity of an organization depend on its personnel and "a happy employee is an effective employee". An organization cannot develop without considering the use of employees' talents and the improvement of their working conditions. Organizations with satisfied employees are probably more successful than other organizations (Basar, 2011).

In the context of an increasingly competitive and complex business environment, the significance of job satisfaction is also growing (Inayat & Khan, 2021). As a result of the literature research, it has been observed that knowing the satisfaction level of employees is important both for employees and institutionally in terms of organizational efficiency (Demiralp, 2006; Sahin, 2015). In addition, due to global competition and technological developments, it is seen that business management has realized that sustaining their existence is not only by attracting a qualified workforce to the business but also by keeping this workforce in the business and ensuring their satisfaction should be the primary goals of the organization (Cabukel, 2008). In addition, in today's volatile business environment, where employee turnover rates are high, organizations need to prioritize employee job satisfaction a priority (Judge, Zhang, & Glerum, 2020). Studies have also reported that increased job satisfaction increases employee performance (Habib et al., 2014; Inuwa, 2016; Phuong & Tran, 2020; Susanto, Syailendra, & Suryawan, 2023). A satisfied workforce not only meets organizational goals but also demonstrates organizational citizenship behaviors, including punctuality, productivity, commitment, and satisfaction with their work and lives. In this contemporary digital era, the role of information technology professionals is becoming increasingly crucial for both public and private sector organizations. An IT employee works with computer systems, software, and networks to ensure the proper functioning and security of an organization's IT systems and technologies. They design, develop, and maintain technology solutions. IT employees may work in programing, database management, cybersecurity, network management, and technical support, and they are responsible for designing, building, and managing information systems, networks, and databases in organizations, as well as developing new software applications and technologies. IT employees are frequently responsible for devising new goods and services, fixing issues, or developing plans of action and strategies to improve company outcomes. Therefore, they are the backbone of organizations. Ensuring that technology works effectively and efficiently in organizations ensures that they have essential positions and are critically important employees.

The first and most important purpose of a business is to make a profit. However, the primary purpose of public institutions is to satisfy citizen satisfaction with a service. In this case, public institution workers' job satisfaction becomes even more important. A positive attitude and behavior on the part of a worker is likely to result from the worker's belief that the employer values, respects, and treats workers well. The present research focuses on job satisfaction as a key outcome. Numerous studies have been conducted on the job satisfaction of employees in Türkiye. The participants in these studies included hospital employees (Gider, Akdere, & Top, 2019), university academic and administrative staff (Küskü, 2003; Toker, 2011), teachers (Aytac, 2020; Baykara & Orhan, 2020; Buyukgoze-Kavas, et al., 2014), municipality employees (Türkyilmaz et al., 2011), manufacturing factory employees (Asgarova, 2019; Demiral, 2017), police officers (Buker & Dolu, 2010), agricultural workers (Çevrimli, 2020), nurses (Kapucu et al., 2017), employees of small and medium-sized companies (Ayranci, 2011; Halil & Selim, 2007), and those working in the tourism sector (Kara, Uysal & Magnini, 2012; Tepeci, 2005; Ozturk & Hancer, 2011). This study makes a notable

contribution to the existing literature because it is the first to investigate job satisfaction among public IT employees in Türkiye.

The objective of this study is to determine the level of job satisfaction among public IT employees. To achieve this objective, the research question can be expressed as follows: "What are the opinions of public institution IT employees regarding job satisfaction?" This study provides a new perspective on the research on employee experience and to assist managers in gaining a deeper understanding of IT employees in public institutions. Since recent studies have considered the views of private sector employees, this study is valuable and unique in its focus on job satisfaction among public IT employees. Furthermore, the findings of this study are crucial for decision makers responsible for public IT employee policies. It is our hope that they will gain insight into the job satisfaction status of IT employees and the factors affecting it and that they will consider these study results when making necessary arrangements. Additionally, the results of this study will contribute to advances in the fields of organizational behavior and human resources management in public institutions.

This study is composed of five sections. The first two sections present the introduction and literature review, and the third section presents the methodology. The fourth section presents the findings of the study, and the fifth and final sections present the conclusions and recommendations. At the end of the study, references are provided.

2. LITERATURE REVIEW

Job satisfaction remains one of the most extensively studied attitudes in industrial and organizational psychology. Scholars and practitioners alike have acknowledged the significance of job satisfaction, citing its utility in predicting crucial organizational effectiveness outcomes (Judge, Zhang, & Glerum, 2020). Research shows that satisfied employees are more likely to become loyal champions, ambassadors, and advocates for their organizations. Research on job satisfaction is of interest to both individuals working in organizations and those studying job satisfaction. Job satisfaction is closely linked to numerous organizational phenomena, such as motivation, performance, leadership, attitude, conflict, and morality. The literature review reveals that researchers have attempted to identify the different components of job satisfaction, evaluate their relative importance, and investigate their impact on worker productivity. This section examines these studies and their findings.

One of the greatest pioneers of job satisfaction studies is the Hawthorne study conducted between 1924 and 1933. Because of these studies, it was first determined that changes in working conditions temporarily increase productivity, which was called the Hawthorne Effect. However, it was later found that this increase was due to observation, not working conditions. Considering this information, other factors related to job satisfaction have begun to be investigated by providing evidence that employees work for purposes other than wages (Mishra, 2013).

Three fundamental relationship impact employee satisfaction. These are the employee's relationship with the organization, the relationship with the manager, and coworkers (Tang, Siu, & Cheung, 2014). Other researchers have also identified the employee-manager relationship as a significant factor that influences employee job satisfaction (Edgar & Geare, 2005; Fila et al., 2014). Furthermore, the employee-organization relationship encompasses the employee's alignment with corporate strategy and company goals, and their dedication to these objectives. This relationship is also influenced by other factors, including the extent to which companies facilitate an optimal work-family balance for their employees (Adams, King, & King, 1996; Allen, Shore & Griffeth, 2003). Because of the literature review, it was found that the job satisfaction of employees in different sectors in different countries was investigated. One of these studies examined the effect of job satisfaction on the performance of employees in private sector organizations in Peshawar, Pakistan. The study findings indicate a significant correlation between an individual's occupation and their job satisfaction level. Additionally, a positive relationship between job satisfaction and employee performance was confirmed. Consequently, it can be concluded that employees who are satisfied with their work perform better than those who are dissatisfied, thereby playing an instrumental role in the growth and success of organizations. Based on the study's insights, the researchers proposed strategies to enhance job satisfaction among bank employees (Inayat & Jahanzeb, 2021). Another study conducted in Pakistan with employees from the telecommunications sector revealed a negative correlation between job stress and employee job satisfaction (Mansoor et al., 2011). A similar study was conducted by Islam et al. (2000) on the job satisfaction levels of women employed in public and private banks in Bangladesh. This study evaluated women's job satisfaction based on factors such as salary, better job opportunities, and the presence of a manager. The results revealed that the factor that most affects female employees' job satisfaction is salary. In a separate study, Kabir and Parvin (2011) examined the job satisfaction of employees of various pharmaceutical companies. The findings indicate that salary, productivity at work, supervision, and coworker relations are the primary determinants of job satisfaction. Overall, employees in the pharmaceutical sector exhibit a favorable level of job satisfaction. A study conducted in Türkiye with bank employees (Güner & Çetinkaya Bozkurt, 2017) identified

several negative factors that affect job satisfaction. These include inadequate physical working conditions, exposure to noise and excessive intensity, long working hours, a stressful work environment, high work intensity, high volume of telephone calls and emails, insufficient wages, lack of overtime pay, lack of necessary financial compensation despite the job involving financial risk, frequent problems with taking leave, and extreme fatigue caused by long working hours. McMurtrey, Grover, Teng, and Lightner (2002) conducted a study on the job satisfaction of IT professionals in an environment where computer-aided software engineering tools were used. The study found that employees with a technical career orientation exhibited a significantly higher level of job satisfaction than those with a managerial career orientation. Another study was conducted by Lim (2008) on IT employees working in libraries. This study examined IT employees' job satisfaction in terms of demographic, socioeconomic, and job-related variables. As a result of this study, it was found that salary, sense of belonging, belief in the desire to belong, and the feeling of acceptance and promotion positively affected job satisfaction.

The literature review presents recent studies on job satisfaction, including Yüksekbilgili's (2022), who emphasized the significance of job satisfaction in retaining efficient employees within an organization. The study highlights that job satisfaction has a positive impact on employee performance. Additionally, a study conducted on 119 managers and civil servants in the Iraqi Ministry of Electricity found that a positive organizational climate had a positive effect on job satisfaction (Al-Ani, 2019). This study investigates the relationships among knowledge management processes, IT employee satisfaction, and organizational performance among 248 academics and administrative personnel working in higher education institutions. The study found that information management processes have a significant impact on employee satisfaction with IT. Furthermore, this study revealed that IT employee satisfaction improves organizational performance in higher education institutions (Sahibzada et al., 2019). In his 2018 master's thesis, Güneş Aydemir identified various factors that determine job satisfaction, such as job quality, work environment, relationships with managers and colleagues, salary, career opportunities, and fringe benefits. Job satisfaction is generally defined as an employee's overall contentment with their job.

The demand for talented IT professionals remains high globally. Therefore, recruiting and retaining IT professionals is vital for many organizations (Metin, 2023). Metin (2023) conducted a study on the mediating role of perceived organizational support and the psychological empowerment retention of public IT specialists in Türkiye. Data for the study were gathered from 361 IT specialists working in 31 public institutions in Türkiye with large-scale IT units. The study revealed that 40.7% of specialists in Türkiye's public IT sector are considering quitting their current positions.

A literature review has revealed that job satisfaction has been researched among employees in various occupations. Nevertheless, there is a significant gap in the existing literature about the job satisfaction of employees in public institutions, particularly in the public IT sector in Turkey. This study aims to address this gap in the literature.

3. METHOD

The methodology section presents in detail the research method, the working group, development of the data collection tool, data collection and analysis processes, and how the validity and reliability of the research were ensured.

3.1. Method of Study

This study was conducted using a qualitative research methodology. This method was preferred because it was planned to conduct interviews with public IT employees in line with the purpose of the study. In addition, it also aimed to collect in-depth information from the participants. This study, which attempts to understand the perspectives of IT sector employees working in public institutions to determine their job satisfaction, is phenomenological research, which is one of the qualitative research method designs and aims to reveal the essence of a person's perception of their experiences.

3.2. Working Group

The population of this study comprises IT employees working in public institutions. In the selection of the study group, convenience sampling and criteria sampling methods were combined. The reason for choosing convenience sampling is that it enables participants who are suitable for the research to be included in the research most easily; in other words, these participants are the easiest for the researcher to access. Criterion sampling was also preferred because it enables selection of participants who meet the criteria appropriate for the research (Patton, 2003). The criteria for this study are that the participants are easily accessible and have worked in the field of informatics at a public institution for at least one year. Since it is impossible to reach and interview all IT sector personnel working in public institutions

in Türkiye, the study group of this study consists of public IT employees working in three public institutions in Ankara who participate in the research voluntarily. Interviews were conducted with 10 participants who met the inclusion criteria and worked in the informatics field. Care was taken to ensure that the participants voluntarily contributed to the study and were informed about the study. The personal information of the participants was not included in the study, and the participants were identified by numbering them as "P1, P2, P3,...P10".

3.3. Data Collection Instrument

In this study, a semi-structured interview form was employed as the instrument for data collection. A draft interview form was created based on the literature review conducted by the researchers. In addition, the "Two-Factor Theory of Motivation" developed by American psychologist Frederick Herzberg guided the preparation of the interview questions. In his theory, Herzberg named the factors related to work-related elements as intrinsic or motivating factors and the factors related to negative attitudes toward work as extrinsic or hygiene factors. Herzberg observed that these factors have a major influence on motivation, happiness, and business success. These two factors are hygiene and motivating factors. The sub-factors of the hygiene factor are company policies, relationships, working conditions, salary, status, and security. The sub-factors of the motivating factor are achievement, recognition, work itself, responsibility, progress, and growth (Nickerson, 2023). Following the literature review and Herzberg's two-factor motivation theory, it was decided that four main factors affect job satisfaction (Al Maqbali, 2015; Indiafreenotes, 2020; Nickerson, 2023; Pandey & Asthana, 2017; Waqas et al., 2015). The interview form was designed to assess four main factors affecting job satisfaction. These factors include organizational factors, the work environment, work itself, and personal factors.

There are 20 questions on the proposed interview form. For the suitability of the questions to the subject, opinions were obtained from two experts who have conducted research in the field of job satisfaction. Based on the feedback received, two questions were added, and two were removed. A preliminary test of the interview form was conducted with five IT employees from a public institution to assess the clarity and comprehension of the interview questions. The participants also indicated that the questions were well-structured and straightforward. Based on expert recommendations and pre-test feedback, a final version of the interview form was obtained. In the final form, there are 20 questions on the interview form. The interview form comprises two parts. In the first part, the participants were informed about the research, and demographic information (gender, age, education level, department, title, years of service) about the participants was included. In the second part, there are 20 interview questions.

Table 1. *Four Factors Affecting Job Satisfaction and Question Distribution*

Factors Affecting Job Satisfaction	Number of Questions	Question Numbers
Institutional factors	5	2., 8., 10., 12., 15.
Work environment	4	5. 9. 11. 13.
Work itself	3	1., 6., 14.
Personal factors	3	3., 4., 7.

Table 1 shows the distribution of questions according to the categories used in the preparation of the interview form and the four factors that affect job satisfaction. As shown in Table 1, there are 15 questions. The remaining five questions ask about the advantages and disadvantages of being an IT employee in a public institution, the aspects they are satisfied with, the aspects they are satisfied with, and what they would like to add. Finally, Ankara Yıldırım Beyazıt University Social and Human Sciences Ethics Committee was applied for the evaluation of the study, and as a result of the evaluation, approval (Decision no. /09-215 Date: 22.11.2023) was obtained for the ethical suitability of the research. Written informed consent was obtained from all participants who participated in this study.

3.4. Data Collection Process

The data were collected using a semi-structured interview form. Six interviews were conducted face-to-face and four were conducted online using the ZOOM program. The researcher conducted all interviews unbiased. The interviews were conducted in the participants' offices. Audio recordings were taken for face-to-face interviews and video recordings were taken for online interviews, with the researcher also taking notes on the responses given. The average interview duration was 45 min. Data collection occurred between December 1 and 30, 2023. Transcripts were created for each participant after the interviews were conducted. The transcripts were written verbatim to capture the participants'

responses accurately. Subsequently, the audio and video recordings were compared with the transcripts. Transcripts were sent to each participant by email for approval. The data analysis began after obtaining approval.

3.5. Data Analysis

Qualitative data analysis approaches, namely descriptive analysis and content analysis, were employed to analyze data obtained from the participants through interviews. Direct quotation marks were used in the descriptive analysis to effectively reflect the views of the participants. Content analysis focused on details and identified concepts and themes. Tables were also created to increase comprehensibility and to summarize the collected data.

3.6. Validity and Reliability

To ensure the validity of this study, Herzberg's two-factor motivation theory and a literature review were used while preparing the interview questions. In addition, expert opinions were obtained before finalizing the questions on the interview form, and pre-tests were conducted with five participants. In the descriptive analysis used in the data analyses, direct quotes were included in the responses given by the participants. The data collected immediately after the interviews were converted into written forms and interview transcripts were prepared. The interview transcripts were checked against audio and video recordings of the interviews. The interview transcripts were sent to the participants. After the participants approved the interview transcripts, the data were analyzed. For the reliability of the study, the researcher conducted the interviews with an impartial attitude. The participants were informed about the research, and the environment and processes in which the interviews were conducted were explained. In addition, all the obtained data (audio and video recordings and notes) were stored in an encrypted environment that only the researchers could access for review when necessary.

4. FINDINGS

This section is divided into two parts. The first part provides an overview of the demographic characteristics of the participants, and the second part presents the findings. The findings were obtained by applying descriptive and content analyses to the participants' responses to each question. In the initial section of the interview form, participants were asked to provide demographic information, including gender, age, education level, department, title, working year, and organizational status. This information is summarized in Table 2, which indicates that the participants are all public sector employees, and they provide a range of demographic information. Table 2 shows that seven participants were female and three were male. Notably, a higher proportion of female IT employees participated in the study. Although most participants work as analysts in the IT sector, the majority have been doing so for 11-15 years. This suggests that they have significant experience in the field.

Table 2. *Demographic Information about the Participants*

No	Gender	Age	Education Level	Department	Title	Working Year
P1	Male	21-30	Bachelor's degree	Computer Engineering	Engineer/Analyst	1-5 Years
P2	Female	31-40	Doctoral Degree	Electronics and Communication	Engineer/Analyst	11-15 Years
P3	Male	21-30	Bachelor's degree	Public Administration	Technical staff	1-5 Years
P4	Female	21-30	Bachelor's degree	Management Information Systems	Engineer/Analyst	1-5 Years
P5	Female	41-50	Doctoral Degree	Electrical and Electronics Engineering	Engineer/Analyst	16-20 Years
P6	Female	31-40	Bachelor's degree	Industrial Engineering	Engineer/Analyst	11-15 Years
P7	Female	31-40	Master's degree	Industrial Engineering	Engineer/Analyst	16-20 Years
P8	Male	31-40	Associate degree	Computer Programing	Technical Staff	6-10 Years
P9	Female	31-40	Bachelor's degree	Physics	Project Manager	11-15 Years
P10	Female	31-40	Master's degree	Business	Engineer/Analyst	11-15 Years

In the first question of the interview form, the participants were asked to explain the relationship between their job and their professional qualifications. Upon analyzing the responses provided by the participants to this question, it was observed that half of them (P2, P3, P6, P8, P9) believed that their current occupation aligned well with their professional abilities. Participant P8 stated "*There is a strong alignment. Working in this sector allows me to use the technical skills and knowledge I have.*" Conversely, three participants (P5, P7, P10) expressed partial compatibility, whereas two participants (P1, P4) pointed out a lack of compatibility. In this context, participant P4 responded to this question as follows, "*My professional qualifications are at a higher level for the job I work in, and they are more than enough for the job I work in. Sometimes it even makes me think that I am blunting myself*".

On the interview form, the second question was about the organization's support for staff's professional development. Participants P1, P4, P5, P7, and P10 stated that no support was provided, while other participants mentioned that training and seminars were organized. Participant P7 stated that "*There is no special support, training, etc. is limited*", while participant P5 responded to the question as follows "*There has been no training provided by my organization so far. I have been improving myself through my education, career, and personal efforts. However, the ordinary flow of work provides a certain amount of development (such as meetings and workshops attended). However, this is at the level inherent in every job.*"

In the third question of the interview form, the participants were asked about institutional factors that make them see themselves as part of the organization they work for. Upon analysis of the participants' responses to this question, two individuals (P1 and P10) indicated that they did not perceive themselves as part of their organizations. On the other hand, other participants stated different organizational factors. For instance, participant P8 responded to this question by stating, "I consider myself a part of the organization I work for because they take our ideas into consideration, communication is effortless, and they support our personal growth." Table 3 summarizes the institutional factors mentioned by the participants and the participants who mentioned them. An analysis of Table 3 reveals that the evaluation and evaluation of employee ideas, collaboration with team members, and valuing and respecting employees are the most prominent factors.

The fourth question on the interview form asked the participants about their job satisfaction in their current position. Half of the participants (P3-P6, P8) reported positive job satisfaction, using words such as "energetic," "happy," and "motivated." However, three participants (P1, P9, P10) expressed negative feelings about their job satisfaction, using words such as "anxious," "unhappy," and "stuck." The two remaining participants (P2, P7) expressed neutral and undecided feelings regarding job satisfaction. Participant P7 reported being satisfied with job security and their interactions with colleagues but expressed a desire for more challenging work. Participant P10 reported feelings of dissatisfaction and pressure in her job.

Table 3. *Organizational factors*

Factors	Participant
To be able to make decisions	P9
Evaluating employee ideas	P2, P8, P5, P4
Healthy communication between employees	P8
Support personal development	P8
Teammates	P6, P7
Valuing and respecting employees	P3, P5

In the fifth question on the interview form, the participants were asked their opinions on how the workload was distributed among the staff at their institution. Six participants (P1, P2, P4, P6, P7, P10) pointed out that the workload was not distributed equally. The participant (P10) expressed this inequality as follows: "*They give more work to those who do the job well or those who do not comply with assigned tasks." They give less work to those who are close to managers or lie to avoid getting the job.*" The other four participants (P3, P5, P8, P9) indicated that their organizations' workloads were evenly distributed. "*In our organization, the workload is evenly distributed*," was the participant's (P9)response.

In the sixth question of the interview form, the participants were asked what they think about the effects of working outside working hours on work/life balance. In response to this question, the participants at P1, P3, P6, P7, P9, and P10 stated that working outside working hours was negative. Participant P8 said, "*Actually, I like working outside working*

hours because it provides an opportunity for my personal development and career advancement. However, this is a problematic situation because it causes intense stress and fatigue, which can cause health problems. In addition, of course, we cannot ignore the fact that it negatively affects both family relationships and the social environment."

In the seventh question of the interview form, participants were asked about the compatibility between their assigned jobs and their abilities. Upon analyzing their responses, seven participants (P1-P3, P6, P8-P10) stated that their assigned jobs were compatible with their abilities, while three participants (P4, P5, P7) noted that they were performing tasks below their skill level. Participant (P4) responded as follows: "*The tasks assigned are not appropriate for my skills, and many tasks were below my skill level.*" The participant (P8) who believed it is compatible replied, "*I think it is appropriate. I have improved myself more with the training and certificates I received after my university education, and I work in this field*."

The eighth question on the interview form was designed to determine participants' opinions of the wage policies implemented in their organizations. Table 4 presents the participants' analyzed responses.

Table 4. *Participants' opinions on the Wage Policy*

Wage Policy	Participant
Reasonable wages	P5, P8, P10
Low wages	P1, P3, P6, P9
Unsure about wages	P2
Not committed to equity	P4, P7

According to Table 4, only three employees were satisfied with their salaries, while others reported receiving low salaries or did not comply with the principles of equality. For example, participant P1 stated, "I think this is wrong. There can be a difference in salary between people with the same title. There should be some standards." P7 expressed this disparity as follows: "*Our salaries are much lower than those of IT workers in the private sector*." Only one participant was uncertain about wages.

The ninth question that the participants asked is whether or not their opinions, suggestions, and ideas as employees are considered in their organizations.

Table 5. *Evaluation Status of the Participants' Ideas*

Evaluation Status of the Participants' Ideas	Participant
It was considered and evaluated	P8, P9
Sometimes considered and evaluated	P6, P7
It was not considered and evaluated.	P1, P2, P3, P4, P5, P10

Table 5 presents the results of examining the participants' responses to the questions. Six participants indicated that their opinions, suggestions, and ideas were disregarded by their organizations. One participant (P1) expressed this situation as follows: "*I do not think that my opinions, suggestions, and ideas are given much importance in our organization. Managers make decisions, and we implement it*."Two participants reported that their ideas were occasionally considered and evaluated. Only two participants believed their ideas, thoughts, and suggestions were valued and evaluated.

The tenth question on the interview form asked participants about their appreciation or reward systems in their organizations. P3, P6, P8, and P10 mentioned receiving small gifts as a reward and expressing their appreciation verbally. The remaining six participants reported not having such a system. For example, P1 participant said, "*Unfortunately, there is no such system. This situation affects employees very negatively. Because people who strive more than other people in the organization and achieve something are discouraged when they do not receive any appreciation or reward and do not work as work-oriented as before.*"

In the 11th question of the interview form, the participants were asked how communication in terms of information exchange and cooperation between employees is ensured in their organizations. The eight participants (P1-P6, P8, P9) responded that they preferred to use e-mail and messaging applications. Only two participants (P7, P10) answered differently, stating that face-to-face meetings were used to communicate.

The impact of the organization's fringe benefits and sociocultural events on job satisfaction was the 12th question on the interview form. When reviewing the participants' responses, eight individuals (P1-P5, P8-P10) had not provided any

feedback on this question because social and cultural activities were not available at their institutions. All participants expressed that fringe benefits contributed to their job satisfaction. P8 participant said, "*It increases our motivation; our loyalty to the organization increases even more. It also brings out the team spirit."* However, all participants also noted that fringe benefits were inadequate.

The 13th question on the interview form asked how participants obtained the necessary tools, supplies, and materials to support their work environment and tasks. Half of the participants reported having access to their work environment and essential tools, equipment, and materials through their managers or relevant administrative units. Three participants (P5-P7) were able to obtain them from the internet, while two participants (P2, P10) were able to access them through their own resources. Participant (P2) responded to this question as follows "*I would expect it to be easier. All the tools/supplies that we may need at the beginning of the work can be provided by default, and information on how to obtain them afterward can be given; in this case, the employees discover them themselves.*", while all other participants stated that they could access them through a relevant person or unit.

In the following question, the participants were asked to describe their working style, including whether they worked from an office, remotely, or in a hybrid setting. They were also asked to explain which working style would increase their job satisfaction. All of the participants, except for P10, reported that they work in an office setting. P5 and P9 mentioned that working remotely was more convenient. They also indicated that if they were allowed to work remotely, this would increase their job satisfaction. One participant (P8) responded to this question as follows, "*I work physically, and since I usually provide technical support, physical work seems suitable." Hybrid working models can also be applied, but they are not completely remote. I think I can work physically when necessary and in emergencies and provide remote support on other days.*"

The interview form's fifteenth question asked participants to describe the personnel selection and recruitment methods used in their institutions. Table 6 presents the results of the participants' responses.

Table 6. *Personnel Selection and Recruitment Process*

Methods	Participant
Reference checks	P2
Assessment tests	P9, P4
In-person interviewing	P2, P3, P6, P7, P8, P9, P10
Resume screening	P8
Application	P2, P3

Upon examining Table 6, it can be seen that interviews, assessment tests, and applications are the top three methods in the recruitment process.

Benefits of working as an IT employee in a public institution is the sixteenth question on the interview form. Table 7 summarizes the participants' responses regarding the benefits of working in public institutions.

Table 7. *Advantages of Working as an IT Personnel in a Public Institution*

Advantages	Participant
Fixed working hours	P1, P8
Learn administrative processes	P2, P4, P5
Job guarantee	P3, P7, P8
Satisfaction with serving the government	P4, P8, P9, P10
Easy to follow developing and changing technology	P6, P7
Participating in large-scale IT projects	P7, P10

Upon analyzing Table 7, it is evident that the participants prioritize their satisfaction with serving the state as their primary advantage. This is followed by job security and learning administrative processes in public institutions, among other advantages. Additionally, IT employees consider fixed working hours, the possibility to easily keep up with new technologies, and the opportunity to participate in large-scale projects as other advantages. One participant (P8) responded to this question as follows, "*There are many advantages. Our working conditions are good and generally stable. There are not many personnel changes; we continue with the people we are used to and with whom we have good communication. There is a job guarantee, and it also makes me happy to contribute to the state and public service.*"

The next question is: What are the disadvantages of working as an IT personnel in a public institution? Participants' responses were analyzed, and the results are summarized in Table 8.

Table 8. *The Disadvantages of Being an IT Employee in a Public Institution*

Disadvantages	Participant
Restricted workspace	P4, P9
Low wages	P1, P2, P3, P5, P8, P9, P10
No professional development support	P1, P5, P6, P7, P9, P10
Bureaucracy	P4, P8, P10
Long working hours	P10
Inflexible work style	P2
Limited fringe benefits	K9

Table 8 shows that the participants identified low wages and lack of support for professional development as the most significant drawbacks of being an IT employee in public institutions. Intense bureaucracy, limited workspace, long hours, inflexible work style, and low benefits were also mentioned as disadvantages. In this question, the P10 coded participant responded to the question as follows "*Low salary policy, not seeing the value you deserve, not being able to advance in your career, educational limitations, long working hours, intense bureaucracy and political pressure, not being agile.*"

Question 18 on the interview form asked participants about the aspects of working at the institution that satisfied them. Table 9 presents a summary of the participants' responses to the questions. When Table 9 is examined, the results indicate that the participants are content with their employment in a public institution and are pleased to contribute to government-related projects. Additionally, they appreciate their coworkers. To this question, P5 responded as follows, "*The fact that it is a state institution that directly follows technological developments and gives direction nationally*", while P2 responded as follows "*The chance to contribute to the projects to be developed as a country in the medium and long term, to transfer the technical skills I have acquired to the policy and strategy level and the chance to work together with people from various backgrounds*".

Table 9. *Participants' Satisfaction Aspects of their Institutions*

Satisfaction Aspects	Participant
Contributing to government projects	P2, P4, P10
Ability to use technical skills	P1, P2
Being a government institution and its prestige	P4, P5, P6, P9
Compliance and good communication skills	P7, P8
Work Climate	P9
Co-workers	P1, P2, P3

In the nineteenth question of the interview form, the participants were asked about aspects of working for public institutions that did not satisfy their needs. The responses were similar to the answers emphasized to the question about the disadvantages of working as an IT staff member in a public institution. The participants identified several issues, including workload distribution, inadequate physical conditions, communication problems, work disruptions, and lack of merit. Participant 4 was particularly dissatisfied with communication problems, being given a workload below their capacity and not being offered training opportunities. The primary concern identified by Participant 1 was the unequal distribution of workloads and the failure to consider employee preferences.

In the final question of the interview form, participants were asked whether they had any additional comments. Participant P8 suggested that private health insurance should be considered due to the risky nature of the occupation. Participant P10 recommended that organizations develop strategies to retain qualified IT personnel to prevent brain drain. The issue of IT personnel leaving the country because of insufficient salaries, long working hours, and inflexibility was highlighted. Action should be taken to address this problem. The remaining participants did not provide any additional comments and expressed their gratitude to the researchers for their efforts in gauging job satisfaction.

5. DISCUSSION AND CONCLUSION

This study reveals public IT employee job satisfaction. In line with this objective, interviews were conducted with 10 IT employees from 3 different public institutions in Ankara. Because of the analysis of the participants' experiences and opinions, it was seen that their qualifications were mostly compatible with their jobs. An analysis of the responses indicated that the organizations were unable to leverage the potential of their employees and that professional development opportunities were limited. An analysis of the attitudes of the participants toward their jobs revealed that most participants demonstrated a high level of commitment and care about their work. However, after facing some problems, they stated that their commitment gradually decreased and they became insensitive to their jobs. Metin (2020) stated that public IT employees have a staff turnover problem, and the main reason for this problem is that they cannot satisfy their need for self-realization.

Another result of the research is that considering employee opinions, facilitating communication, and contributing to professional development increase IT employees' commitment to the organization, job satisfaction, and performance. In addition, it has been revealed that good communication, cooperation, and coordination positively affect job satisfaction, and in cases of incompatibility and lack of communication, teamwork should be encouraged and efforts should be made to achieve harmony. Other research results also show that job satisfaction positively affects employee productivity. It is also negatively related to the employee turnover rate (Egan, Yang, & Bartlet, 2004; Korkmaz & Erdogan, 2014; Silverthorne, 2004). Employees' job dissatisfaction can cause disasters for an organization, such as apathy toward their responsibilities, lateness to work, disengagement from their jobs, and decreased job performance, all of which negatively affect the daily functioning of the organization. These can result in employees leaving the organization, leading to high employee turnover (Gregory, 2011, Hee et al., 2018). This situation should be carefully evaluated for public institutions that have had difficulty employing IT employees for many years. It should be considered that IT employees' work styles, satisfaction, and commitment differ from other employees (Drucker, 2002). It should be kept in mind that employee satisfaction and satisfaction levels with their jobs depend on various factors, and working environments should be improved. In particular, the hierarchical organizational structure in public institutions does not meet the needs of IT employees.

It has been observed that workload distribution is generally not fair and balanced, work is assigned based on performance, and employees need equal opportunities and a fair working environment. In addition, this balance needs to be established because working outside of working hours will cause problems if it is continuous, negatively affect the work-life balance, and reduce the time allocated to their private lives. On the other hand, flexible working hours that can be implemented can positively affect the satisfaction and performance of IT employees.

The study revealed that public IT employees are generally paid low wages with no fair wage mechanism in place. The additional research findings indicate that salaries are a significant factor that affects job satisfaction (Hee et al., 2018; Utriainen & Kyngäs, 2009). It would be beneficial to review public institutions' salary policies for IT employees. At the very least, they should be equalized with private sector IT personnel who perform the same job functions. This can help reduce the probability of public institutions' IT employee turnover rates. Furthermore, a suitable salary policy based on performance-based evaluations can be implemented.

The public sector is an essential pillar of any functioning society. However, it faces challenges despite providing services aimed at improving people's lives. reputation problems among IT employees, communication difficulties, and institutional resistance can alienate IT workers from public institutions. Managers and decision makers in public organizations must understand that IT workers have different needs than other employees. Managers should simplify

IT workers' bureaucracies and provide support to eliminate barriers between functional areas. Transforming the work style to a digital-first system can motivate IT workers to work in public institutions. In addition, IT workers can work seamlessly between the workplace and home with consistent and secure network access. The physical presence of IT workers in the workplace does not necessarily guarantee productivity. Therefore, hybrid working opportunities should be provided for IT workers.

Researchers argue that job satisfaction is essential for employee motivation (Rao, 2005). Employee motivation and performance are even more important, especially considering that digital transformation has affected the entire world and organizations. Improvement of working environments, communication, and the creation of motivational and encouraging environments are at the forefront. Thus, public IT employees' performance, motivation, and job satisfaction can increase. Job satisfaction and service quality can be enhanced by workers who are satisfied with their work. In this regard, managers and policymakers at public institutions should focus on providing various opportunities to their employees to keep them satisfied (Kabir & Parvin, 2011). Employees should be provided advancement opportunities, such as salary increases and participation in policymaking, to improve job satisfaction and performance. In addition, security and good relationships with managers and co-workers are important for job satisfaction.

This study has some limitations. Interviews were conducted with 10 IT employees. In future studies, a questionnaire can be applied to public IT employees using a quantitative research method on the same subject, and interviews can be conducted with IT employees from different cities. In addition, the effect of demographic variables such as gender, age, years of service, and educational status on the satisfaction of public IT employees can be investigated.

Ethics Committee Approval: Ethics committee approval was received for this study from the Ankara Yıldırım Beyazıt University Social and Human Sciences Ethics Committee (Date: 22.11.2023, Decision no. 09-215).

Informed Consent: Written informed consent was obtained from all participants who participated in this study.

Peer Review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- V.A., H.G.; Data Acquisition- H.G.; Data Analysis/Interpretation- V.A., H.G.; Drafting Manuscript- V.A., H.G.; Critical Revision of Manuscript- V.A.; Final Approval and Accountability- V.A.; Technical or Material Support- H.G.; Supervision- V.A.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

ORCID IDs of the authors

Vildan ATEŞ 0000-0002-8855-8556
Hatice GÜNAY 0000-0003-4590-6888

REFERENCES

Adams, G. A., King, L.A. and King, D. W. (1996). Relationships of job and family involvement, family social support, and work–family conflict with job and life satisfaction. *Journal of Applied Psychology, 81*(4), 411–420.

Alegre, I., Mas-Machuca, M., and Berbegal-Mirabent, J. (2016). Antecedents of employee job satisfaction: Do they matter? *Journal of Business Research, 69*(4), 1390-1395.

Allen, D. G., Shore, L. M., & Griffeth, R. W. (2003). The role of perceived organizational support and supportive human resource practices in the turnover process. *Journal of Management, 29*(1), 99–118.

Al Maqbali, M. A. (2015). Factors that influence nurses' job satisfaction: a literature review. *Nursing Management, 22*(2).

Asgarova, A. (2019). Impact of employee training on employee job satisfaction and achievement: a case of Turkish manufacturing industry. *Available at SSRN* 3398117.

Ayranci, E. (2011). A study on the factors of job satisfaction among owners of small and medium-sized Turkish businesses. *International Journal of Business and Social Science, 2*(5), 87-100.

Aytac, T. (2020). The effects of working in public or private schools on job satisfaction of teachers in Turkey: A meta-analysis study. *Eurasian Journal of Educational Research, 20*(89), 179-200.

Basar, U. (2011). Örgütsel Adalet Algısı, Örgütsel Özdeşleşme ve İş Tatmini Arasındaki İlişkilere Yönelik Görgül Bir Araştırma [An empirical research investigating relationships between organizational justice perception, organizational identification and job satisfaction] (Masters Thesis). Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp.

Baykara, Z. E., & Orhan, R. (2020). An Analysis of Job Satisfaction Levels of Physical Education Teachers in Ankara, Turkey. E*ducational Research and Reviews, 15*(2), 65-71.

Bhardwaj, A., Mishra, S., & Jain, T. K. (2021). An analysis to understanding the job satisfaction of employees in the banking industry. *Materials Today: Proceedings, 37*, 170-174.

Buker, H. and Dolu, O. (2010). Police job satisfaction in Turkey: Effects of demographic, organizational and jurisdictional factors. *International Journal of Comparative and Applied Criminal Justice, 34*(1), 25-51.

Buyukgoze-Kavas, A., Duffy, R. D., Güneri, O. Y., and Autin, K. L. (2014). Job satisfaction among Turkish teachers: Exploring differences by school level. *Journal of Career Assessment, 22*(2), 261-273.

Çevrimli, M. B., Günlü, A., Mat, B., Akın, A. C., and Sipahi, C. A. (2020). A study on the validity and reliability of the level of employee satisfaction at livestock farming enterprises in Turkey. *Acta Vet Eurasia, 46*, 45-53.

Demiral, Ö. (2017). Effects of training on employee job satisfaction and achievement: Train to gain evidence from manufacturing businesses in Turkey. Business & Management Studies: *An International Journal, 5*(3), 765-785.

Demiralp, O. (2006). *Kırklareli ili milli eğitim müdürlüğünde çalışanların memnuniyet düzeyi The gladness rate of the workers who works for the manegement of national education in Kirklareli.* (Masters Thesis). Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp

Drucker, P. (2002), *Managing in the Next Society,* New York: St. Martin's Press.

Dorta-Afonso, D., Romero-Domínguez, L., & Benítez-Núñez, C. (2023). It's worth it. High-performance work systems for employee job satisfaction: The mediational role of burnout. *International Journal of Hospitality Management, 108*, 103364.

Edgar, F., & Geare, A. (2005). HRM practice and employee attitudes: Different measures—Different results. Personnel Review, 34(5), 534–549

Egan, T. M., Yang, B., & Bartlett, K. R. (2004). The effects of organizational learning culture and job satisfaction on motivation to transfer learning and turnover intention. *Human Resource Development Quarterly, 15*(3), 279–301.

Gider, Ö., Akdere, M., & Top, M. (2019). Organizational trust, employee commitment and job satisfaction in Turkish hospitals: implications for public policy and health. *Eastern Mediterranean Health Journal, 25*(9).

Habib, S., Aslam, S., Hussain, A., Yasmeen, S., and Ibrahim, M. (2014). The impact of organizational culture on job satisfaction, employee commitment, and turnover intention. *Advances in Economics and Business, 2*(6), 215-222.

Halil, Z., and Selim, Z. (2007, December). Measuring employee satisfaction in small and medium sized enterprises. In *Congress on knowledge, management and economy*, Istanbul, Turkey.

Hee, O. C., Yan, L. H., Rizal, A. M., Kowang, T. O., and Fei, G. C. (2018). Factors influencing employee job satisfaction: A conceptual analysis. *International Journal of Academic Research in Business and Social Sciences, 8*(6), 331-340.

Inayat, W. and Khan, M. (2021). A study of job satisfaction and its effect on the performance of employees working in private sector organizations, Peshawar. *Education Research International, 2021*(1), 1751495.

Indianafreenotes (2020). Job Satisfaction. https://indiafreenotes.com/job-satisfaction/

Inuwa, M. (2016). Job satisfaction and employee performance: An empirical approach. *The Millennium University Journal, 1*(1), 90-103.

Judge, T. A., Zhang, S.C. and Glerum, D. R. (2020). Job satisfaction. *Essentials of job attitudes and other workplace psychological constructs*, 207-241.

Kabir, M. N., and M. M. Parvin (2011). Factors affecting employee job satisfaction of the pharmaceutical sector. *Australian Journal of Business and Management Research, 1*(9), 113-123.

Kara, D., Uysal, M., & Magnini, V. P. (2012). Gender differences on job satisfaction of the five-star hotel employees: The case of the Turkish hotel industry. *International Journal of Contemporary Hospitality Management, 24*(7), 1047-1065.

Kapucu, S., Kose, T. K., Kurt, B., & Ersoy, N. A. (2017). Attitudes and satisfaction of nurses with the work environment in Turkey. *International Journal of Caring Sciences, 10*(2), 771-780.

Kong, H., Jiang, X., Chan, W., & Zhou, X. (2018). Job satisfaction research in the field of hospitality and tourism. *International Journal of Contemporary Hospitality Management, 30*(5), 2178-2194.

Korkmaz, O. and Erdogan, E. (2014). İş yaşam dengesinin örgütsel bağlılık ve çalışan memnuniyetine etkisi. *Ege Academic Review, 14*(4), 541-557.

Küskü, F. (2003). Employee satisfaction in higher education: the case of academic and administrative staff in Turkey. *Career Development International, 8*(7), 347-356.

Mansoor, M., Fida, S., Nasir, S., & Ahmad, Z. (2011). The impact of job stress on employee job satisfaction: a study on the telecommunication sector of Pakistan. *Journal of Business Studies Quarterly, 2*(3), 50.

Metin, F. (2023). Retention of public IT professionals: The mediating roles of perceived organizational support and psychological empowerment. *Sosyoekonomi, 31*(57), 101-129.

Metin, F. (2020). Kamuda bilişim uzmanı personel devir hızı sorunu ve çözüm önerileri: TÜİK vaka çalışması [Problem Of Turnover Rate Of IT Experts In Public Sector And Some Solutions: TUIK Case Study]. *Hacettepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 38*(2), 243-266.

Miner (1992). *Industrial Organizational Psychology*. McGraw-Hill, Inc. Singapore.

Ozturk, A. B., & Hancer, M. (2011). The effect of demographics on job satisfaction: a study of hotel managers in Turkey. *International Journal of Hospitality & Tourism Administration, 12*(3), 189-201.

Pandey, P. and Asthana, P. K. (2017). An empirical study of factors influencing job satisfaction. *Indian Journal of Commerce and Management Studies, 8*(3), 96-105.

Patton, M. Q. (2003). Qualitative evaluation checklist. Evaluation checklists project, 21(1), 1-13.

Phuong, T. T. K. & Tran, T. V. (2020). Job satisfaction, employee loyalty and job performance in the hospitality industry: A moderated model. *Asian Economic and Financial Review, 10*(6), 698.

Riyanto, S., Endri, E., and Herlisha, N. (2021). Effect of work motivation and job satisfaction on employee performance: Mediating role of employee engagement. Problems and Perspectives in Management, 19(3), 162.

Sahin, L. (2015). Çalışan memnuniyetinin sağlanmasında temel dinamikler: dışsal faktörlerin çalışanların memnuniyet algıları üzerindeki etkilerine yönelik bir alan araştırması. *ISGUC The Journal of Industrial Relations and Human Resources, 17*(2), 28-44.

Silverthorne, C. (2004). The impact of organizational culture and person-organization fit on organizational commitment and job satisfaction in Taiwan. *Leadership and Organization Development Journal, 25*(7), 592–599.

Susanto, P. C., S. Syailendra, and R. F. (2023). Determination of motivation and performance: analysis of job satisfaction, employee engagement and leadership. *International Journal of Business and Applied Economics, 2*(2), 59-68.

Tang, S., Siu, O., & Cheung, F. (2014). A study of work-family enrichment among Chinese employees: The mediating role of work support and job satisfaction. Applied Psychology, 63(1), 130–150.

Tepeci, M. (2005). The dimensions and impacts of organizational culture on employee job satisfaction and intent to remain in the hospitality and tourism industry in Turkey. *Journal of Travel and Tourism Research, 5*(1/2), 21-39.

Turkyilmaz, A., Akman, G., Ozkan, C., and Pastuszak, Z. (2011). Empirical study of public sector employee loyalty and satisfaction. *Industrial Management and Data Systems, 111*(5), 675-696.

Toker, B. (2011). Job satisfaction of academic staff: an empirical study on Turkey. *Quality Assurance in Education, 19*(2), 156-169.

Utriainen, K., and H. Kyngäs (2009). Hospital nurses' job satisfaction: a literature review. *Journal of Nursing Management, 17*(8), 1002-1010.

Waqas, A., U. Bashir, M. F. Sattar, H. M. Abdullah, I. Hussain, W. Anjum, and R. Arshad (2014). Factors influencing job satisfaction and its impact on job loyalty. *International Journal of Learning and Development, 4*(2), 141-161.

How cite this article

Ateş, V., & Günay, H. (2024). A Qualitative Study on Job Satisfaction of Public Information Workers. *Acta Infologica, 8*(2), 118-132. https://doi.org/10.26650/acin.1422463

İSTANBUL UNIVERSITY PRESS

Acta Infologica
acin 2024, 8 (2), 133–156
DOI: 10.26650/acin.1447456

Research Article

Turkish Lira Banknote Classification using Transfer Learning and Deep Learning

Mirsat Yeşiltepe[1], Harun Elkıran[2], Jawad Rasheed[2]

[1]Yıldız Technical University, Department of Mathematical Engineering, İstanbul, Türkiye
[2]İstanbul Sabahattin Zaim University, Department of Computer Engineering, İstanbul, Türkiye

Corresponding author : Mirsat Yeşiltepe
E-mail : mirsaty@yildiz.edu.tr

ABSTRACT

With the increasing exchange of foreign currencies due to globalization, there is a need for systems that can recognize and validate multiple currencies in real time. Such systems facilitate smooth international transactions and support the finance sector in dealing with diverse currencies. This study focuses on classifying Turkish banknotes using deep learning models. The dataset comprises 6901 images of six different denominations (5 TL, 10 TL, 20 TL, 50 TL, 100 TL, and 200 TL) under various conditions, such as flat, angled, curved, and bent. The proposed model implements pre-trained models, including VGG16, VGG19, DenseNet121, DenseNet169, DenseNet201, MobileNet, and MobileNetV2, to classify the images. Different image sizes (50x50, 100x100, 150x150, and 200x200) and optimizers (SGD, RMSprop, Adam, Adamax, etc.) were tested to determine the most effective combinations. The best result was achieved with DenseNet201 with an image size of 200 and the SGD optimizer, achieving an accuracy of 98.84% in 12 epochs. Smaller image sizes (50x50) resulted in reduced performance for all models. In addition, models such as DenseNet169 and DenseNet121 also demonstrated high performance; however, MobileNetV2 struggled with smaller images.

Keywords: DenseNet201, Optimizer, Banknote, Convolution, Accuracy.

Submitted : 06.03.2024
Revision Requested : 26.08.2024
Last Revision Received : 17.10.2024
Accepted : 18.10.2024
Published Online : 03.12 .2024

1. INTRODUCTION

Banknote classification systems are increasingly used in applications such as self-service kiosks, vending machines, and public transportation systems. With the increasing exchange of foreign currencies due to globalization, the rise in counterfeit banknotes poses a significant threat to financial systems worldwide. Moreover, many developing countries have limited access to reliable financial transaction infrastructure. Automated banknote classification systems, especially when designed to work on mobile platforms, provide affordable and accessible solutions for these regions and promote financial inclusion.

The classification and detection of banknotes are essential tasks in the financial and retail sectors, particularly as global currency circulation and automated cash-handling processes increase in complexity. With the increasing threat of counterfeit banknotes infiltrating markets, developing advanced systems to accurately classify and authenticate currency has become a pressing need. Traditionally, manual banknote inspection has been used; however, it is prone to human error and is inefficient in environments that demand high-speed processing. Automated banknote classification driven by recent advances in artificial intelligence (AI) and machine learning offers robust solutions for real-time and large-scale cash-handling applications.

Deep learning, a subset of machine learning, has gained attention as a powerful tool for image classification tasks, including banknote recognition. Convolutional Neural Networks (CNNs), in particular, have proven effective in learning complex features from banknote images, such as intricate patterns, watermarks, and holograms, that distinguish real currency from counterfeit. The ability of deep learning models to automatically extract hierarchical features from raw data makes them well-suited for handling the variations and challenges in currency classification. Transfer learning, where pre-trained models are fine-tuned for banknote classification tasks, further enhances system performance, particularly when limited datasets are available.

The need for accurate and efficient banknote classification is underscored by its diverse applications in automated teller machines (ATMs), vending machines, and cashier systems, as well as in global foreign exchange operations. Furthermore, counterfeit detection is of paramount importance because the proliferation of fake currency can disrupt economies and undermine trust in financial systems. By leveraging deep learning techniques, automated systems can not only identify legitimate currency and provide a defense against fraud by providing real-time analysis in high-volume environments, such as banks, retail stores, and transportation hubs.

The classification of banknotes using machine learning and deep learning techniques has attracted considerable attention in recent years due to the increasing need for automated systems capable of distinguishing between real and counterfeit notes. Early research in this area primarily relied on traditional machine learning techniques, such as support vector machines (SVMs) and decision trees, which manually extracted features like edge detection, color histograms, and texture analysis for classification tasks. Although these methods provided moderate success, their reliance on manual feature engineering limited their scalability and accuracy when confronted with the complex patterns present in modern currency designs. These limitations have paved the way for more advanced techniques that leverage the power of deep learning models, particularly CNNs, to address the shortcomings of earlier approaches.

One of the key breakthroughs in this field was the application of neural networks to banknote recognition, as introduced by (Baek et al., 2018). Their study demonstrated that CNNs can effectively classify banknotes by automatically learning hierarchical features directly from raw images without requiring manual feature extraction. This study demonstrated that deep learning models can outperform traditional methods in terms of both accuracy and efficiency. The authors employed a multilayer CNN architecture that captured spatial hierarchies, making it highly suitable for identifying intricate patterns on banknotes, such as watermarks, holographic images, and fine textures. This approach provides a foundation for the subsequent adoption of deep learning models in the field of banknote classification and counterfeit detection.

Subsequently, other researchers explored the use of transfer learning to enhance the performance of banknote classification systems, particularly when limited data were available. (Prakash et al., 2023) introduced a deep learning approach combining CNNs for image analysis and recurrent neural networks (RNNs) for security feature assessment in counterfeit banknote detection. Testing on a dataset of real and fake notes, the ensemble model shows superior accuracy of 98.36% and precision of 96.8% compared to traditional methods. The transfer learning approach enables the system to generalize well to unseen data, thereby making it more robust and scalable across different currencies and denominations

In addition, multispectral imaging combined with deep learning has been explored as a way to further improve counterfeit detection. (Wang et al., 2022) presented an automated approach using optical coherence tomography (OCT) and machine learning to classify counterfeit banknotes by analyzing internal features. By training classifiers on OCT

image-derived features, the study achieved high accuracy in detecting and categorizing counterfeits, with the support vector machine model reaching a sensitivity of 96.55% and specificity of 98.85%. Their research underscored the potential of integrating multispectral data with CNNs, expanding the scope of banknote classification systems to better handle sophisticated counterfeiting techniques that rely on non-visible features

In addition to CNNs and multispectral imaging, other studies have focused on lightweight architectures for real-time banknote classification in mobile and embedded systems. (Linkon et al., 2020) evaluated lightweight CNN models with transfer learning for Bangladeshi banknote classification using ResNet152v2, MobileNet, and NASNetMobile on two datasets with 8000 and 1970 images, respectively. The experiment shows that MobileNet achieved 98.88% on the larger dataset, NASNetMobile reached 100% on the smaller dataset, and MobileNet achieved 97.77% on the combined dataset. This development has broadened the applicability of automated banknote classification, making it accessible in areas where high-end hardware is unavailable.

Several hybrid approaches that combine image processing techniques with machine learning models have also been proposed for counterfeit detection. (Pachón, Ballesteros, and Renza, 2023) presented a pruning technique for sequential CNNs that reshaped network layers and was tested on models including AlexNet, VGG11, and VGG16, using a dataset of Colombian peso banknotes. The pruned models achieved up to a 75% reduction in parameters and computational load (FLOPs) with minimal accuracy loss, while models with higher pruning rates (up to 95%) showed more significant accuracy drops, especially for AlexNet and VGG16. Although more computationally intensive, these hybrid approaches show promise in handling difficult cases where counterfeiters use advanced printing techniques.

Similarly, much research has been conducted on the classification and recognition of banknotes from different countries. Authors in (Galeana Pérez and Bayro Corrochano, 2018) proposed a scheme to classify Mexican and Euro banknotes. In contrast, some work on Indian banknotes has also been done, such as (Mittal and Mittal, 2018) proposed a CNN-based deep learning model to classify Indian banknotes according to various states and positions. Their work achieved an accuracy of 0.966. In (Veeramsetty, Singal, and Badal, 2020), the authors tested various transfer learning models with several parameters to achieve an accuracy of 84.4

Research into automated banknote classification helps develop more sophisticated detection techniques that are vital for preventing fraud, maintaining currency integrity, and ensuring trust in financial systems. Researchers worldwide have implemented various automated banknote recognition and classification systems that reduce the need for manual intervention in industries that handle cash, such as retail, banking, and public transportation. Thus, such an automatic system can lower operational costs, streamlines workflows, and ensure high-speed processing of cash transactions. However, few studies on Turkish banknote classification have been reported. For example, Researchers in (Baykal et al., 2018) exploited a dataset, created using images of banknotes used in Turkey obtained under different conditions, and the relationship between classes was examined in terms of color and features. They used a pretrained DenseNet121 and other models to classify Turkish banknotes to achieve an accuracy of 93.15%. In another study (Khashman, Ahmed, and Mammadli, 2019), the authors focused on classifying countries based on banknotes used in different countries, including those used in Turkey. Their aim was to classify countries, not banknotes. The banknotes in the dataset's images are flat; thus, different shapes of banknotes were not tested. Similar studies like (Khashman and Sekeroglu, 2005) and (Khashman, Sekeroglu, and Dimililer, 2005) used Turkish notes to classify deformation rates of Turkish and Euro banknotes using artificial neural networks. However, the banknotes used in these two studies are no longer used in Turkey. In (İyikesici and Erçelebi, 2023), researchers exploited deep learning techniques to identify and detect counterfeit Turkish banknotes. However, they considered limited classes, including 5- and 10-unit Turkish banknotes.

These studies provide a comprehensive overview of the advances in banknote classification using deep learning and related techniques. The transition from traditional machine learning models to deep learning, along with innovations in transfer learning, multispectral imaging, and lightweight architectures, has significantly enhanced the accuracy and scalability of banknote classification systems. These studies collectively emphasize the importance of continuously improving automated systems to handle the evolving complexities of currency fraud and classification, thereby ensuring financial security and operational efficiency across industries.

Existing studies on banknote classification and counterfeit detection using deep learning have several shortcomings. First, less work is required to classify new Turkish banknotes. Second, most research does not focus on all units of banknotes (such as banknotes of 5 Turkish Lira (TL) and 10 TL), limiting their scalability to multi-unit scenarios, which are crucial in global financial systems. Third, prior studies considered a limited amount of data due to data availability and scarcity. Fourth, although some efforts have been made to develop lightweight models for mobile and embedded platforms, the real-time performance of such models, especially in low-resource environments, remains underexplored. In addition, counterfeit detection techniques primarily focus on visible features, often failing to detect more sophisticated counterfeits that use non-visible features such as ultraviolet ink and watermarks. Moreover, there are

limited emphasis on making these deep learning models interpretable and explainable, which is essential for regulatory compliance and trust in financial applications. These gaps suggest further research to address the complexity and scalability of banknote classification systems.

This study explores the application of deep learning models, such as CNNs and transfer learning, to develop a robust system for banknote classification. This study investigates various architectures and optimization techniques to enhance the accuracy and efficiency of banknote recognition systems. Ultimately, the goal is to contribute to advancing secure, scalable, and automated solutions that can be deployed across various industries to ensure operational efficiency, financial security, and improved customer experiences when handling physical currency. This study makes the following contributions:

- Merges two existing datasets to create a comprehensive dataset of Turkish banknotes, totaling 6901 images. These images represent banknotes in various conditions (flat, angled, curved) across six denominations (5 TL, 10 TL, 20 TL, 50 TL, 100 TL, and 200 TL), providing diversity in angles, sides, and shapes.
- We tested various pretrained deep learning models for banknote classification, including VGG16, VGG19, DenseNet121, DenseNet169, DenseNet201, MobileNet, and MobileNetV2.
- Examined the impact of different optimizers (SGD, RMSprop, Adam, etc.) and image sizes (50×50, 100×100, 150×150, and 200×200) to identify the most efficient combinations for accurate classification.
- In this study, the performance of different deep learning models in classifying Turkish banknotes is analyzed, and the most effective model is determined.
- By analyzing the confusion patterns between different banknote values, the learning process of the model and potential improvement areas can be evaluated.
- The training efficiency and overlearning problem will be addressed, and the necessary measures will be determined to optimize the performance of the model.
- Provides detailed analysis of how different combinations of model architectures, optimizers, and image sizes influence classification accuracy. It was found that no single parameter alone could guarantee high efficiency, and testing different parameter combinations was crucial for high performance.

The remainder of this paper is divided as follows: Section 2 describes the dataset and outlines the methods and models. Section 3 presents the experimental results, and Section 4 analyzes the results. Section 5 concludes the study.

2. DATA AND METHODOLOGY

With the increasing exchange of foreign currencies due to globalization, systems that can recognize multiple denominations of currency in real time are required. Such systems facilitate smooth international transactions and support the finance sector in dealing with diverse currencies. Thus, this study exploited various state-of-the-art deep learning models to classify Turkish banknotes. Figure 1 shows the proposed scheme in this study. The information from two distinct data sources is combined to create the dataset. In addition to separating the data into training and test sets, the system performs preprocessing tasks, such as resizing images and other similar tasks. Then, it trains seven technologically advanced models, including VGG16, VGG19, DenseNet121, DenseNet169, DenseNet201, MobileNet, and MobileNetV2, in a separate manner. The trained models were then validated with a test set to determine whether the proposed scheme is valid.

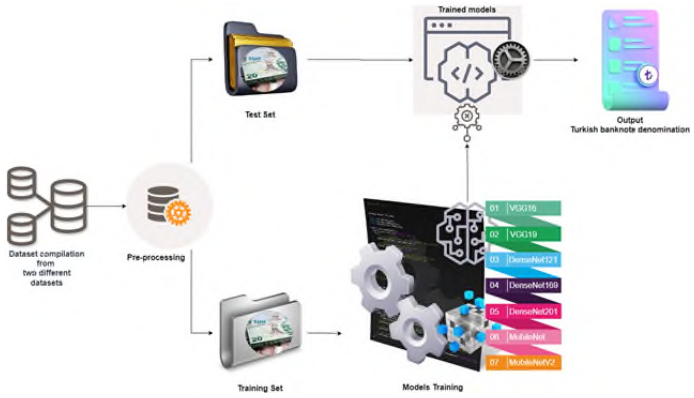


Figure 1. The workflow of the proposed system. The dataset was created by merging data from two different sources. The system performs pre-processing, such as image resizing, etc., and splits the data into training and test sets. Then, seven different customized state-of-the-art models are trained separately. The trained models were then tested on a test set to evaluate the validity of the proposed scheme.

2.1. Dataset

The dataset used in this study was created by merging two existing datasets from previous studies, namely (Baltaci, 2020) and (Sahin, 2018). resulting in a comprehensive collection of Turkish banknotes. The final dataset comprised a total of 6,901 images representing six Turkish Lira denominations: 5, 10, 20, 50, 100, and 200 TL. Each class is represented by a substantial number of images, providing a well-balanced dataset for training and testing deep learning models. The images feature banknotes under various conditions, including flat, angled, curved, and bent conditions, providing a challenging and realistic dataset that simulates real-world conditions in banknote handling. Figure 2 shows sample images of banknotes under various conditions.

Figure 2. The sample banknote images with various conditions (angle, curved, bent, and flat) were taken in different brightness environments. Turkish Banknotes are represented by 6 classes: 5, 10, 20, 50, 100, and 200 TL.

The diversity of the dataset is a significant strength because it includes images captured from different sides, angles, and orientations of the banknotes. This ensures that deep learning models trained on these data can generalize well to banknotes that may be presented under non-ideal conditions. The variations in the dataset—such as bends or curved notes—mimic how banknotes appear in everyday transactions, which is critical for real-world applicability. In addition, the representation of multiple denominations in the dataset allows for classification across a range of values, ensuring that models can distinguish between banknotes of different monetary values.

Moreover, the proposed dataset provides a valuable contribution to the field of currency classification, especially with the inclusion of challenging conditions such as different angles and bent banknotes. Such features increase the dataset's complexity, which makes it ideal for training deep learning models to be robust to variations in banknote presentation. This dataset could be useful for not only this particular study but also future research in the areas of currency recognition, counterfeit detection, and financial automation systems, where variability in banknote appearance poses a critical challenge.

2.2. Methodology

VGG16: The VGG16 is a deep convolutional neural network model comprising 16 weight layers (13 convolutional and 3 fully connected). The architecture follows a simple, uniform design in which convolutional layers are stacked with small 3x3 kernels and ReLU activation functions. Max pooling is applied after groups of convolutional layers to reduce the spatial dimensions. The network ends with three fully-connected layers and a softmax output. Despite its simplicity, VGG16 has high computational and memory costs due to its numerous parameters.

VGG19: VGG19 is an extension of VGG16 with 19 weight layers (16 convolutional and 3 fully connected). Like VGG16, it uses small 3×3 filters and ReLU activations in each layer but contains more convolutional layers, increasing the depth. The additional layers improve the capacity to learn complex features but further increase the computational requirements. VGG19 retains the same design philosophy, with max pooling applied after blocks of convolutions and fully connected layers at the end.

DenseNet121: DenseNet121 is a densely connected convolutional network in which each layer is directly connected to each other in a feedforward manner. Instead of learning redundant feature maps, DenseNet concatenates feature maps from all preceding layers, which improves network efficiency and allows for better feature reuse. The architecture consists of dense blocks separated by transition layers that downslope feature maps using convolution and pooling. DenseNet121 has 121 layers, with relatively fewer parameters than traditional networks like VGG, making it more computationally efficient.

DenseNet169: DenseNet169 follows the same architectural design as DenseNet121 but has 169 layers, thereby offering deeper feature extraction. The dense connectivity pattern helps mitigate the vanishing gradient problem; thus, deep networks can be trained more easily. The increased depth allows DenseNet169 to capture more complex features while maintaining higher efficiency in terms of parameter count and computational cost compared to similarly deep networks.

DenseNet201: DenseNet201 further extends the DenseNet architecture to 201 layers, providing deeper feature learning capabilities. Like its counterparts, DenseNet201 employs dense connections and transition layers between dense blocks. The deeper architecture allows for more complex hierarchical feature extraction, making it suitable for tasks requiring fine-grained classification, although there is a trade-off in increased computational load compared to smaller DenseNet models.

MobileNet: MobileNet is designed for mobile and embedded applications, focusing on efficiency and speed. This method uses depthwise separable convolutions, where a standard convolution is factorized into a depthwise convolution followed by a pointwise convolution, drastically reducing the number of parameters and computational cost. The proposed structure allows MobileNet to maintain high accuracy while being computationally lightweight, which makes it ideal for low-power devices and real-time applications.

MobileNetV2: MobileNetV2 improves upon MobileNet by introducing inverted residuals and linear bottlenecks. The inverted residuals use a shortcut connection between thin bottleneck layers. In contrast, the linear bottleneck layer ensures that the activation function does not destroy significant information during the bottleneck stage. MobileNetV2 retains the use of depthwise separable convolutions for computational efficiency; however, it improves feature representation, especially for complex tasks, by reducing information loss via the linear bottleneck structure. This resulted in better performance on resource-limited devices.

3. WORKING ENVIRONMENT AND EXPERIMENTAL RESULTS

The working environment for this study was implemented using Python as the primary programing language, and the Keras library was employed for building and training the deep learning models. Keras, a high-level API, was selected for its ease of use and integration with TensorFlow, making it ideal for conducting experiments with various deep learning models. The hardware setup included a system with GPU acceleration to facilitate faster model training and testing, allowing for efficient experimentation across multiple models, optimizers, and image sizes (Keras Team, 2023a). The optimization functions are SGD, RMSprop, Adam, Adadelta, Adagrad, Adamax, Nadam, and Ftrl (Keras Team 2023b)(Filter 2022). The development and testing were conducted in a controlled environment to ensure the reproducibility of results and to measure each model's performance accurately.

The dataset used in the experiments was created by merging two existing datasets, which resulted in 6901 images of Turkish banknotes. These images were classified into six categories: 5, 10, 20, 50, 100, and 200 TL. The dataset was highly diverse and contained images of banknotes in various orientations and conditions, such as flat, angled, curved, and bent. The images were preprocessed by resizing them to four different sizes (50×50, 100×100, 150×150, and 200×200), which allowed the models to be trained and tested on varying image dimensions. The dataset was split into training and testing sets with ratios of 80% and 20%, respectively. Table 1 presents the dataset details. This preprocessing step ensured that the models could handle different image resolutions and identified the optimal size for maximum classification accuracy.

Table 1. *Dataset description, number of samples in dataset, and train-test split for each class label in dataset*

Class	Train set	Test set	Total
5 TL	920	230	1150
10 TL	920	230	1150
20 TL	920	231	1151
50 TL	920	230	1150
100 TL	920	230	1150
200 TL	920	230	1150
		Total	**6901**

The experiments used several pretrained deep learning models, including VGG16, VGG19, DenseNet121, DenseNet169, DenseNet201, MobileNet, and MobileNetV2. Each model was fine-tuned using various optimizers, such as SGD, RMSprop, Adam, Adadelta, Adagrad, Adamax, Nadam, and Ftrl. These optimizers were applied with their default parameter values to assess their impact on model performance. Additionally, different image sizes were tested to determine the optimal resolution for efficient model training. The classification task involved training CNNs with a combination of predefined and custom layers, including two hidden layers of 1024 neurons with ReLU activation and a final output layer with a SoftMax activation function that matches the number of banknote classes.

The training process was conducted for 100 epochs with an early stopping mechanism based on validation accuracy to avoid overfitting. If validation accuracy did not improve after a certain number of epochs, training was stopped to obtain the best model. The validation accuracy served as a control point, and only the best performing model was retained in each experiment. Various metrics can be used to measure the success of the classification process. This study considered accuracy as an evaluation measure, which is the proportion of the total data that is in the correct class. It is the most widely used metric, especially when the dataset is balanced (Foody, 2023). In addition, this study measured the confusion matrix against each model exploited. Accuracy was the primary evaluation metric, and it was monitored throughout the training process. Accuracy and loss values for both the training and validation datasets were plotted in each experiment to assess the performance of the models at each epoch.

DenseNet201 achieved the highest accuracy of 98.84% with an image size of 200×200 and the SGD optimizer, which was completed in 12 epochs. The reason for keeping the epoch value at 12 is that overlearning occurred at this value. The model's efficiency increased significantly after the third epoch and was maintained after the eighth epoch. The performance of DenseNet201 decreased slightly with smaller image sizes; however, it still performed well with an accuracy of 97.54% for both 150×150 and 100×100 image sizes using the Adamax optimizer. However, for the smallest image size of 50, the performance decreased to 91.82% with the Adam optimizer, indicating that the model benefits from larger image inputs that can extract more detailed features. Table 2 lists the performances of the DenseNet201 model when trained on varying image sizes and optimizers. Figure 3 shows the accuracy and loss curves of the DenseNet201 model, and Figure 4 shows the confusion matrix for each.

Table 2. *Accuracy performance analysis of DenseNet201 trained on varying input sizes and optimizers*

Image Size	Optimizer	Accuracy Rate	Epoch number
200×200	SGD	0.9884	12
150×150	Adamax	0.9754	16
100×100	Adamax	0.9754	14
50×50	Adam	0.9182	17

DenseNet169 also performed well, with its best accuracy of 98.62%, which was achieved using an image size of 200×200 and the SGD optimizer, completing in 17 epochs. With a slightly smaller image size of 150×150, it maintained strong performance, achieving 97.39% accuracy with the RMSprop optimizer. Like DenseNet201, its accuracy decreased with smaller image sizes, achieving 97.47% at 100×100 and 90.15% at 50×50, indicating that DenseNet169, like its deeper counterpart, benefits from larger images and performs better with optimizers like SGD and RMSprop. Table 3 lists the performances of the DenseNet169 model when trained on varying image sizes and optimizers. Figure 5 shows the accuracy and loss curves of the DenseNet169 model, and Figure 6 shows the confusion matrix for each.

Table 3. *Accuracy performance analysis of DenseNet169 trained on varying input sizes and optimizers*

Image Size	Optimizer	Accuracy Rate	Epoch number
200×200	SGD	0.9862	17
150×150	RMSprop	0.9739	19
100×100	Adamax	0.9747	11
50×50	RMSprop	0.9015	20

DenseNet121 reached its highest accuracy of 97.68% with an image size of 150×150 and the Adamax optimizer, which completed in 11 epochs. For an image size of 200×200, it performed slightly lower with 97.38% using the RMSprop optimizer; however, this performance was still strong. For smaller image sizes of 100×100 and 50×50, the model's accuracy decreased to 96.31% and 90.30%, respectively, confirming that, like other DenseNet models, DenseNet121 performs better with larger image sizes but can still maintain decent performance on mid-sized images. Table 4 lists the performances of the DenseNet121 model when trained on varying image sizes and optimizers. Figure 7 shows the accuracy and loss curves of the DenseNet121 model, and Figure 8 shows the confusion matrix for each.

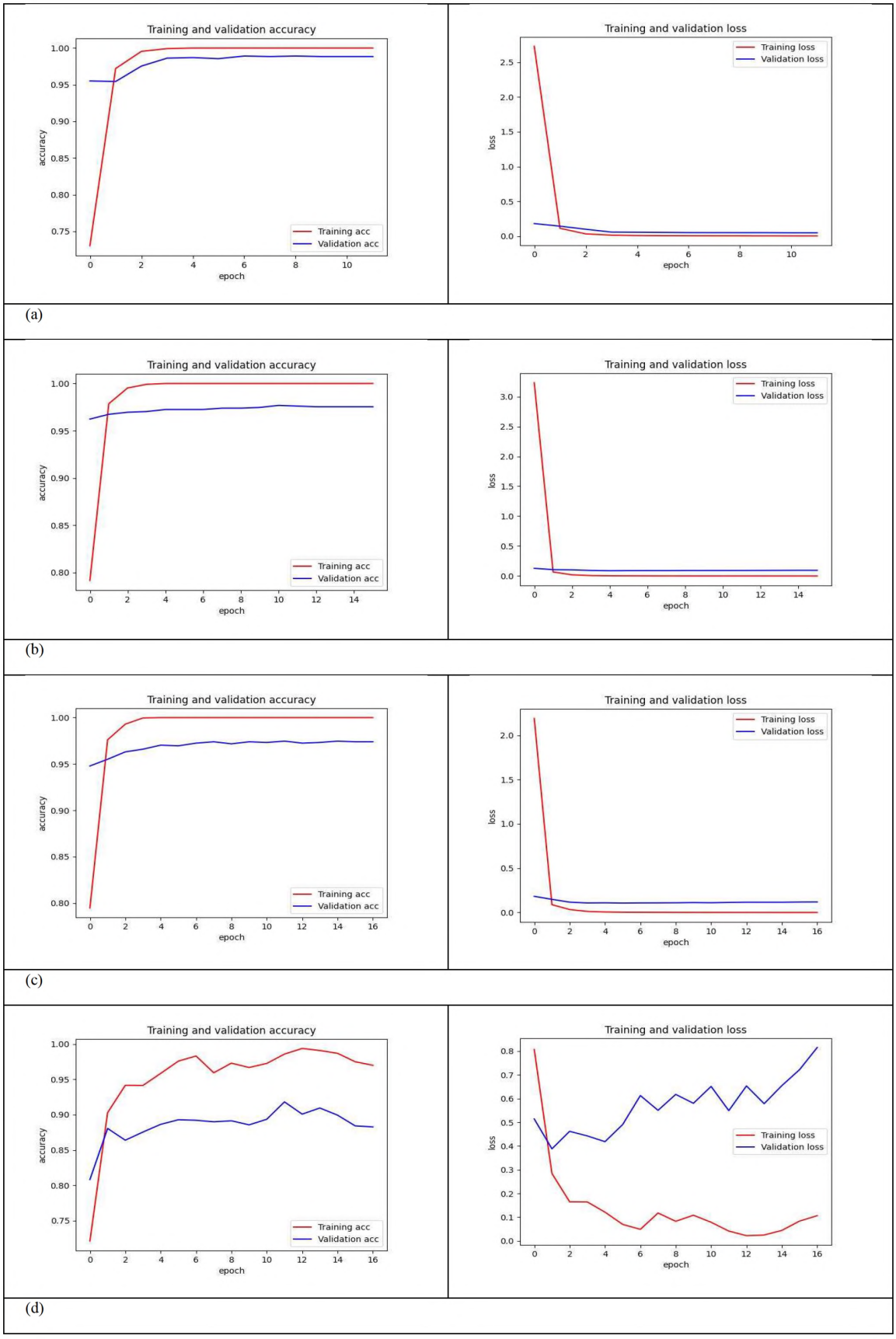


Figure 3. Accuracy and loss curves of DenseNet201 when trained over varying image sizes with several optimizers. Trained with (a) image size of 200×200 and SGD optimizer, (b) image size of 150×150 and Adamax optimizer, (c) image size of 100×100 and Adamax optimizer, (d) image size of 50×50 and Adam optimizer.

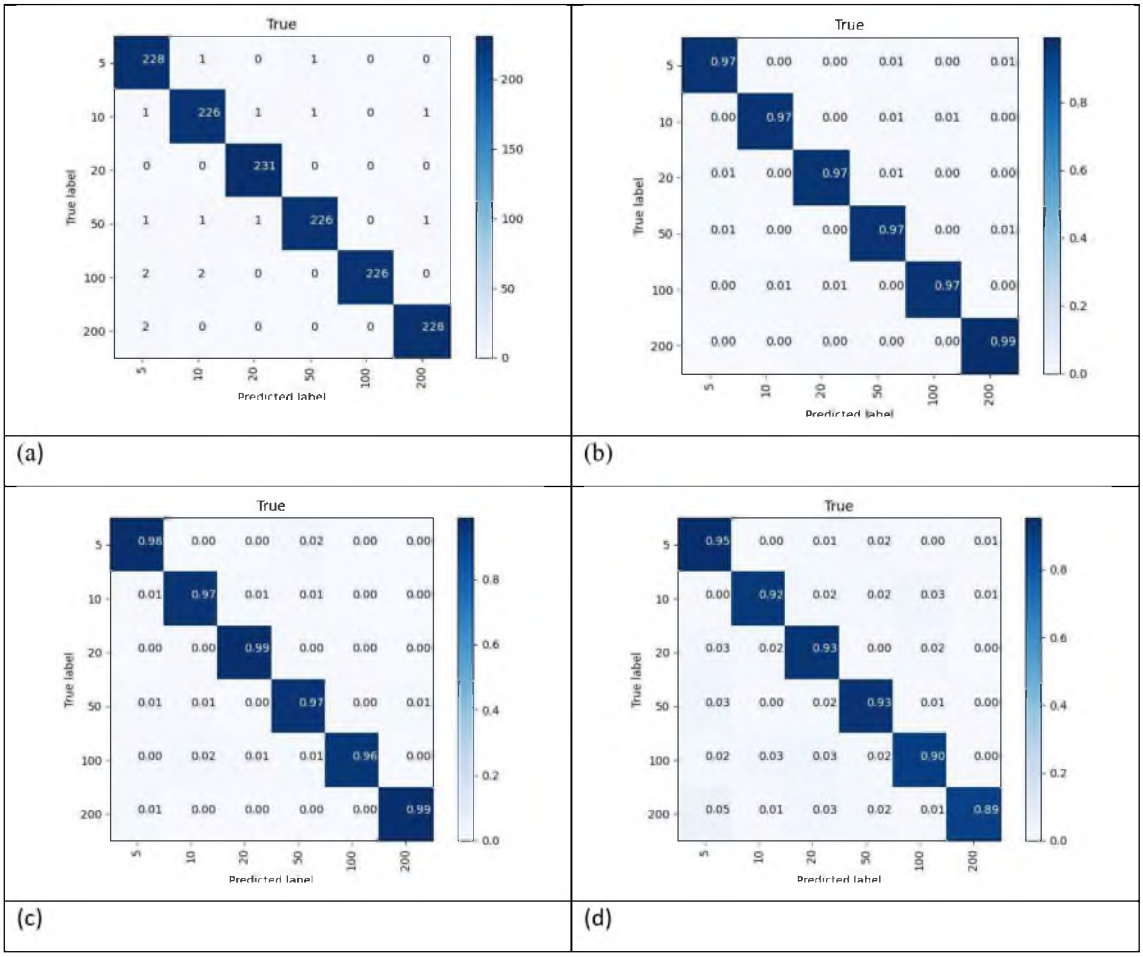


Figure 4. The confusion matrix of DenseNet201 when trained on images of various sizes with several optimizers. Trained with (a) image size of 200×200 and SGD optimizer, (b) image size of 150×150 and Adamax optimizer, (c) image size of 100×100 and Adamax optimizer, (d) image size of 50×50 and Adam optimizer.

Table 4. *Accuracy performance analysis of DenseNet121 trained on varying input sizes and optimizers*

Image Size	Optimizer	Accuracy Rate	Epoch number
200×200	RMSprop	0.9738	22
150×150	Adamax	0.9768	11
100×100	Adamax	0.9631	17
50×50	RMSprop	0.9030	23

MobileNet achieved its highest accuracy of 97.47% with an image size of 200×200 and the Adamax optimizer, completing in 20 epochs. For smaller image sizes of 150×150, 100×100, and 50×50, the model's performance gradually decreased to 96.81%, 95.51%, and a very low 19.55%, respectively, indicating that MobileNet struggles significantly with very small image sizes. MobileNet's performance suggests it is a good option for efficient classification with larger image sizes; however, it becomes less reliable when image resolution is reduced. While conducting the experiments, we also tried various other image sizes, such as 128×128 and 96×96. Table 5 lists the performance obtained by the MobileNet model when trained on varying image sizes and optimizers. Figure 9 shows the accuracy and loss curves of the MobileNet model, and Figure 10 shows the confusion matrix for each.

MobileNetV2 demonstrated moderate performance compared to the other models, with the highest accuracy of 95.51% at an image size of 150 using the Adamax optimizer. At the largest image size of 200, this value reached 95.37% with the Ftrl optimizer. However, with smaller image sizes of 100 and 50, MobileNetV2's performance decreased to 93.63% and 17.45%, respectively. This suggests that while MobileNetV2 can achieve decent performance with mid-sized images, it struggles significantly with negligible images, which is similar to MobileNet. Table 6 lists the performance obtained by the MobileNetV2 model when trained on varying image sizes and optimizers. Figure 11 shows the accuracy and loss curves for the MobileNetV2 model, and Figure 12 shows the confusion matrix for each.

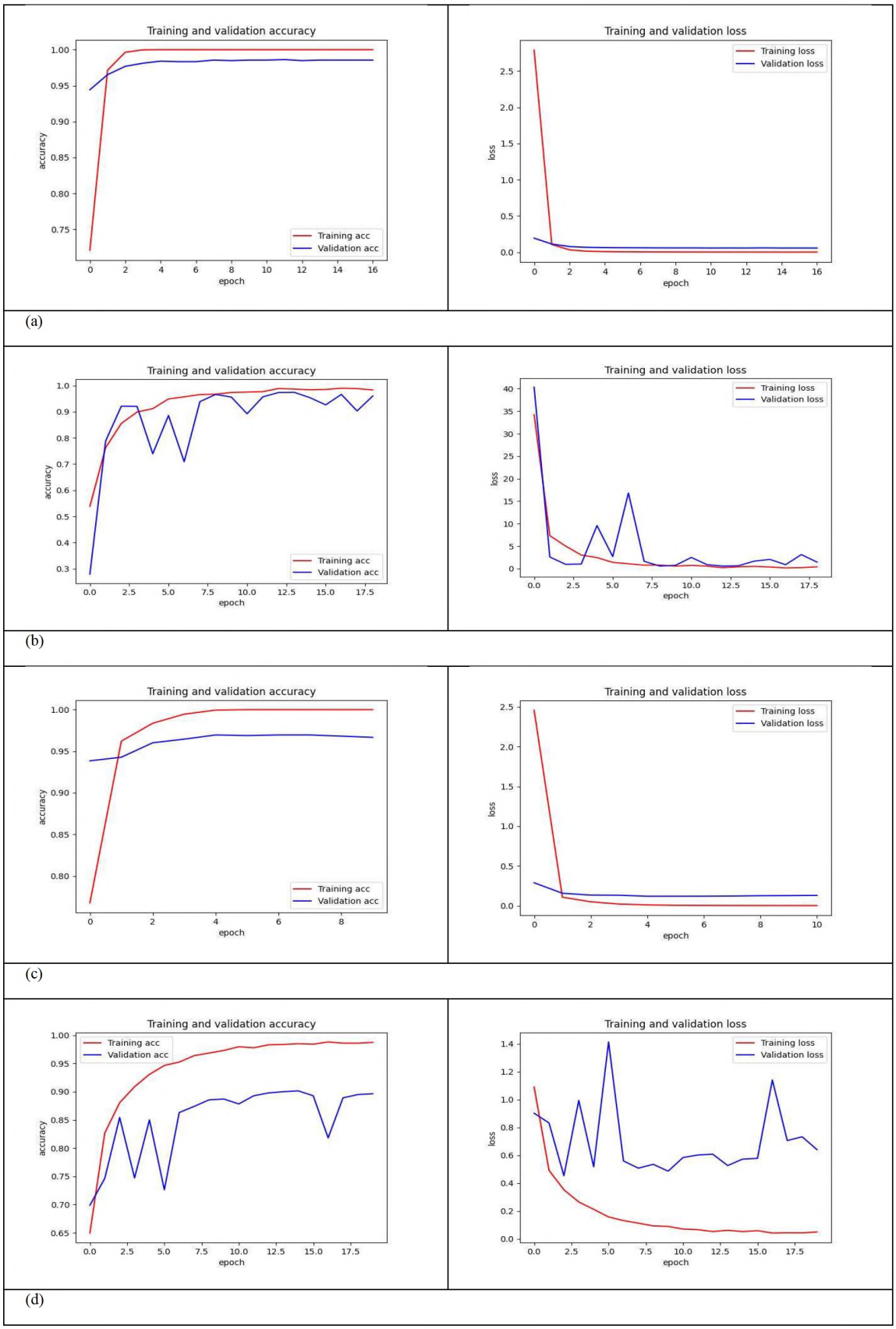


Figure 5. Accuracy and loss curves of DenseNet169 when trained over varying image sizes with several optimizers. Trained with (a) image size of 200×200 and SGD optimizer, (b) image size of 150×150 and RMSprop optimizer, (c) image size of 100×100 and Adamax optimizer, (d) image size of 50×50 and RMSprop optimizer.

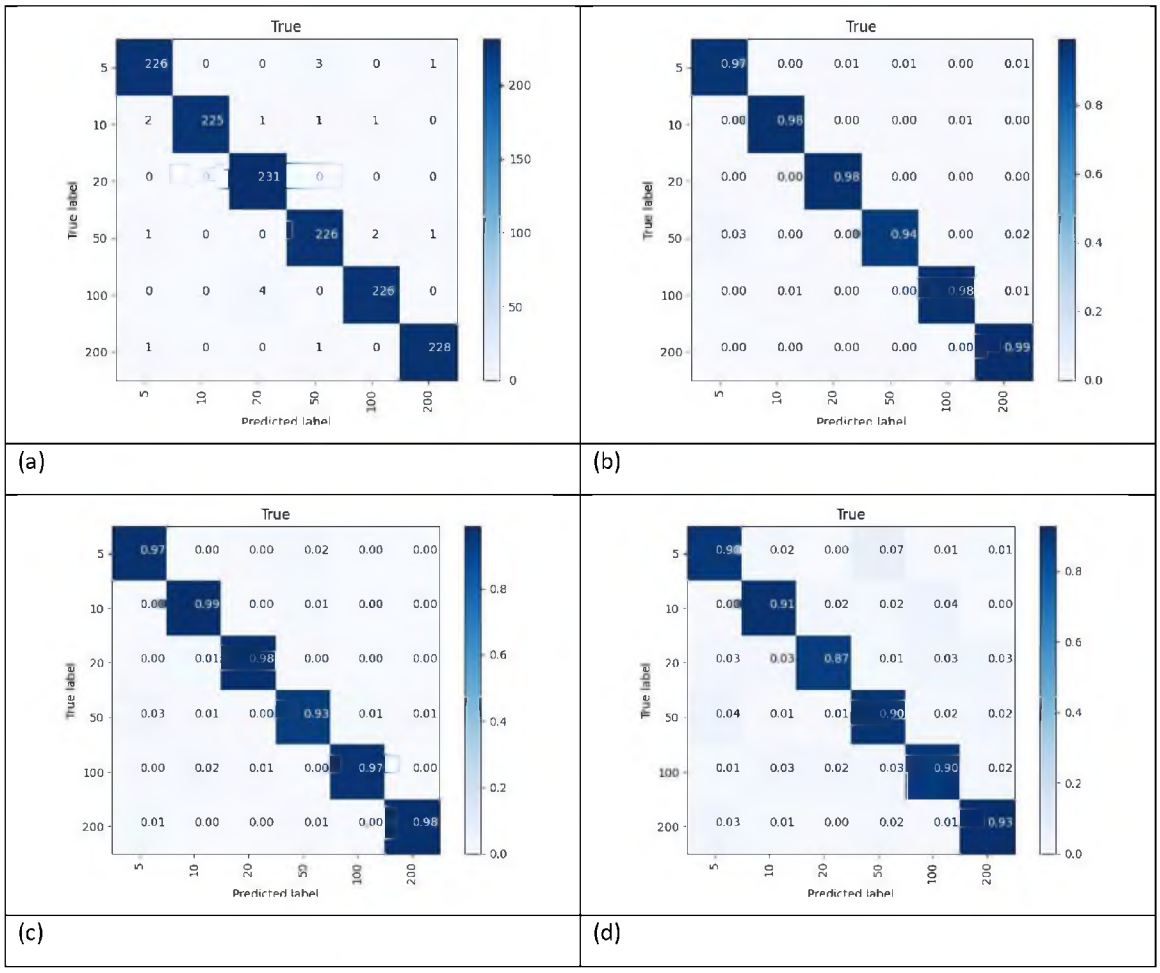


Figure 6. The confusion matrix of DenseNet169 when trained on images of various sizes with several optimizers. Trained with (a) image size of 200×200 and SGD optimizer, (b) image size of 150×150 and RMSprop optimizer, (c) image size of 100×100 and Adamax optimizer, (d) image size of 50×50 and RMSprop optimizer.

Table 5. *Accuracy performance analysis of MobileNet trained on varying input sizes and optimizers*

Image Size	Optimizer	Accuracy Rate	Epoch number
200×200	Adamax	0.9747	20
150×150	Adamax	0.9681	15
100×100	Adamax	0.9551	17
50×50	Adamax	0.1955	26

Table 6. *Accuracy performance analysis of MobileNetV2 trained on varying input sizes and optimizers*

Image Size	Optimizer	Accuracy Rate	Epoch number
200×200	Ftrl	0.9537	6
150×150	Adamax	0.9551	18
100×100	SGD	0.9363	22
50×50	Nadam	0.1745	15

VGG16 achieved the highest accuracy of 96.69% at an image size of 150×150 using the RMSprop optimizer. With an image size of 200×200, it performed similarly, achieving 96.38% with the same optimizer. For smaller image sizes of 100×100 and 50×50, the model accuracy decreased to 95.37% and 90.15%, respectively. VGG16 performed reasonably well across different image sizes, although its performance was slightly lower compared to the DenseNet models, indicating that it may not be as efficient in extracting intricate features from banknote images. Table 7 lists the

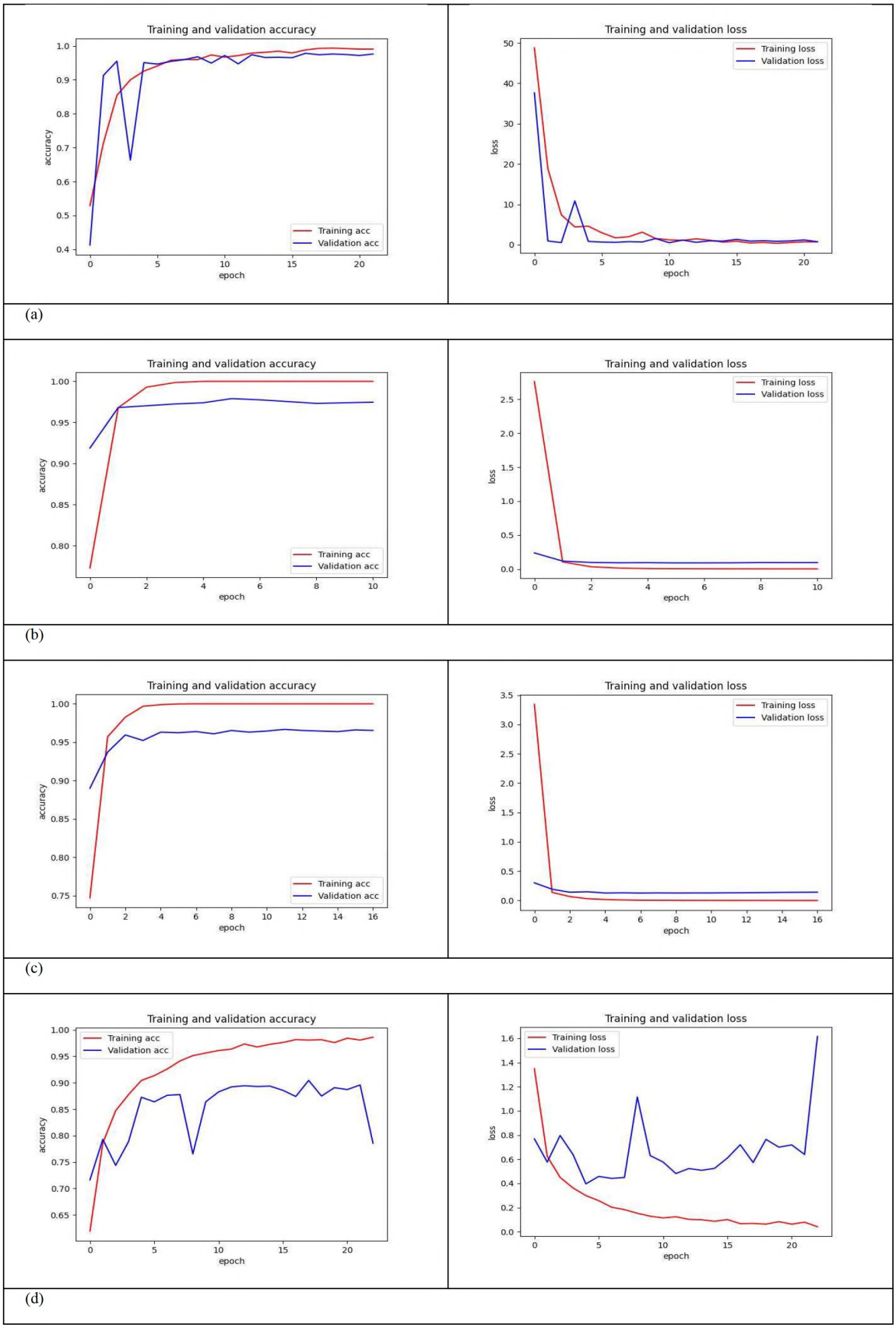


Figure 7. Accuracy and loss curves of DenseNet121 when trained over varying image sizes with several optimizers. Trained with (a) image size of 200×200 and RMSprop optimizer, (b) image size of 150×150 and Adamax optimizer, (c) image size of 100×100 and Adamax optimizer, (d) image size of 50×50 and RMSprop optimizer.

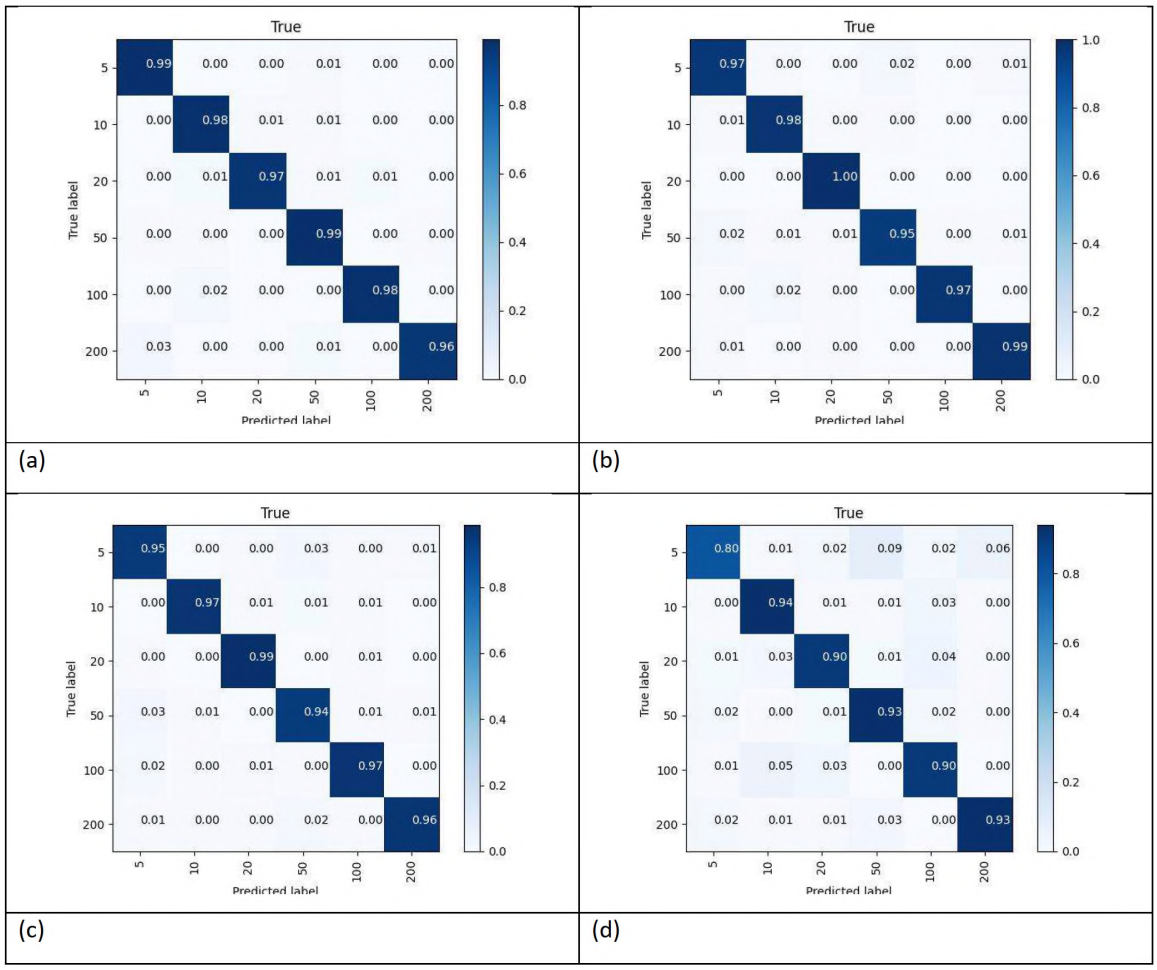


Figure 8. The confusion matrix of DenseNet121 when trained on images of various sizes with several optimizers. Trained with (a) image size of 200×200 and RMSprop optimizer, (b) image size of 150×150 and Adamax optimizer, (c) image size of 100×100 and Adamax optimizer, (d) image size of 50×50 and RMSprop optimizer.

performances of the VGG19 model when trained on varying image sizes and optimizers. Figure 13 shows the accuracy and loss curves for the VGG19 model, and Figure 14 shows the confusion matrix for each.

Table 7. *Analysis of accuracy of VGG16 trained over varying input sizes and optimizers*

Image Size	Optimizer	Accuracy Rate	Epoch number
200×200	RMSprop	0.9638	24
150×150	RMSprop	0.9669	20
100×100	Adamax	0.9537	13
50×50	RMSprop	0.9015	24

VGG19 achieved its highest accuracy of 96.89% with an image size of 200×200 using the Adamax optimizer, and it completed in 21 epochs. For smaller image sizes of 150×150 and 100×100, it achieved 96.02% and 94.57%, respectively. At the smallest image size of 50×50, the performance dropped to 89.28%, suggesting that VGG19 also benefits from larger image sizes but struggles more than VGG16 with smaller images. Table 8 lists the performance obtained by the MobileNetV2 model when trained on varying image sizes and optimizers. Figure 15 shows the accuracy and loss curves for the VGG19 model, and Figure 16 shows the confusion matrix for each.

The experimental results revealed that image size plays a significant role in determining model efficiency. For smaller images (50×50), no model or optimizer achieved satisfactory results, indicating that the image resolution was too low for the models to extract meaningful features. Some models performed reasonably well for image sizes of 100×100 and 150×150, especially when faster training times were prioritized. However, the best performance was observed with an image size of 200×200, particularly when combined with the DenseNet201 model and the SGD optimizer. This

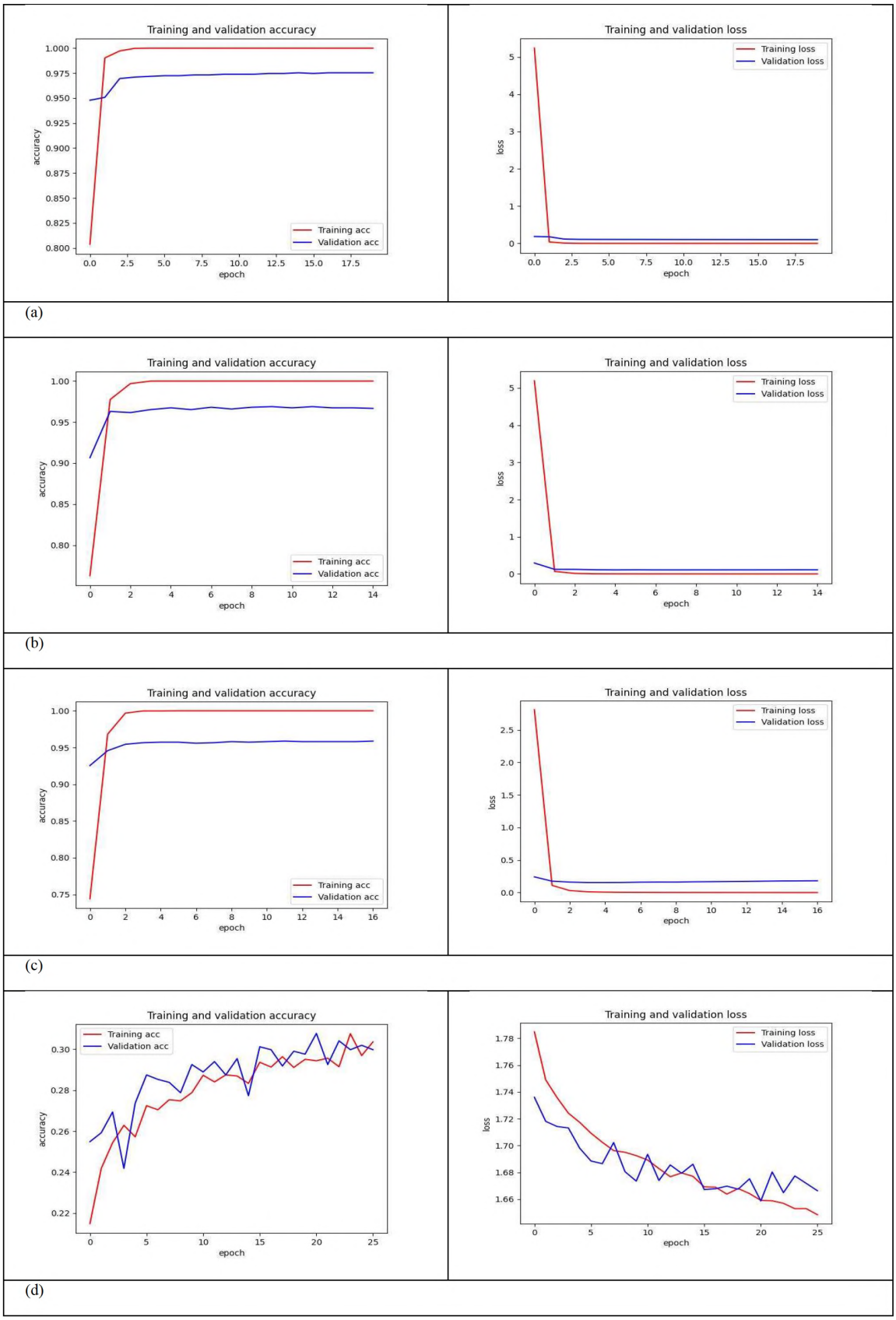


Figure 9. Accuracy and loss curves of MobileNet when trained over varying image sizes with several optimizers. Trained with (a) image size of 200×200 and Adamax optimizer, (b) image size of 150×150 and Adamax optimizer, (c) image size of 100×100 and Adamax optimizer, (d) image size of 50×50 and Adamax optimizer.

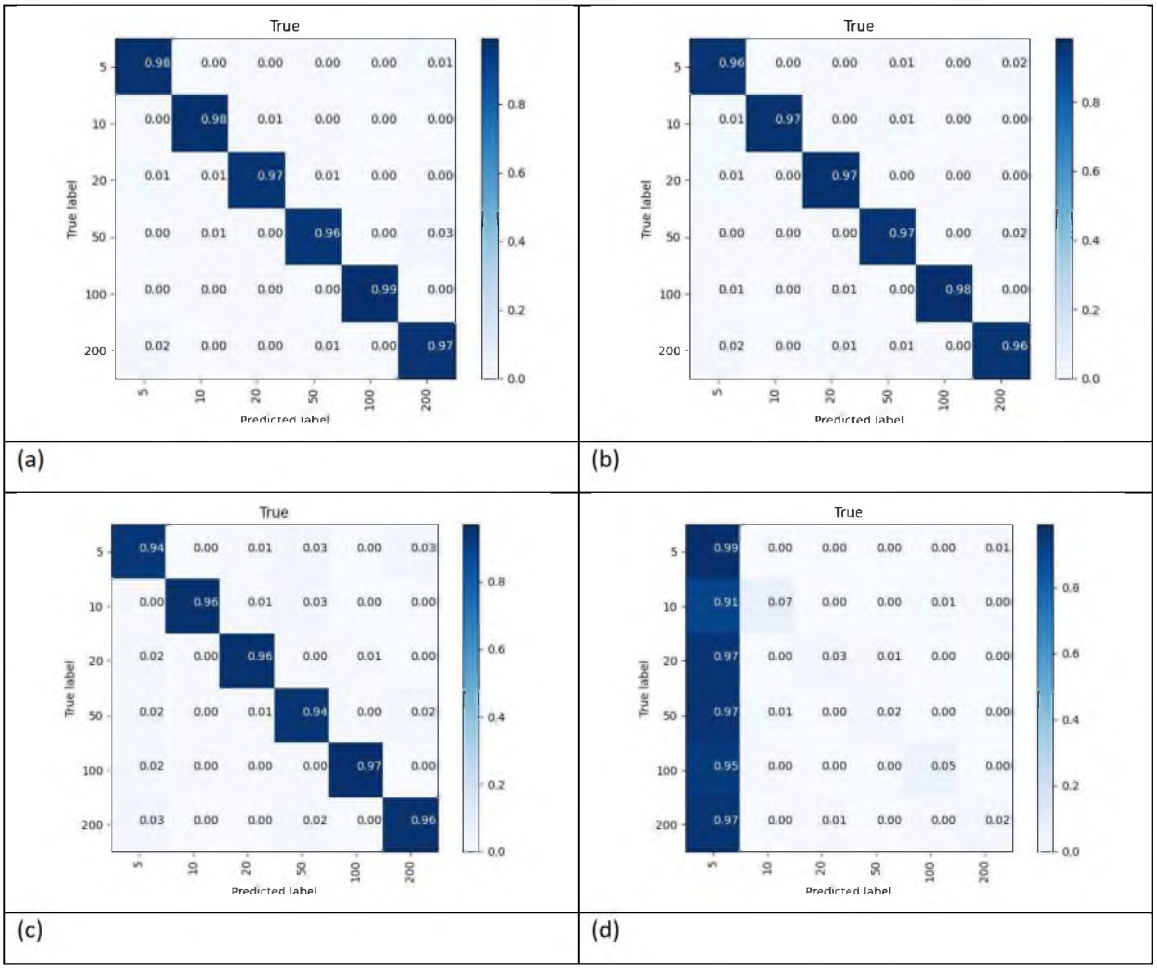


Figure 10. The confusion matrix of MobileNet when trained on images of various sizes with several optimizers. Trained with (a) image size of 200×200 and Adamax optimizer, (b) image size of 150×150 and Adamax optimizer, (c) image size of 100×100 and Adamax optimizer, (d) image size of 50×50 and Adamax optimizer.

Table 8. *Analysis of accuracy of VGG19 trained over varying input sizes and optimizers*

Image Size	Optimizer	Accuracy Rate	Epoch number
200×200	Adamax	0.9689	21
150×150	Adamax	0.9602	16
100×100	Adamax	0.9457	13
50×50	Adamax	0.8928	28

combination achieved an impressive accuracy of 98.84%, which improved efficiency significantly after the third epoch and stabilized by the eighth epoch, demonstrating that larger image sizes provide models with more detailed features for classification.

4. DISCUSSION

In this section, the results of the tests performed in the test environment are analyzed for the three variables studied, two of which are fixed variables. The data in the graphs were sorted from smallest to largest according to the appropriate cases, and the value analyzed was the accuracy value.

The image size parameter is a parameter tested in the test environment and is the first step of the feature detection process in the learning process. In the test environment, the width and height of the images were set to be the same. Although it seems to be an advantage to choose a high dimension, it can be observed that after a certain value, it is not useful and sometimes decreases the performance value.

The VGG, Mobile, and DenseNet models were selected as the types of transfer learning to be tested. Mobile

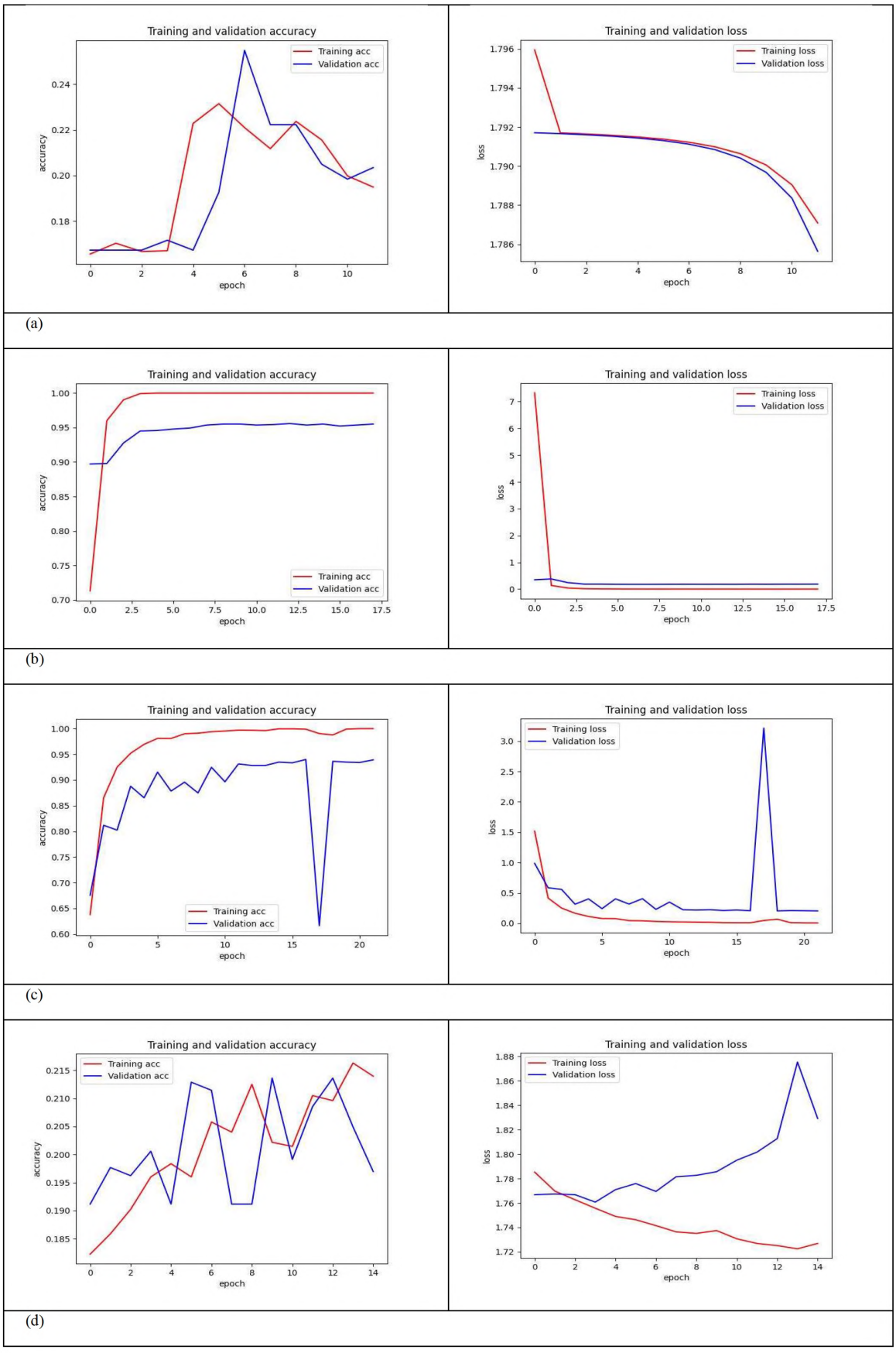


Figure 11. Accuracy and loss curves of MobileNetV2 when trained over varying image sizes with several optimizers. Trained with (a) image size of 200×200 and Ftrl optimizer, (b) image size of 150×150 and Adamax optimizer, (c) image size of 100×100 and SGD optimizer, (d) image size of 50×50 and Nadam optimizer.

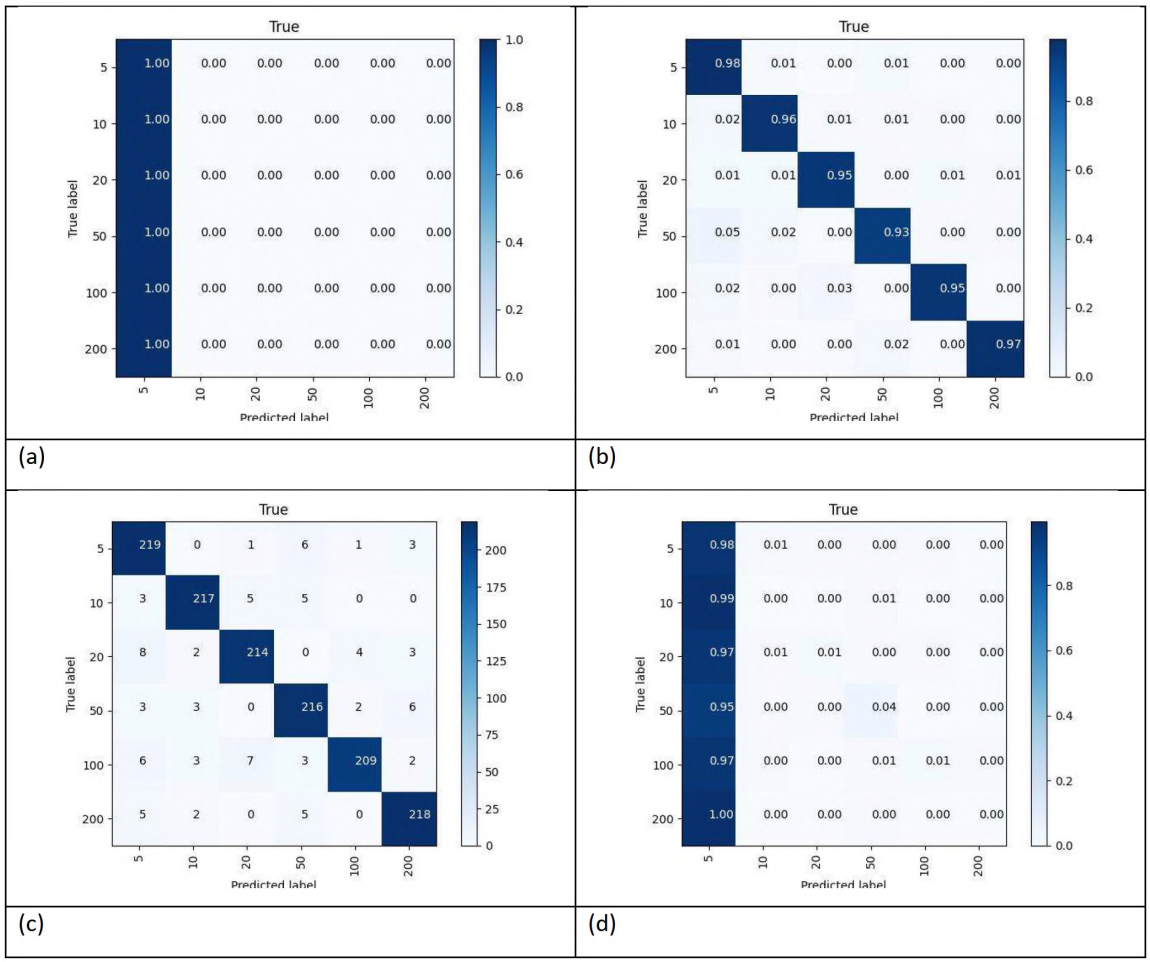


Figure 12. The confusion matrix of MobileNetV2 when trained on images of various sizes with several optimizers. Trained with (a) image size of 200×200 and Ftrl optimizer, (b) image size of 150×150 and Adamax optimizer, (c) image size of 100×100 and SGD optimizer, (d) image size of 50×50 and Nadam optimizer.

models were selected because they are fast and have fewer layers, VGG models have more layers, and Dense models have a balanced number of layers but run slower than other models. In this study, another tested parameter was the optimizer function. These functions affect learning speed and weight. For this reason, they also affect the performance measurement value.

From Figure 17, it can be observed that there is no efficiency with any transfer model or optimizer function when images with a width and height of 50 are used in the test environment. For the other image sizes, low efficiency was observed in some cases, but high efficiency was observed in most cases. A size of 100 is good if the training process must be rapid, and a size of 200 is good if performance is important. Increasing the size is not an important parameter alone; however, it is advantageous when used in conjunction with other parameters in the test environment.

In Figure 18, the effect of the examined model on the performance accuracy value was not significantly different from that of the appropriate parameters. The most efficient models were DenseNet201 and DenseNet121, and the least efficient models were MobileNetV2. It can be seen that the selection of the appropriate model is not the only condition that provides sufficient efficiency; the appropriate optimizer function and appropriate image size should also be determined. It can be observed that appropriate accuracy values can occur under appropriate conditions for all models.

Figure 19 shows that Ftrl is the least efficient optimizer function. With the other functions, efficient results can be obtained in an appropriate test environment. The efficiency criterion can be reached when the parameters are tested using various combinations of parameters that do not have appropriate values for the parameters alone for efficient operation in the test environment.

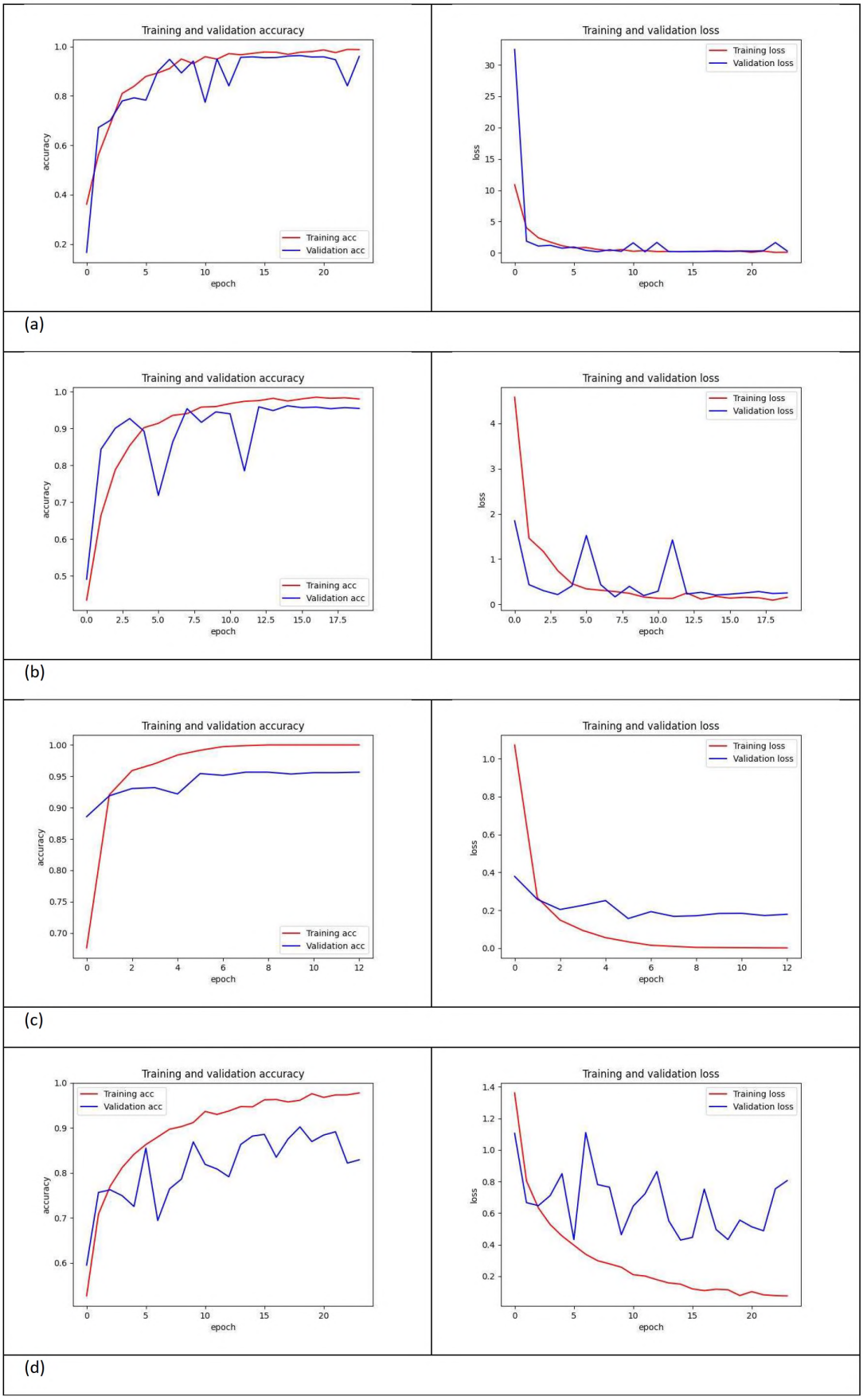


Figure 13. Accuracy and loss curves of VGG16 when training for various image sizes with several optimizers. Trained with (a) image size of 200×200 and RMSprop optimizer, (b) image size of 150×150 and RMSprop optimizer, (c) image size of 100×100 and Adamax optimizer, (d) image size of 50×50 and RMSprop optimizer.

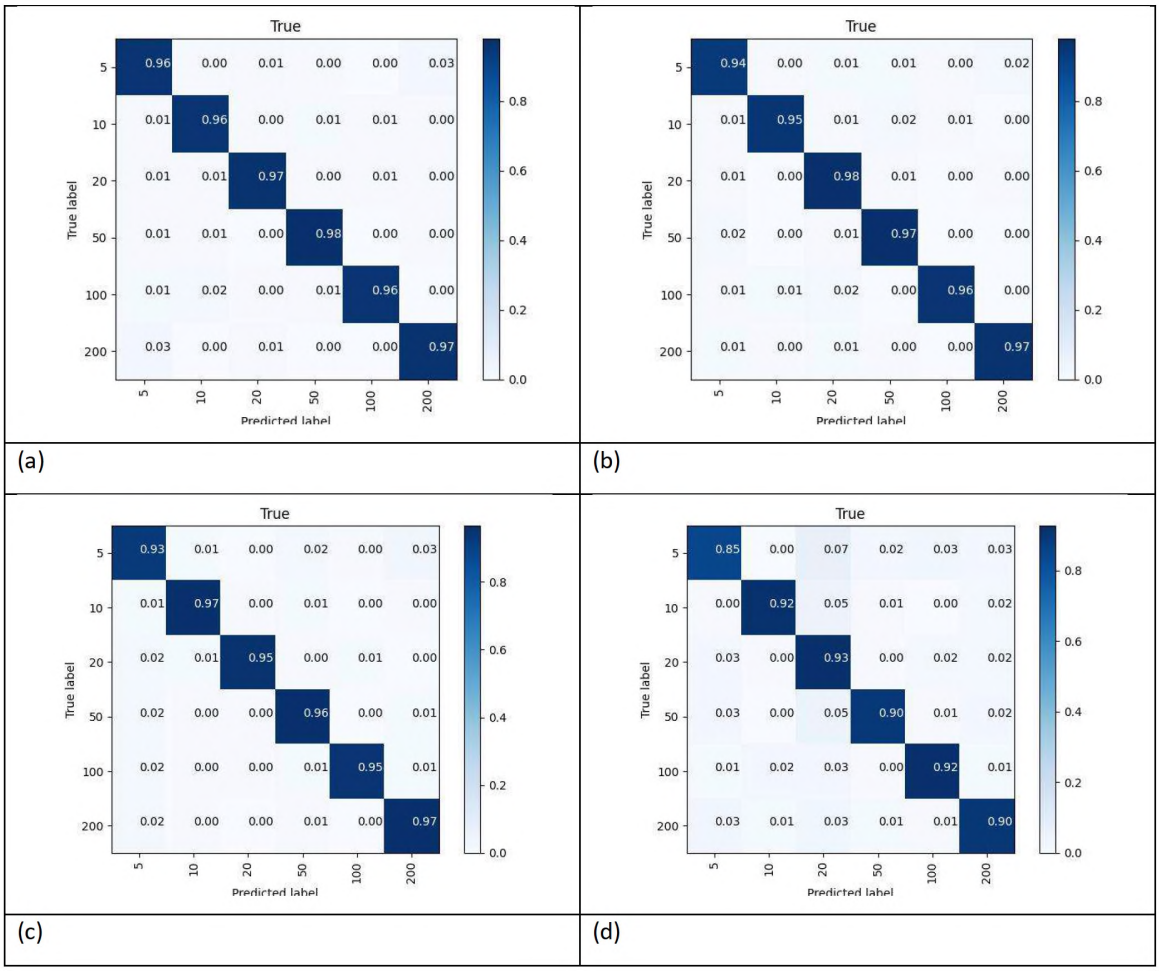


Figure 14. Accuracy and loss curves of VGG16 when training for various image sizes with several optimizers. Trained with (a) image size of 200×200 and RMSprop optimizer, (b) image size of 150×150 and RMSprop optimizer, (c) image size of 100×100 and Adamax optimizer, (d) image size of 50×50 and RMSprop optimizer.

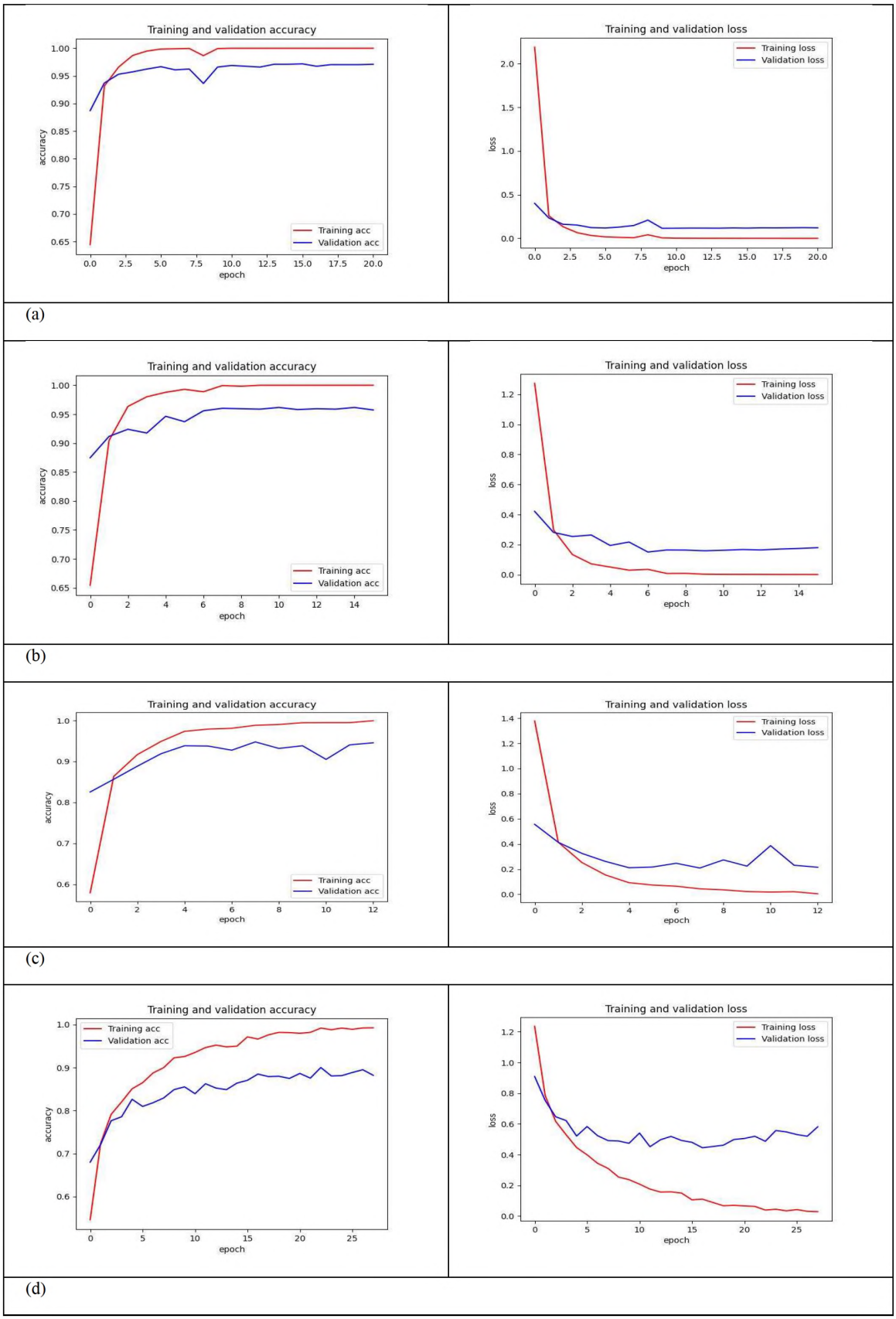


Figure 15. Accuracy and loss curves of VGG19 after training on various image sizes with several optimizers. Trained with (a) image size of 200×200 and Adamax optimizer, (b) image size of 150×150 and Adamax optimizer, (c) image size of 100×100 and Adamax optimizer, (d) image size of 50×50 and Adamax optimizer.

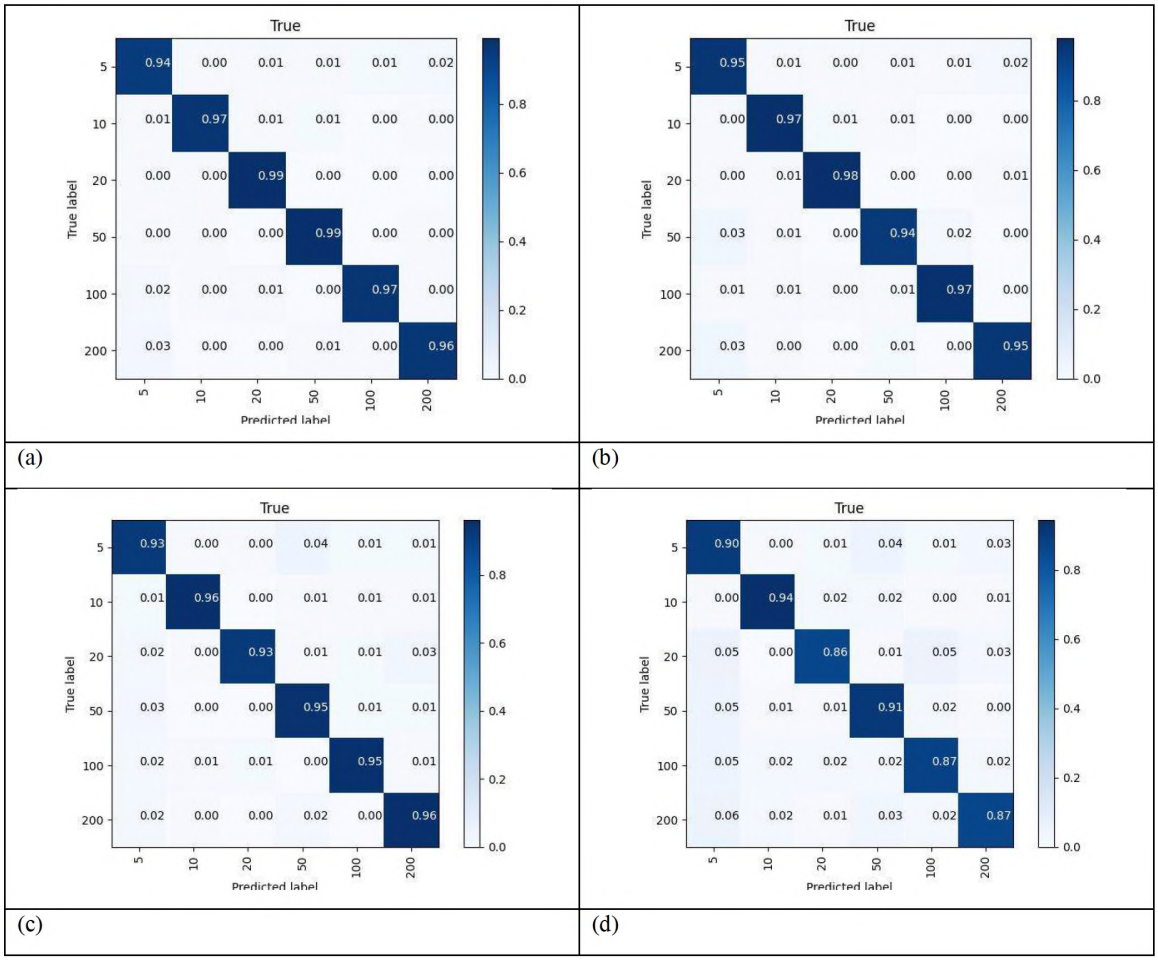


Figure 16. The confusion matrix of VGG19 when training for various image sizes with several optimizers. Trained with (a) image size of 200×200 and Adamax optimizer, (b) image size of 150×150 and Adamax optimizer, (c) image size of 100×100 and Adamax optimizer, (d) image size of 50×50 and Adamax optimizer.

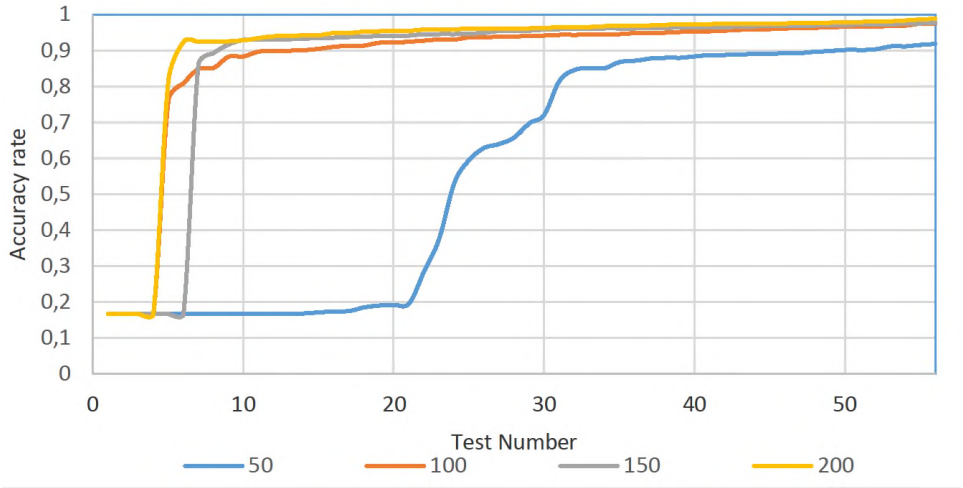


Figure 17. The image size and test performance relationship.

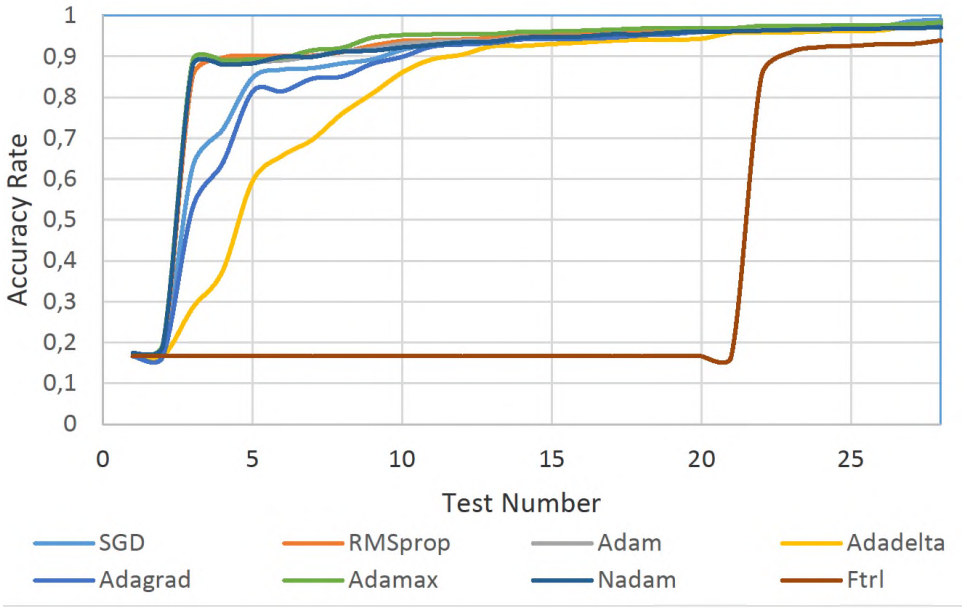


Figure 18. The accuracy comparison relationship between transfer learning models with respect to the test set.

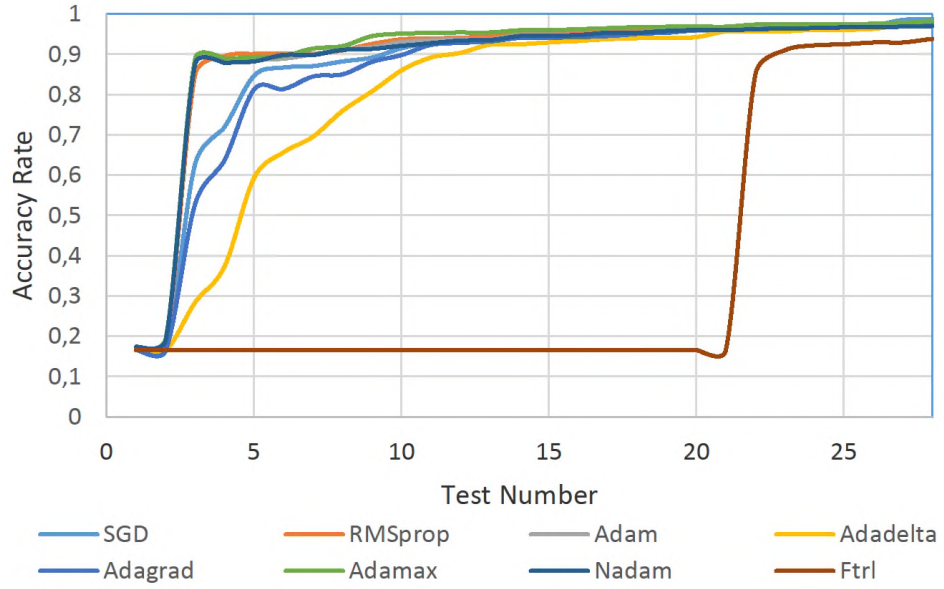


Figure 19. Figure 19 shows that Ftrl is the least efficient optimizer function. With the other functions, efficient results can be obtained in an appropriate test environment. The efficiency criterion can be reached when the parameters are tested using various combinations of parameters that do not have appropriate values for the parameters alone for efficient operation in the test environment.

5. CONCLUSION

With advancements in artificial intelligence and deep learning, research into banknote classification has provided new opportunities to realize more accurate, reliable, and scalable systems. These systems can leverage new technologies such as edge computing and Internet of Things, to improve the efficiency of real-time recognition systems. Efficient real-time classification ensures faster transactions and enhances user experience. This study investigated the performance of various deep learning models in the classification of Turkish banknotes using a comprehensive dataset of 6901 images across six denominations (5 TL, 10 TL, 20 TL, 50 TL, 100 TL, and 200 TL). The dataset includes banknotes under different conditions such as flat, angled, and bent conditions, providing a realistic testing environment. The study evaluates the performance of pre-trained models, including VGG16, VGG19, DenseNet121, DenseNet169, DenseNet201, MobileNet, and MobileNetV2, by varying key parameters, such as image size (50×50, 100×100, 150×150, 200×200) and optimizers (SGD, RMSprop, Adamax, Adam, and others). The best results were obtained using the DenseNet201 model, which achieved an accuracy of 98.84% with an image size of 200 and the SGD optimizer in just 12 epochs. DenseNet169 also performed well, achieving 98.62% accuracy under similar conditions. The study highlights the importance of selecting appropriate image sizes and optimizers because smaller image sizes significantly reduced model performance, particularly for MobileNet and MobileNetV2, which showed poor results with images sized 50×50. In addition, models using the Ftrl optimizer consistently exhibited lower performance. These findings demonstrate the effectiveness of DenseNet models for banknote classification, especially when combined with larger image sizes and optimizers like SGD. The results provide valuable insights into developing robust and efficient automated currency recognition systems, especially for real-time applications in financial sector.

In future work, it is planned to conduct tests using banknotes from different countries. Another field of study is to work with a test environment in which the effect of optimizing functions as variables can be better observed.

Peer Review: Externally peer-reviewed.

Conflict of Interest: Authors declared no conflict of interest.

Grant Support: The authors received no specific funding for this work.

ORCID IDs of the authors

Mirsat Yeşiltepe 0000-0003-4433-5606
Harun Elkıran 0000-0002-5834-6210
Jawad Rasheed 0000-0003-3761-1641

REFERENCES

Baek, S., Choi, E., Baek, Y., & Lee, C. (2018). Detection of counterfeit banknotes using multispectral images. *Digital Signal Processing, 78*, 294–304. https://doi.org/10.1016/j.dsp.2018.03.015

Baltaci, F. (2020). Turkish lira banknote dataset. Retrieved January 2, 2024, from https://www.kaggle.com/datasets/baltacifatih/turkish-lira-banknote-dataset

Baykal, G., Demir, U., Shyti, I., & Unal, G. (2018). Turkish lira banknotes classification using deep convolutional neural networks. In 2018 26th Signal Processing and Communications Applications Conference (SIU) (pp. 1–4). IEEE.

Filter, J. (2022). Split-folders - PyPI. Retrieved January 15, 2024. from https://pypi.org/project/split-folders/

Foody, G. M. (2023). Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient. *PLOS ONE, 18*(10), e0291908. https://doi.org/10.1371/journal.pone.0291908

Galeana Pérez, D., & Bayro Corrochano, E. (2018). Recognition system for euro and Mexican banknotes based on deep learning with real scene images. *Computación y Sistemas, 22*(4). https://doi.org/10.13053/cys-22-4-3079

İyikesici, B., & Erçelebi, E. (2023). An efficient deep learning architecture for Turkish lira recognition and counterfeit detection. *Turkish Journal of Electrical Engineering and Computer Sciences, 31*(3), 678–692. https://doi.org/10.55730/1300-0632.4009

Keras Team. (2023a). Keras application. Keras. Retrieved January 15, 2024, from https://keras.io/api/applications/

Keras Team. (2023b). Optimizers. Keras. Retrieved January 17, 2024, from https://keras.io/api/optimizers/

Khashman, A., Ahmed, W., & Mammadli, S. (2019). Banknote issuing country identification using image processing and neural networks. In Proceedings (pp. 746–753). https://doi.org/10.1007/978-3-030-04164-9_98

Khashman, A., & Sekeroglu, B. (2005). *Multi-banknote identification using a single neural network*. In Proceedings (pp. 123–129).

Khashman, A., Sekeroglu, B., & Dimililer, K. (2005). Deformed banknote identification using pattern averaging and neural networks. In Proceedings of the 4th WSEAS International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics (pp. 233–237).

Linkon, A. H. M., Labib, M. M., Bappy, F. H., Sarker, S., Jannat, M.-E., & Islam, M. S. (2020). Deep learning approach combining lightweight CNN architecture with transfer learning: An automatic approach for the detection and recognition of Bangladeshi banknotes. In 2020 11th International Conference on Electrical and Computer Engineering (ICECE) (pp. 214–217). IEEE. https://doi.org/10.1109/icece51571.2020.9393113

Mittal, S., & Mittal, S. (2018). Indian banknote recognition using convolutional neural network. In 2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU) (pp. 1–6). IEEE. https://doi.org/10.1109/iot-siu.2018.8519888

Pachón, C. G., Ballesteros, D. M., & Renza, D. (2023). An efficient deep learning model using network pruning for fake banknote recognition. Expert Systems with Applications, 233, 120961. https://doi.org/10.1016/j.eswa.2023.120961

Prakash, H., Yadav, A., Ushashree, P., Jha, C., Sah, G. K., & Naik, A. (2023). *Deep learning approaches for automated detection of fake Indian banknotes*. In 2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS) (pp. 1–5). IEEE. https://doi.org/10.1109/icicacs57338.2023.10100265

Sahin, O. (2018). *GitHub - Ozgurshn/TurkishBanknoteDataset.* Retrieved January 2, 2024, from https://github.com/ozgurshn/TurkishBanknoteDataset.

Veeramsetty, V., Singal, G., & Badal, T. (2020). CoinNet: Platform independent application to recognize Indian currency notes using deep learning techniques. *Multimedia Tools and Applications, 79*(31–32), 22569–22594. https://doi.org/10.1007/s11042-020-09031-0

Wang, L., Zhang, Y., Lanchi, X., Zhang, X., Guang, X., Li, Z., Li, Z., Shi, G., Hu, X., & Zhang, N. (2022). Automated detection and classification of counterfeit banknotes using quantitative features captured by spectral-domain optical coherence tomography. *Science & Justice, 62*(5), 624–631.https://doi.org/10.1016/j.scijus.2022.09.004.

How cite this article

Yeşiltepe, M., Eklıran, H., & Rasheed, J. (2024). Turkish Lira Banknote Classification using Transfer Learning and Deep Learning. *Acta Infologica, 8*(2), 133-156. https://doi.org/10.26650/acin.1447456

İSTANBUL UNIVERSITY PRESS

Acta Infologica
acin 2024, 8 (2), 157–175
DOI: 10.26650/acin.1462978

Research Article

Evaluation of the Quality of Second-Hand Clothes Trading Websites using the VIKOR and MOORA Methods

Sinem Büyüksaatçı Kiriş[1]

[1]İstanbul University-Cerrahpaşa, Faculty of Engineering, Department of Industrial Engineering, İstanbul, Türkiye

Corresponding author : Sinem Büyüksaatçı Kiriş
E-mail : s.buyuksaatcikiris@iuc.edu.tr

ABSTRACT

Along with technology, people's living standards have also improved in the same direction. Long-distance shopping trips are no longer preferred by people. Many consumers now find it more appealing to shop while sitting at home and avoiding the crowds while also finding what they are looking for more quickly. However, this impressive opportunity provided by technology can sometimes create difficulties for customers. In the online realm, there are various alternative e-commerce websites. This led to frequent consideration of the query, "Which website is better for me to shop from?". Here convenience means price, security, product variety, etc. such criteria. This rightful consideration of consumers has led many researchers to enter this field and have enabled them to use decision-making methods during the selection of e-commerce sites. From the 2000s to the present, second-hand goods have gained popularity again. The fact that second-hand goods are generally cheaper than new ones encourages thrifty consumers to buy second-hand goods because of their price advantages. In addition, many consumers may turn to second-hand sources because they are concerned about the scarcity of natural resources and the volume of waste generated. The trend toward buying second-hand products is driven by the issue of sustainability, which has recently received significant attention. In this study, the quality of three e-commerce sites selling second-hand clothes in Turkey was evaluated using various multi-criteria decision-making methods. The criteria taken into consideration while shopping on the e-commerce sites in question were determined from studies in the literature and from a website called "Şikayetvar" that is frequently visited by internet users. The weights of the determined criteria were then calculated using the AHP method. The sites were then compared using the VIKOR and MOORA methods.

Keywords: Second Hand Clothes, Sustainability, VIKOR, MOORA

Submitted	: 01.04.2024
Revision Requested	: 03.09.2024
Last Revision Received	: 03.10.2024
Accepted	: 20.10.2024
Published Online	: 31.10.2024

1. INTRODUCTION

"The acquisition of second-hand objects through methods and venues of exchange that are often different from new products" is the definition of second-hand shopping. Digital channels have witnessed a surge in the popularity and reach of second-hand shopping, particularly in consumer-to-consumer e-commerce.

In fact, buying second-hand has existed as a form of purchasing alternative since the mid-fourteenth century. People's poverty levels gradually increased due to the deep economic depression caused by the plague in Europe in the 16th century, increasing population, political and social uprisings, and severe famine. The deterioration of the economy has led people to give the clothes they use to each other, expecting something in return and exchanging them. The development of second-hand product trade continued until the 18th century, during which time the Industrial Revolution occurred. From the 2000s to the present, second-hand goods have regained popularity, and the used product market has begun to grow and attract much attention. (Borusiak et al., 2020).

Although the second-hand market has many features and benefits compared to traditional sales, it can be more complex in some respects. For example, a second-hand store can compete in three different markets at the same time. These include rivalry from consignors, buyers, and dealings with the waste industry. Furthermore, many customers anticipate that second-hand shops will be of a quality comparable to chain stores or retail establishments that offer brand-new merchandise. What is expected from a decent-quality store is that it is well organized and that the goods and shopping experience are provided with the help of a friendly staff. Therefore, offline channels offer the chance to browse the store, examine products, and engage in social interaction while shopping. When using online channels, it cannot be adjusted in this way for reasons closely related to physical stores. When evaluating product condition during second-hand online shopping, customers should benefit from images and product descriptions rather than touching the actual product. Therefore, when assessing products in online channels, additional factors such as brand name, price, store, and country of origin are taken into consideration. In addition, online stores enable you to reach a wider audience geographically. Although it is claimed that some geographical restrictions may still exist in e-commerce, in theory, any seller can sell their products to users anywhere in the world. Additionally, online stores do not restrict the time a customer can shop (Kassinen and Koivumäki, 2019).

Consumers and site users make purchases when purchasing second-hand products for various reasons. According to Kassinen and Koivumäki (2019), these are known as economic and critical motivations. Many researchers have emphasized economic reasoning when purchasing second-hand products. It is observed that in most cases, second-hand goods are not purchased as the first choice; instead, consumers are forced to buy second-hand goods due to economic constraints. Saving money is an important factor for people in general. Second-hand markets are important for people with very little income (Napompech and Kuawiriyapan, 2011). In a study investigating the main reasons for buying second-hand goods, it was concluded that the main reasons for the group between the ages of 18 and 24 were saving money (Nieminen, 2016).

There are three main categories of critical reasons for purchasing second-hand goods. These are listed as avoiding traditional methods, ethical and ecological motivations, and anti-vanity. The inclination of people to distance themselves from the current consumption system is linked to their avoidance of traditional methods. Due to the rapid phase of consumption, many people are unusually turning to second-hand goods as they realize that second-hand items are generally relatively less used and are therefore still usable and often in good condition.

According to ethical and ecological motivations, many people may turn to second-hand resources because they are concerned about the scarcity of natural resources and the amount of waste produced. Waste is considered a negative factor. When a used product is purchased instead of a new one, there is no waste, thus allowing the balanced use of production resources (Dengin, 2012). Research has shown that students who purchase items from second-hand stores are more aware of environmental issues (Borusiak et al., 2020). According to anti-vanity motivation, many people are not interested in the latest trends in their purchases. It has been observed that many people with this view tend to purchase functional products that add value to the person rather than show off.

Secondhand markets provide a focused solution to the issues raised by fast fashion. Clothes are frequently destroyed before they reach the end of their life cycle because people consume fashion so quickly. Many people believe they can act environmentally responsibly without sacrificing the satisfaction of their true needs and desires when buying used clothing. Regardless of the item's age, second-hand clothing is defined as clothing that has been worn previously. Nonetheless, whether or not an item of clothing determines whether it qualifies as vintage. Unlike many other second-hand goods, used clothing is made of a unique material or size.

Although the use of second-hand clothes was seen as a style of clothing inspired by poverty in previous periods, this idea gained a different meaning in the 90s. In the 90s, fashion became a popular way to recreate the 1970s. The

fashion trend has shifted people to second-hand shopping markets. Over time, consumers have begun to realize the benefits of second-hand clothing and have turned to sustainability issues. The fact that second-hand goods are generally cheaper than new goods encourages frugal consumers to buy second-hand goods because of their price advantages. However, consumers who buy second-hand clothes with low budgets have adopted this situation as a sociological conflict avoidance strategy and a means of escaping the burden of poverty (Tóta, 2016).

The sale of second-hand clothes on e-commerce sites has increased considerably. The increase in the number of such sites leads users to confusion about which sites to shop from. In this study, it is aimed to evaluate the quality of three e-commerce sites operating for the sale of second-hand clothing in Turkey using various multi-criteria decision-making methods. The weights of the criteria were calculated using the AHP method. Subsequently, an evaluation was performed between the sites using the VIKOR and MOORA methods. It has been observed that e-commerce sites operating in different fields, such as Limango, Markofoni, Trendyol, Morhipo, D & R, Hepsiburada, and LCWaikiki, are frequently discussed in the literature. Only one study has been found to evaluate sites where second-hand clothes are sold (Karadayı Usta and Kadaifci, 2022). However, this study differs from the aforementioned study in terms of the criteria used, method for obtaining criteria, and methods used for ranking alternatives.

This paper is organized as follows. Section 2 discusses the methods used. Section 3 presents the application in detail. The obtained findings are also reported at the end of each relevant step in this section. Section 4 concludes the paper with recommendations for further research.

2. METHODOLOGY

2.1. Analytical Hierarchy Process (AHP)

The AHP is a multi-criteria decision-making tool that can make pairwise comparisons between criteria using the eigenvalue approach and calibrate the numerical scale used in quantitative and qualitative performance measurements. According to the definition in the Operations Management book written by Russell and Taylor (2003), it is a quantitative method used to rank decision alternatives and make selections according to multiple criteria (Çelikbilek, 2018).

Myers and Alpert first presented the AHP idea in 1968. Saaty then developed the model and found it useful for solving decision-making problems in 1977. When the decision hierarchy can be identified, the Analytic Hierarchy Process (AHP) is utilized as a decision-making and estimation technique that provides the percentage distribution of decision points in terms of the criteria influencing the decision (Yaralıoğlu and Köksal, 2003).

The steps of the AHP Method are presented below:

STEP 1. Defining the Problem: As in every problem, the first stage of the AHP method defines the problem. It is very important that this step is performed well. Any mistake made at this point will directly affect the course of the transaction.

STEP 2. Determination of Criteria/Alternatives: After a clear definition of the problem, it will be much easier to determine the criteria to be used for the solution and alternatives to be evaluated.

STEP 3. Creating the Hierarchical Structure: The hierarchical structure of the target problem is created at this stage, as shown as an example in Fig. 1.

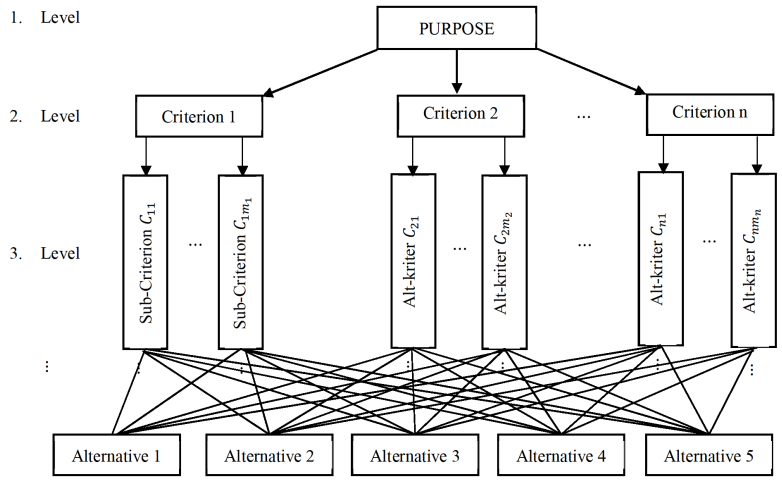


Figure 1. Hierarchical Structure Example

STEP 4. Pairwise Comparisons: In this step, pairwise comparison matrices are prepared for each level, adhering to the hierarchical structure.

When performing pairwise comparisons, the pairwise comparison scale recommended by Wind and Saaty (1980) and given in Table 1 is used. Pairwise comparisons are performed for only one side of the principal diagonal of the comparison matrix. In common use, the upper part of the prime diagonal is filled in by decision-makers, and the values below the prime diagonal are completed accordingly.

Table 1. Fundamental scale for pairwise comparisons (Wind and Saaty, 1980)

Importance intensity	Definition	Explanation
1	Equal Importance	The two activities contribute equally to the objective of
3	Moderate Importance	Experience and judgment slightly favor one activity over another.
5	Strong Importance	Experience and judgment strongly favor one activity over another
7	Very Strong Importance	An activity is very strongly favored over another; its dominance is demonstrated in practice
9	Absolute Importance	The evidence-favoring one activity over another is of the highest possible order of affirmation
2,4,6,8	For compromise between the above values	Sometimes, a compromise judgment must be interpolated numerically. Because there is no good word to describe.
Reciprocals Of the above		If activity i has one of the above nonzero numbers assigned to it when compared with activity j, then j has the reciprocal value compared with i.

STEP 5. Normalization of Pairwise Comparison Matrices: Normalization is done by taking the row totals. A pairwise comparison matrix containing n criteria is denoted by $[x_{ij}]_{n\times n}$, normalization calculations are carried out with the help of Eq. (1), and the normalized matrix $[y_{ij}]_{n\times n}$ is obtained.

$$y_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}} \tag{1}$$

STEP 6. Calculation of Priority Vectors: Priority vectors are obtained by averaging the rows of the normalized matrix. In a problem with n criteria, the priority vector is denoted by $A = [a_i]_{n\times 1}$. The priority value of the first criterion is calculated with the help of Eq. (2).

$$a_1 = \frac{\sum_{j=1}^{n} y_{1j}}{n} \tag{2}$$

STEP 7. Consistency Tests: These tests are applied to determine whether pairwise comparisons are consistent. Inconsistent test results require a re-comparison. To perform the consistency test, the normalized pairwise comparison matrix must first be multiplied by the priority vector. The Eq. valid for the consistency test of the pairwise comparison of n criteria is given in (3).

$$[t_i]_{n\times 1} = [x_{ij}]_{n\times n}.[a_i]_{n\times 1} \tag{3}$$

Because of Eq. (3), each element of the obtained vector is divided into priority vector elements, respectively.

$$\frac{t_i}{a_i} \quad For\ \forall i \in n \tag{4}$$

After Eq. (4), the largest eigenvalue, , λ_{maks} is calculated by averaging the obtained values.

$$\lambda_{maks} = \frac{\sum_{i=1}^{n} \frac{t_i}{a_i}}{n} \tag{5}$$

The consistency index (CI) is then calculated with the help of λ_{maks}.

$$CI = \frac{\lambda_{maks} - n}{n - 1} \tag{6}$$

After applying Eq. (6), the consistency ratio result (CR) is obtained by dividing the fit index (CI) and random index (RI) values together.

$$CR = \frac{CI}{RI} \tag{7}$$

The random index value to be used during the calculation in Eq. (7) varies according to the number of criteria and is shown in Table 2.

Table 2. RI Values for Consistency Ratio Calculation

n	RI
3	0,58
4	0,90
5	1,12
6	1,24
7	1,32
8	1,41
9	1,45
10	1,49
11	1,51
12	1,53

The consistency ratio obtained because of these calculations is expected to be less than 0.1. If CR<0.1, the discrepancy rate is considered to be at an acceptable level.

2.2. VIKOR Method

The VIKOR method, developed for the optimization of multi-criteria complex systems, is based on choosing and ranking alternatives when there are conflicting criteria. Considering that each alternative is evaluated according to each criterion function, the multi-criteria measure for compromise ranking is developed from the L_p criterion, which is used as the aggregation function in compromise programing. Various j alternatives are shown as $c_1, c_2, \ldots, c_n$. The measurement of criterion i of alternative c_j is expressed as f_{ij}. The VIKOR method developed with the L_p form is described in Eq. (8) (Yıldırım and Önder, 2014).

$$L_{pj} = \left\{ \sum_{i=1}^{n} [w_i (f_i^* - f_{ij}) / (f_i^* - f_i^-)]^p \right\}^{(1/p)} , 1 \leq p \leq \infty; j = 1, 2, \ldots, j \tag{8}$$

STEP 1. Creating a Decision Matrix: As in other methods, the decision process in the VIKOR method begins with a definition of the decision problems. In decision problems, the alternatives to be evaluated and the criteria that enable the selection of alternatives are determined. Although the criteria are determined intuitively to meet the expectations of the decision-maker, they can also be obtained by consulting experts on the subject. When evaluated according to the relevant criterion, the values obtained from the alternatives indicate the scores of the alternatives. After the criteria of the decision problem and the scores of the alternatives are determined according to the criteria, the scores are converted into a decision matrix, as shown in Eq. (9). The rows of the decision matrix (i=1,2,. . . ,m) show the alternatives, and the columns (j=1,2,. . . ,n) show the criteria.

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{1n} \\ . & & & . \\ . & & & . \\ . & & & . \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \tag{9}$$

STEP 2. Determination of ideal solution values: After creating the decision matrix, the best f_j^+ and worst f_j^- values are determined for each criterion. In determining the f_j^+ and the f_j^- values, calculations were performed in two ways, considering the criterion features. If criterion j has a benefit feature, f_j^+ and the f_j^- values are;

$$\begin{aligned} f_j^+ &= \frac{max}{i} x_{ij} \\ f_j^- &= \frac{max}{i} x_{ij} \end{aligned} \tag{10}$$

If criterion j represents a cost, that is a loss, f_j^+ and the f_j^- values are;

$$\begin{aligned} f_j^+ &= \frac{max}{i} x_{ij} \\ f_j^- &= \frac{max}{i} x_{ij} \end{aligned} \tag{11}$$

STEP 2. Normalization Process and Creation of Normalization Matrix: A linear normalization process is applied to purify the values that make up the decision matrix from their similarities and keep them at a comparable level. The decision matrix of a decision problem consisting of m alternatives and n criteria is transformed into the R normalization matrix which is $m \times n$ dimensions. The elements of the R matrix are calculated with the help of Eq. (12).

$$r_{ij} = \frac{f_j^+ - x_{ij}}{f_j^+ - f_j^-} \tag{12}$$

$$X = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{12} & r_{22} & \dots & r_{2n} \\ . & & & . \\ . & & & . \\ . & & & . \\ r_{m1} & r_{m1} & \dots & r_{mn} \end{bmatrix} \tag{13}$$

STEP 3. Weighting of Normalized Decision Matrix: By multiplying the criteria shown in the columns of the normalized decision matrix with the relevant w_j weights, the V weighted normalized decision matrix. is obtained. The weighted normalized decision matrix elements are calculated using Eq. (14).

$$v_{ij} = r_{ij} \times w_j \tag{14}$$

$$X = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1n} \\ v_{12} & v_{22} & \dots & v_{2n} \\ . & & & . \\ . & & & . \\ . & & & . \\ v_{m1} & v_{m1} & \dots & v_{mn} \end{bmatrix} \tag{15}$$

STEP 4. Calculation of S_i and R_i Values: S_i and R_i values are calculated for (j=1,2,. . . ,n). S_i and R_i values show the average and worst group scores for the alternative i.

$$\begin{aligned} S_i &= \sum_{j=1}^{n} v_{ij} \\ S_i &= \sum_{j=1}^{n} w_j \times r_{ij} \\ S_i &= \sum_{j=1}^{n} w_j \times \frac{f_j^* - x_{ij}}{f_j^* - f_j^-} \end{aligned} \tag{16}$$

$$\begin{aligned} R_j &= \frac{max}{j} v_{ij} \\ R_j &= \frac{max}{j} (w_j \times r_{ij} \\ R_j &= \frac{max}{j} \left(w_j \times \frac{f_j^* - x_{ij}}{f_j^* - f_j^-} \right) \end{aligned} \tag{17}$$

STEP 5. Calculation of Q_i Values: S^*, S^-, R^* and R^- parameters used in the calculation step of Q_i values are shown in Eq. (18).

$$\begin{aligned} S^* &= \frac{min}{i} S_i \\ S^- &= \frac{max}{i} S_i \\ S^* &= \frac{min}{i} R_i \\ S^- &= \frac{max}{i} R_i \end{aligned} \tag{18}$$

The q parameter used in the calculation of Q_i values shows the weight (maximum group benefit) of the majority of the criteria. While the q value expresses the weight for the strategy that provides the maximum group benefit, (1-q) expresses the weight of the minimum regret of the opponents. Consensus is achieved by "majority vote" (q>0,5), "consensus" (q=0,5) or "veto" (q<0,5).

$$Q_i = \frac{q \times (S_i - S^*)}{(S^- - S^*} + \frac{(1-q) \times (R_i - R^*)}{R^- - R^*} \tag{19}$$

STEP 6. Ranking Alternatives and Checking Conditions: By ordering S_i, R_i and Q_i values from smallest to largest, it is checked whether the alternative with the minimum Q_i value meets the following two conditions to test the accuracy of the rankings of the alternatives.

Condition 1. Acceptable advantage: When Q_i values are listed from smallest to largest and the first alternative is shown as A^1 and the second alternative is shown as A^2, acceptable advantage connected to condition 20.

$$Q(A^2) - Q(A^1) \geq DQ \tag{20}$$

The ***DQ*** parameter used in Eq. (20) depends on the number of alternatives and is calculated using Eq. (21) to indicate the number of alternatives ***m***.

$$DQ = \frac{1}{m-1} \tag{21}$$

Condition 2. Acceptable stability condition: Alternative A^1, which comes first when Q_i values are ranked from smallest to largest, is the best alternative with the minimum value when ranked from smallest to largest according to S and R values. In cases where any of these two conditions are not met, the following solution set is considered. If the acceptable stability condition is not met, both alternatives A^1 and $A^2$are accepted as the compromise solution. If an acceptable advantage is not provided, all alternatives $A^1, A^2, \ldots, A^m$ are included in the compromise best common solution set.

2.3. MOORA Method

The MOORA method was first introduced in 2006 by Brauers and Zavadskas as "a new method proposed for multi-objective optimization with distinct alternatives". The normalization step of the MOORA method was the same as that of TOPSIS. In this method, unlike the TOPSIS and VIKOR methods, non-ideal solutions are considered. The solutions are based solely on the reference point. The relationship between each alternative and the reference point is determined by taking the difference for each criterion (Çelikbilek, 2018).

The MOORA method refers to the matrix of responses of alternatives for the purposes for which the rates are applied. The MOORA method is mostly applied in two methods: the ratio method and the reference point approach. The method begins by writing the data in the form of a matrix, in which alternatives form rows, and criteria (objectives) form the columns (Yıldırım and Önder, 2014).

2.3.1. Ratio Method

The normalization process, which is performed by dividing the criteria by the square root of the sum of the squares of each alternative and where i=1,2,. . . ,m is the number of alternatives and j=1,2,. . . ,n is the number of criteria, is shown in Eq. (22).

$$x_{ij}^{*} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^{2}}} \tag{22}$$

x_{ij}^{*} is used as the normalized value of alternative i for criterion j. $x_{ij}^{*} \in [0,1]$. After the normalization process is completed, the criteria in the prepared table are determined and collected according to their maximum or minimum status, and the collected minimum values are subtracted from the collected maximum values.

$$y_{i^{*}} = \sum_{j=1}^{g} x_{ij^{*}} - \sum_{j=g+1}^{n} x_{ij^{*}} \tag{23}$$

2.3.2. Reference-Point Approach

In this approach, in addition to the ratio method, reference points (r_j'**s**) are determined for each criterion, which are maximum points for maximization and minimum points for minimization. Subsequently, the distances of these determined points to each x_{ij}^{*} are calculated.

$$r_j - x_{ij^{*}} \tag{24}$$

Where
i=1,2,. . . ,m number of alternatives,
j=1,2,. . . ,n number of criteria,
x_{ij}^{*}, The normalized value of alternative i in criterion j,
r_j, reference point of the criterion j

3. APPLICATION

3.1. Determining Alternatives and Criteria to be Used

This study aimed to evaluate the quality of e-commerce sites operating for the sale of second-hand clothes in Turkey using various multi-criteria decision-making methods. For this purpose, first, the trading sites in Turkey that deal with second-hand goods were researched to identify alternatives. The existence of many sites on the subject has supported the work in this field. The alternatives to be used in the content of this study were three e-commerce sites that mainly focused on women's clothing. These sites are "ModaCruz.com", "Dolap.com" and "Gardrops.com" respectively.

After the alternatives were identified, the opportunities offered by these sites and the opportunities they provided were investigated on the basis of quality criteria. During this research, a site called "Şikayetvar", which is frequently visited by internet users, was used. On this site, users can express one or more dissatisfaction related to the area they wish to explore. The three e-commerce sites that will be discussed within the scope of this study were researched, and the complaints of the users about the mentioned sites were evaluated. These complaints were then listed for use as evaluation criteria. The listed criteria are also classified in detail according to criteria in previous studies on this field.

A detailed view of the criteria determined from the reviews is provided in Table 3.

As shown in Table 3, there are many criteria stated by customers regarding the three e-commerce sites. Since using all these criteria in the evaluation phase will create implementation difficulties, the following 12 criteria were determined to be used in practice. The notation C here means criterion.

Table 3. Criteria list stated by customers on Şikayetvar website

Websites Criteria	Gardrops.com	ModaCruz.com	Dolap.com	Matching Criterion from Literature Reviews	Reference
Product Return Issues	+	+	+	Return Policy	Wang ve Huarng (2002), Ramanathan (2010), Ömürbek & Şimşek (2014).
Accessibility	+	+	+	Accessibility	Cebi (2013), Negash et al. (2003), Nilashi et al. (2012), Parasuraman et al. (2005), Wang & Huarng (2002), Ramanathan (2010), Ecer (2014), Alptekin et al. (2015), Kang et al. (2016), İlkbahar & Cebi (2017), Jiang et al. (2022).
After Sales Customer Support	+	+	+	Accessibility	
Support line problems	+	+	+	Customer Support Customer Service	Wang & Huarng (2002), Dündar et al. (2007), Ramanathan (2010), Dey et al. (2015), İlkbahar & Cebi (2017), Kahraman et al. (2018), Li & Sun (2020), Rekik (2021), Mohamed (2024)
Not Canceling The Order	+		+	System Quality	İçtenbaş and Rouyendegh (2012), Vatansever and Akgul (2014), Rouyendegh et al. (2019)
Shipping Fee	+	+		Service Quality	İçtenbaş & Rouyendegh (2012), Vatansever & Akgul (2014), Rouyendegh et al. (2019), Mohamed (2024)
Fake Product	+	+	+	Accuracy Trust Reliability	Janda et al. (2002), Nilashi et al. (2012), Cebi (2013), Alptekin et al. (2015), Li & Sun (2020), Mohamed (2024)
Defective and missing products	+	+	+	Trust Relevance	Wang & Huarng (2002), DeLone & McLean (2003), Lee & Kozar (2006), Ecer (2014).
Shopping Security	+			Security	Parasuraman et al. (2005), Sun & Lin (2009), Ramanathan (2010), Alptekin et al. (2015), Kaya (2010), Aydın & Kahraman (2012), Santouridis et al. (2012), Dey et al. (2015), Kang et al. (2016), Li & Sun (2020)
Disorders in the System	+	+	+	System Quality	İçtenbaş & Rouyendegh (2012), Vatansever & Akgul (2014), Rouyendegh et al. (2019)
Receiving Commission	+	+	+	Service Quality	İçtenbaş & Rouyendegh (2012), Vatansever & Akgul (2014), Rouyendegh et al. (2019)
Failure to Ensure Buyer and Seller Rights	+	+	+	Service Quality	İçtenbaş & Rouyendegh (2012), Vatansever & Akgul (2014), Rouyendegh et al. (2019)
IBAN not Accepted	+		+	System Quality	İçtenbaş & Rouyendegh (2012), Vatansever & Akgul (2014), Rouyendegh et al. (2019)

Table 3. Continued

Shipping Problem	+	+	+	System Quality Service Quality	İçtenbaş & Rouyendegh (2012), Vatansever & Akgul (2014), Rouyendegh et al. (2019)
Failure load money into account	+	+	+	System Quality Service Quality	İçtenbaş & Rouyendegh (2012), Vatansever & Akgul (2014), Rouyendegh et al. (2019)
Money deduction on product returns	+	+	+	Return Policy	Wang & Huarng (2002), Ramanathan (2010), Ömürbek & Şimşek (2014).
Insufficient live support		+	+	Online Support (Help)	Cebi (2013), Lee & Kozar (2006)
Lack of trust	+	+		Reliability	Devaraj et al. (2002), Kim & Lim (2001), Negash et al. (2003), Yüksel (2007), Nilashi et al. (2012), Parasuraman et al. (2005), Sun & Lin (2009), Kassim & Abdullah (2010), Ömürbek & Şimşek (2014), Alptekin et al. (2015).
Arbitrary refunds		+	+	Service Quality	İçtenbaş & Rouyendegh (2012), Vatansever & Akgul (2014), Rouyendegh et al. (2019)
Selling products in violation of rules		+	+	Accuracy	Wang & Huarng (2002), DeLone & McLean (2003), Lee & Kozar (2006), Ecer (2014).
Problems encountered in product returns			+	System Quality Service Quality Return Policy	Wang & Huarng (2002), Ramanathan (2010), İçtenbaş & Rouyendegh (2012), Vatansever & Akgul (2014).Ömürbek & Şimşek (2014), Rouyendegh et al. (2019), Li & Sun (2020), Mohamed (2024)

*C*1: Product Return Convenience
*C*2: Support Line Presence (Adequate Live Support)
*C*3: Preventing the Sale of Illegal Products (Fake Products)
*C*4: Shortage of Defective and Missing Product Shipments
*C*5: After Sales Customer Support
*C*6: Solution to Cargo Problem
*C*7: Money is loaded into account in a short period
*C*8: Fewer System Problems
*C*9: Security
*C*10: Accuracy
*C*11: Ease of Use of the Web Page
*C*12: Reputation

3.2. Preparation of the Survey and Data Collection

After the alternatives and criteria were determined, an evaluation survey was prepared in the EXCEL environment, and 12 decision-makers using three e-commerce sites were asked to evaluate both the criteria and the alternative e-

commerce sites among themselves. To help decision makers better understand the purpose of the survey, the information given in Tables 4 and 5 was shared with them before the evaluations.

When comparing criteria individually, only one side of the principal diagonal of the matrix is filled by decision-makers. In common use, the upper part of the prime diagonal is filled in by decision-makers, and the values below the prime diagonal are completed accordingly.

Table 4. Example criterion pairwise comparison matrix

EXAMPLE-1				
QUALITY OF E-COMMERCE SITES FOR SELLING SECOND-HAND CLOTHES	Criterion A	Criterion B	Criterion C	Criterion D
Criterion A	---	Less Important	Equally Important	Absolutely Important
Criterion B		---	Much More Important	Absolutely Unimportant
Criterion C			---	Equally Important
Criterion D				---

The expressions in the matrix given in EXAMPLE-1 are as follows. (Less important) Criterion A is less important than Criterion B when evaluating the quality of e-commerce sites selling second-hand clothes. (Equally important) Criterion A is equally important as Criterion C when evaluating the quality of e-commerce sites selling second-hand clothes. (Much more important) Criterion B is much more important than Criterion C when evaluating the quality of e-commerce sites selling second-hand clothes. (Absolutely unimportant) Criterion A is Absolutely unimportant compared to Criterion D when evaluating the quality of websites selling second-hand clothes.

Table 5. Evaluation of alternative example matrix

EXAMPLE-2				
QUALITY OF E-COMMERCE SITES FOR SELLING SECOND-HAND CLOTHES	Criterion A	Criterion B	Criterion C	Criterion D
Dolap	Poor	Good	Good	Very Poor
Moda Cruz	Good	Excellent	Excellent	Good
Gardrops	Average	Good	Average	Good

The expressions in the matrix given in EXAMPLE-2 are expressed as follows. “Dolap” second-hand e-commerce site is Poor in Criterion A, Good in Criterion B, Good in Criterion C, and Very Poor in Criterion D. “ModaCruz” second-hand e-commerce site is Good in terms of Criterion A, Excellent in Criterion B, Excellent in Criterion C, and Good in Criterion D. “Gardrops” second-hand e-commerce site is Average in terms of Criterion A, Good in terms of Criterion B, Average in terms of Criterion C, and Good in terms of Criterion D.

3.3. Importance Weights of Criteria Determined by the AHP Method

In this step, first, the verbal data obtained from the 12 decision-makers were converted into numerical equivalents for pairwise comparisons of the criteria. The numerical equivalents are given in Table 6.

Then, a single matrix is obtained by taking the average of all responses. In the obtained pairwise comparison matrix, row totals were taken, and normalization operations were carried out with the help of Eq. (1), and the pairwise comparison normalized matrix in Table 7 was created.

After the criterion pairwise comparison normalized matrix was obtained, the priority vector was obtained by taking the row averages of the normalized matrix. The priority vector obtained in this study are listed in Table 8.

In order to perform the consistency test, first, the normalized pairwise comparison matrix must be multiplied by the priority vector. The consistency calculations were carried out following Eq.s (2.4)-(2.7) and the consistency ratio was calculated as -0.614. This means that the matrix is consistent, and the results obtained through this matrix are usable. As a result, the priority vector of the criteria calculated by the AHP method (in other words the importance weights of the criteria) can be used in the VIKOR and MOORA methods.

Table 6. Scales used to evaluate criteria

VERBAL EVALUATION	NUMERICAL EQUIVALENT
Absolutely Unimportnt	1/7
Much Less Important	1/5
Less Important	1/3
Equally Important	1
More Important	3
Much More Important	5
Absolutely Important	7

Table 7. Criterion pairwise comparison normalized matrix

QUALITY OF E-COMMERCE SITES FOR SELLING SECOND-HAND CLOTHES	Product Return Convenience	Support Line Presenc(Adequate Live Support)	Preventing the Sale of Illegal Products (Fake Products)	Shortage of Defective and Missing Product Shipments	After Sales Customer Support	Solutions to Cargo Problems	Money is loaded into the account in a short period	Fewer System Problems	Security	Accuracy	Ease of Use of the Web Page	Reputation
Product Return Convenience	0,044	0,133	0,084	0.067	0.108	0.056	0.094	0.076	0.089	0.115	0.063	0.124
Support Line Presence (Adequate Live Support)	0.123	0.048	0.066	0.089	0.077	0.071	0.098	0.096	0.091	0.052	0.077	0.078
Preventing the Sale of Illegal Products (Fake Products)	0.090	0.076	0.041	0.076	0.161	0.092	0.128	0.135	0.060	0.104	0.109	0.055
Shortage of Defective and Missing Product Shipments	0.060	0.086	0.064	0.049	0.126	0.075	0.101	0.122	0.038	0.058	0.083	0.085
After Sales Customer Support	0.093	0.071	0.129	0.120	0.052	0.056	0.063	0.077	0.068	0.042	0.064	0.055
Solutions to Cargo Problems	0.046	0.063	0.071	0.068	0.054	0.054	0.104	0.075	0.088	0.062	0.092	0.071
Money is loaded into the account in a short period	0.096	0.108	0.122	0.114	0.075	0.129	0.043	0.085	0.065	0.057	0.095	0.080
Fewer System Problems	0.068	0.092	0.112	0.120	0.080	0.081	0.074	0.050	0.048	0.072	0.061	0.076
Security	0.084	0.094	0.053	0.040	0.075	0.101	0.060	0.051	0.047	0.113	0.137	0.127
Accuracy	0.105	0.051	0.089	0.059	0.045	0.069	0.051	0.073	0.109	0.048	0.119	0.121
Ease of Use of the Web Page	0.075	0.100	0.122	0.111	0.090	0.135	0.112	0.083	0.174	0.156	0.037	0.081
Reputation	0.115	0.078	0.048	0.087	0.060	0.080	0.073	0.079	0.125	0.123	0.062	0.048

Table 8. Priority Vector

		Priority Vektor
$C1$	Product Return Convenience	0.088
$C2$	Support Line Presence (Adequate Live Support)	0.080
$C3$	Preventing the Sale of Illegal Products (Fake Products)	0.094
$C4$	Shortage of Defective and Missing Product Shipments	0.079
$C5$	After Sales Customer Support	0.074
$C6$	Solutions to Cargo Problems	0.071
$C7$	Money is loaded into the account in a short period	0.089
$C8$	Fewer System Problems	0.078
$C9$	Security	0.082
$C10$	Accuracy	0.078
$C11$	Ease of Use of the Web Page	0.106
$C12$	Reputation	0.081
TOTAL		1.000

3.4. Evaluation of Alternatives using the VIKOR Method

To apply the VIKOR method, a decision matrix must first be created. For this purpose, the verbal values that decision-makers assign to alternatives by evaluating them according to the relevant criteria must be converted into numerical equivalents and the average of 12 decision-makers must be taken, as in the AHP method. The scale used to evaluate alternatives is shown in Table 9, and the resulting decision matrix is shown in Table 10.

Table 9. Scales used to evaluate alternatives

VERBAL EVALUATION	NUMERICAL EQUIVALENT
VERY POOR	1
POOR	2
AVERAGE	3
GOOD	4
EXCELLENT	5

Table 10. Decision matrix

QUALITY OF E-COMMERCE SITES FOR SELLING SECOND-HAND CLOTHES	Product Return Convenience	Support Line Presence (Adequate Live Support)	Preventing the Sale of Illegal Products (Fake Products)	Shortage of Defective and Missing Product Shipments	After Sales Customer Support	Solutions to Cargo Problems	Money is loaded into the account in a short period	Fewer System Problems	Security	Accuracy	Ease of Use of the Web Page	Reputation
Dolap	2.846	3.154	3.077	3.308	3.231	3.615	3.923	3.385	3.769	3.308	3.846	3.750
ModaCruz	2.769	2.692	2.615	2.846	2.846	3.000	2.923	2.846	2.846	3.077	3.154	3.250
Gardrops	2.462	2.692	2.692	3.000	2.769	2.923	3.154	3.077	3.000	3.000	3.154	3.333

After creating the decision matrix, the best and worst values for each criterion are determined. Here, it is important to determine whether the criteria have benefits or costs. Therefore, is the criterion a desirable or undesirable situation?

The 12 criteria considered in this study have benefits, and accordingly, the ideal solution values in Table 11 were determined by taking Eq. (10).

Table 11. Ideal Solution Values

QUALITY OF E-COMMERCE SITES FOR SELLING SECOND-HAND CLOTHES	Product Return Convenience	Support Line Presence (Adequate Live Support)	Preventing the Sale of Illegal Products (Fake Products)	Shortage of Defective and Missing Product Shipments	After Sales Customer Support	Solutions to Cargo Problems	Money is loaded into the account in a short period	Fewer System Problems	Security	Accuracy	Ease of Use of the Web Page	Reputation
f_j^+	2.846	3.154	3.077	3.308	3.231	3.615	3.923	3.385	3.769	3.308	3.846	3.750
f_j^-	2.462	2.692	2.615	2.846	2.769	2.923	2.923	2.846	2.846	3.000	3.154	3.250

In the next stage, decision matrix normalization was performed depending on the ideal solution values. Subsequently, the weighted normalized decision matrix in Table 12 was obtained by considering the weights calculated using the AHP method.

Table 12. Weighted normalized decision matrix

QUALITY OF E-COMMERCE SITES FOR SELLING SECOND-HAND CLOTHES	Product Return Convenience	Support Line Presence (Adequate Live Support)	Preventing the Sale of Illegal Products (Fake Products)	Shortage of Defective and Missing Product Shipments	After Sales Customer Support	Solutions to Cargo Problems	Money is loaded into the account in a short period	Fewer System Problems	Security	Accuracy	Ease of Use of the Web Page	Reputation
Dolap	0	0	0	0	0	0	0	0	0	0	0	0
ModaCruz	0.018	0.080	0.094	0.079	0.062	0.063	0.089	0.078	0.082	0.059	0.106	0.081
Gardrops	0.088	0.080	0.078	0.053	0.074	0.071	0.068	0.044	0.068	0.078	0.106	0.068

Then, considering Eq. (18) and Eq. (19), compromise solution values were calculated, and alternative e-commerce sites were ranked in terms of quality. The results obtained are stated in table 13.

Table 13. Ranking of alternatives using the VIKOR method

	Q_i	Ranking
Dolap	0	1
ModaCruz	1	3
Gardrops	0,9926	2

The results in Table 13 clearly show that the e-commerce site found to be of the highest quality by decision makers is "Dolap". Gardrops ranks second. ModaCruz was ranked last in terms of quality.

3.5. Evaluation of Alternatives using the MOORA Method

As in the other methods, the decision matrix in the MOORA method must first be created. The decision matrix shown in Table 10 is the main input of the proposed method. To normalize the decision matrix, the square root of the sum of the squares of the values of the alternatives under each criterion was taken; the value in the relevant column was divided by this value, and a normalized decision matrix was created. The relevant matrix is shown in Table 14.

Table 14. The normalized decision matrix

QUALITY OF E-COMMERCE SITES FOR SELLING SECOND-HAND CLOTHES	Product Return Convenience	Support Line Presence (Adequate Live Support)	Preventing the Sale of Illegal Products (Fake Products)	Shortage of Defective and Missing Product Shipments	After Sales Customer Support	Solutions to Cargo Problems	Money is loaded into the account in a short period	Fewer System Problems	Security	Accuracy	Ease of Use of the Web Page	Reputation
Dolap	0.609	0.638	0.634	0.625	0.631	0.653	0.674	0.628	0.674	0.610	0.653	0.627
ModaCruz	0.593	0.545	0.539	0.537	0.556	0.542	0.502	0.528	0.509	0.567	0.536	0.544
Gardrops	0.527	0.545	0.555	0.567	0.541	0.528	0.542	0.571	0.536	0.553	0.536	0.558

In the next stage, the weighted normalized decision matrix shown in Table 15 was obtained by considering the weights calculated using the AHP method.

Table 15. Weighted normalized decision matrix

QUALITY OF E-COMMERCE SITES FOR SELLING SECOND-HAND CLOTHES	Product Return Convenience	Support Line Presence (Adequate Live Support)	Preventing the Sale of Illegal Products (Fake Products)	Shortage of Defective and Missing Product Shipments	After Sales Customer Support	Solutions to Cargo Problems	Money is loaded into the account in a short period	Fewer System Problems	Security	Accuracy	Ease of Use of the Web Page	Reputation
Dolap	0.053	0.051	0.060	0.049	0.047	0.046	0.060	0.049	0.055	0.048	0.069	0.051
ModaCruz	0.052	0.044	0.051	0.042	0.041	0.038	0.045	0.041	0.042	0.044	0.057	0.044
Gardrops	0.046	0.044	0.052	0.045	0.040	0.037	0.048	0.044	0.044	0.043	0.057	0.045

After obtaining the weighted normalized matrix, the reference points and distances of each alternative to these reference points were calculated, and the overall distances were obtained by taking the row totals. These distance values were also considered when ranking alternatives. The obtained results are shown in Table 16.

Table 16. Ranking of alternatives using the MOORA method

	d_i	Ranking
Dolap	0,000	1
ModaCruz	0,097	3
Gardrops	0,092	2

As a result of all these steps, the same results were obtained using the VIKOR method, and it was calculated that the "Dolap" e-commerce site was of higher quality in light of the considered criteria.

4. DISCUSSION AND CONCLUSION

Second-hand shopping is becoming increasingly popular and widespread across digital channels, especially in the context of consumer-to-consumer e-commerce. This initiative, which aims to extend the product life cycle, can include other options, such as rental, exchange, redesign, and repair, in addition to simply selling used products. From an environmental perspective, such alternative forms of consumption are needed to simultaneously reduce the use of material resources, meet consumers' need for innovation, and respond to the revenue demands of a global industry.

The sale of second-hand clothes on websites has increased considerably. The large number of sites for this purpose causes confusion among consumers regarding which site to shop from. Determining the most suitable website in terms of many criteria, such as price, security, after-sales support, and accessibility, is a problem that consumers need to solve. In this study, the quality of three e-commerce sites operating for the sale of second-hand clothes in Turkey was evaluated using the AHP, VIKOR, and MOORA methods. The criteria most taken into consideration when shopping on e-commerce sites were determined from studies in the literature and from a site called "Şikayetvar", which is frequently visited by internet users. The weights of the determined criteria were then calculated using the AHP method. Evaluations were performed among the websites using these weights in the VIKOR and MOORA methods.

According to the evaluation results, "ease of use of the web page", "preventing the sale of illegal products (fake products)", "money being deposited into the account in a short time", "product return convenience" and "security" are the five criteria that users most take into consideration, respectively. The "solution for cargo problem" is often considered the least significant criterion by decision-makers. These details are clearly shown in Fig. 2.

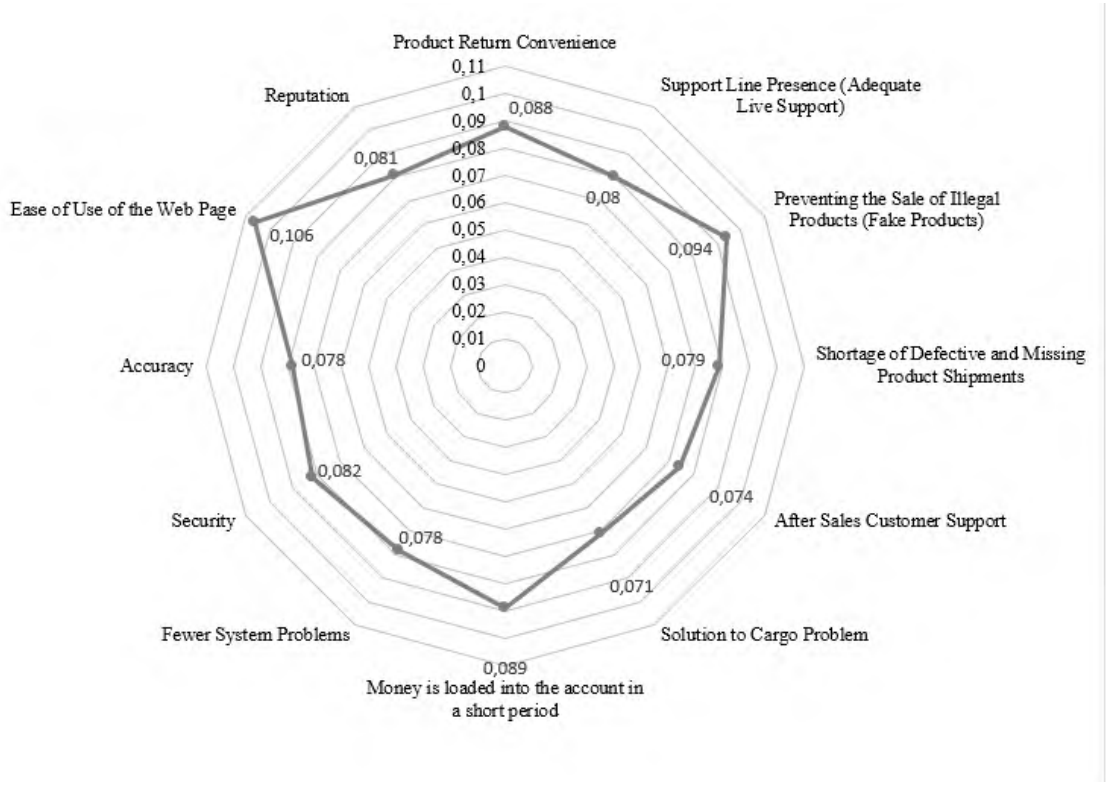


Figure 2. Importance Weights of Criteria

In addition, it was determined that the Dolap website was of higher quality than the ModaCruz and Gardrops websites considering 12 criteria and evaluations of only 12 decision makers. This is because the Dolap website exhibits the lowest value among the distances calculated using both the VIKOR and MOORA methods, as shown in Fig. 3. These findings guide how second-hand clothing-selling websites should improve their service and system quality. In today's world, where online shopping is increasingly favored due to considerations of time efficiency and speed, it is unavoidable that users seek to conduct their transactions with ease. Consequently, the criterion of " ease of use of the website"

carries the highest weight in this context. The issue of "fake products," which poses significant challenges in the sale of second-hand clothes, ranks second in importance. The primary concern is that inflated prices may be charged for items that do not genuinely belong to the specified brand, thereby deceiving buyers. As a result, users anticipate that the websites from which they purchase products will provide assurances and implement measures to prevent the sale of illegal (fake) products.

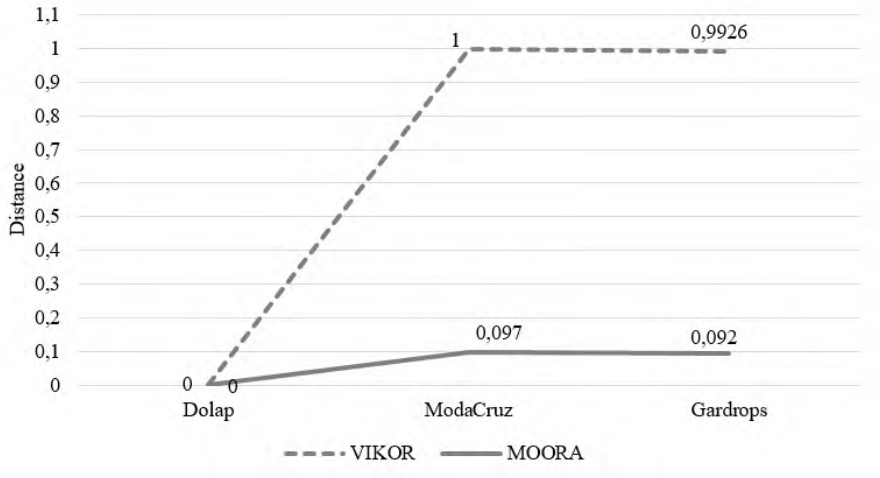


Figure 3. Distance Values of Alternatives

It is important to note that the findings of this study are derived from the personal evaluations of only 12 decision-makers. Consequently, the performance rankings of the websites identified in this study may vary depending on the numbers and characteristics of the decision-makers. This limitation is a significant aspect of the research. This study reflects a specific time frame, a particular user group and a single country context. Nonetheless, the individuals interviewed were selected for their impartiality and lack of any material or ethical interests or relationships with the websites under evaluation, and their opinions were presented in the study without any modification. It is also essential to emphasize that the primary objective of this study is to demonstrate the application of various multi-criteria decision-making methods to address this particular problem.

In this study, it is assumed that decision-makers possess the same knowledge and experience, an assumption critiqued in the literature as a limitation of MCDM methods. Scholars argue that it is unrealistic for decision-makers to expect consensus on knowledge, experience, and perspectives. Consequently, assigning equal weights to individuals may undermine the reliability and effectiveness of the outcomes. (Koksalmis and Kabak, 2019; Ayasrah and Turan, 2021; Škoda et al., 2021). Therefore, determining the weights of decision-makers using different methods will be beneficial in future studies.

The topic discussed in this study contributes to the literature, especially because it is related to sustainable fashion, which is currently highly emphasized. In addition, since this study includes more than one method, it proposes an integrated solution to the literature. In addition, a holistic perspective was provided by matching the real user comments received from the "Şikayetvar" website with the evaluation criteria described in the literature.

It is possible to update considering different methods and using more evaluators in future studies. The use of evaluation methods, especially those involving fuzzy datasets, will increase the reliability of the findings. It is also possible to update the criteria considered.

Peer Review: Externally peer-reviewed.

Conflict of Interest: The author have no conflict of interest to declare.

Grant Support: The author declared that this study has received no financial support.

ORCID ID of the author

Sinem Büyüksaatçı Kiriş 0000-0001-7697-3018

REFERENCES

Alptekin, N., Hall, E. E., & Sevim, N. (2015). Evaluation of websitewebsites quality using fuzzy TOPSIS method. *International Journal of Academic Research in Business and Social Sciences*, 5(8), 221-242. http://dx.doi.org/10.6007/IJARBSS/v5-i8/1778

Ayasrah, O., & Turan, F. M. (2021). A review of multi-criteria decision-making methods using application of variable weight theory and IF-TOPSIS-EF. *Recent Trends in Manufacturing and Materials Towards Industry 4.0: Selected Articles from iM3F 2020*, Malaysia, 13-24. https://doi.org/10.1007/978-981-15-9505-9_2

Aydin, S., & Kahraman, C. (2012). Evaluation of e-commerce website quality using fuzzy multi-criteria decision making approach. *IAENG International Journal of Computer Science*, 39(1). Retrieved from https://www.iaeng.org/IJCS/issues_v39/issue_1/IJCS_39_1_07.pdf

Borusiak, B., Szymkowiak, A., Horska, E., Raszka, N., & Żelichowska, E. (2020). Towards building sustainable consumption: A study of second-hand buying intentions. *Sustainability, 12*(3), 875. https://doi.org/10.3390/su12030875.

Cebi, S., 2013, A quality evaluation model for the design quality of online shopping websites. *Electronic Commerce Research and Application*s, 12(2), 124-135. https://doi.org/10.1016/j.elerap.2012.12.001

Çelikbilek, Y., (2018). *Çok Kriterli Karar Verme Yöntemleri, Açıklamalı ve Karşılaştırmalı Sağlık Bilimleri Uygulamaları ile*, Nobel Yayınevi, ANKARA,978-605-320-857-0.

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of management information systems*, 19(4), 9-30. https://doi.org/10.1080/07421222.2003.11045748

Dengin, H. P. (2011). İkinci El Giysi Pazarları ve Türkiye'deki Durumu. *Akdeniz Sanat*, 4(8), 0-0. Retrieved from https://dergipark.org.tr/tr/pub/akdenizsanat/issue/27655/291503

Devaraj, S., Fan, M., & Kohli, R. (2002). Antecedents of B2C channel satisfaction and preference: validating e-commerce metrics. *Information systems research*, 13(3), 316-333. https://doi.org/10.1287/isre.13.3.316.77

Dey, S., Jana, B., Gourisaria, M. K., Mohanty, S. N., & Chatterjee, R. (2015). Evaluation of Indian B2C E-shopping websites under multi criteria decision-making using fuzzy hybrid technique. International Journal of Applied Engineering *Research*, 10(9), 24551-24580.

Dündar, S., Ecer, F., & Özdemir, Ş. (2007). Fuzzy Topsis Yöntemi ile Sanal Mağazalarin Web Sitelerinin Değerlendirilmesi. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 21(1), 287-305. Retrieved from https://dergipark.org.tr/tr/pub/atauniiibd/issue/2691/35409#article_cite

Ecer, F. (2014). A hybrid banking websites quality evaluation model using AHP and COPRAS-G: a Turkey case. *Technological and Economic Development of Economy*, 20(4), 758-782. https://doi.org/10.3846/20294913.2014.915596

Içtenbaş, B. D., & Rouyendegh, B. D. (2012, April). A fuzzy AHP for evaluation of ecommerce websites performance. In *Proceedings of the 3rd International Conference on Information Management and Evaluation: ICIME 2012* (pp. 16-17).

Ilbahar, E., & Cebi, S. (2017). Classification of design parameters for E-commerce websites: A novel fuzzy Kano approach. *Telematics and Informatics*, 34(8), 1814-1825. https://doi.org/10.1016/j.tele.2017.09.004

Janda, S., Trocchia, P. J., & Gwinner, K. P. (2002). Consumer perceptions of Internet retail service quality. *International journal of service industry management*, 13(5), 412-431. https://doi.org/10.1108/09564230210447913

Jiang, H., Lin, Y., Luo, X., & Shao, T. (2022). Understanding the selection of cross-border import e-commerce platforms through the DANP and TOPSIS techniques: a multi-study analysis. *Journal of Global Information Technology Management*, 25(1), 26-53. https://doi.org/10.1080/1097198X.2021.2022397

Kahraman, C., Onar, S. Ç., & Öztayşi, B. (2018). B2C marketplace prioritization using hesitant fuzzy linguistic AHP. International Journal of Fuzzy Systems, 20(7), 2202-2215. https://doi.org/10.1007/s40815-017-0429-4

Kang, D., Jang, W., & Park, Y. (2016). Evaluation of e-commerce websites using fuzzy hierarchical TOPSIS based on ES-QUAL. *Applied Soft Computing*, 42, 53-65. https://doi.org/10.1016/j.asoc.2016.01.017

Karadayı-usta, S. & Kadaifci, C. (2022). Sürdürülebilir moda için bütünleşik bir yaklaşım önerisi: İkinci el kıyafet alışveri-şinde etkili olan kriterlerin anlaşılması ve analizi. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 37 (2), 873-888. https://doi.org/10.17341/gazimmfd.945327

Kassim, N., & Abdullah, N. A. (2010). The effect of perceived service quality dimensions on customer satisfaction, trust, and loyalty in e-commerce settings: A cross cultural analysis. *Asia pacific journal of marketing and logistics*, 22(3), 351-371. https://doi.org/10.1108/13555851011062269

Kassinen, I., & Koivumäki, T. (2019). *Customer value propositions in second-hand clothing e-commerce.* (Master's Thesis, University of Oulu). Retrieved from http://jultika.oulu.fi/files/nbnfioulu-201906052368.pdf

Kaya, T. (2010). Multi-attribute evaluation of website quality in E-business using an integrated fuzzy AHPTOPSIS methodology. *International Journal of Computational Intelligence Systems*, 3(3), 301-314. https://doi.org/10.1080/18756891.2010.9727701

Kim, S. Y., & Lim, Y. J. (2001). Consumers' perceived importance of and satisfaction with internet shopping. *Electronic Markets*, 11(3), 148-154. https://doi.org/10.1080/101967801681007988

Koksalmis, E., & Kabak, Ö. (2019). Deriving decision makers' weights in group decision making: An overview of objective methods. *Information Fusion*, 49, 146-160. https://doi.org/10.1016/j.inffus.2018.11.009

Lee, Y., & Kozar, K. A. (2006). Investigating the effect of website quality on e-business success: An analytic hierarchy process (AHP) approach. *Decision support systems*, 42(3), 1383-1401. https://doi.org/10.1016/j.dss.2005.11.005

Li, R., & Sun, T. (2020). Assessing factors for designing a successful B2C E-Commerce website using fuzzy AHP and TOPSIS-Grey methodology. *Symmetry*, 12(3), 363. https://doi.org/10.3390/sym12030363

Mohamed, M. (2024). BHARAT Decision Making Model: Harness an Innovative MCDM Methodology for Recommending Beneficial E-Commerce Website. *Multicriteria Algorithms with Applications*, 2, 53-64. https://doi.org/10.61356/j.mawa.2024.26761

Napompech, K., & Kuawiriyapan, S. (2011). Factors affecting the second hand clothes buying decision of Thai women. *International Journal of Arts & Sciences*, 4(9), 201-209. https://www.proquest.com/openview/17a719c3d020d5425cd332203160833a/1?pq-origsite=gscholar&cbl=626342

Negash, S., Ryan, T., & Igbaria, M. (2003). Quality and effectiveness in web-based customer support systems. *Information & management*, 40(8), 757-768. https://doi.org/10.1016/S0378-7206(02)00101-5

Nieminen, R. (2016). Consumer perceptions and attitudes towards second-hand C2C online buying. (Master's Thesis, Lappeenranta Unıversıty of Technology). Retrieved from https://lutpub.lut.fi/bitstream/handle/10024/129910/Pro_gradu_Nieminen.pdf?sequence=2&isAllowed=y

Nilashi, M., Bagherifard, K., Ibrahim, O., Janahmadi, N., & Ebrahimi, L. (2012). Ranking parameters on quality of online shopping websites using multi-criteria method. *Research Journal of Applied Sciences, Engineering and Technology*, 4(21), 4380-4396. Retrieved from https://www.semanticscholar.org/paper/Ranking-Parameters-on-Quality-of-Online-Shopping-Nilashi-Bagherifard/1282e9d9506950c9001e87029de6fcac77163b65

Ömürbek, N., & Şimşek, A. (2014). Analitik hiyerarşi süreci ve analitik ağ süreci yöntemleri ile online alışveriş site seçimi. *Yönetim ve Ekonomi Araştırmaları Dergisi*, 12(22), 306-327. Retrieved from https://dergipark.org.tr/en/pub/yead/issue/21809/234450

Parasuraman, A., Zeithaml, V. A., & Malhotra, A. (2005). ES-QUAL: A multiple-item scale for assessing electronic service quality. Journal of service research, 7(3), 213-233. https://doi.org/10.1177/1094670504271156

Ramanathan, R. (2010). E-commerce success criteria: determining which criteria count most. *Electronic Commerce Research*, 10(2), 191-208. https://doi.org/10.1007/s10660-010-9051-3

Rekik, R. (2021). An integrated fuzzy anp-topsis approach to rank and assess e-commerce web sites. In *Hybrid Intelligent Systems: 19th International Conference on Hybrid Intelligent Systems (HIS 2019)* held in Bhopal, India, December 10-12, 2019 19 (pp. 197-209). Springer International Publishing. https://doi.org/10.1007/978-3-030-49336-3_20

Rouyendegh, B. D., Topuz, K., Dag, A., & Oztekin, A. (2019). An AHP-IFT integrated model for performance evaluation of E-commerce web sites. *Information Systems Frontiers,* 21, 1345-1355. https://doi.org/10.1007/s10796-018-9825-z

Santourıdıs, I., Trıvellas, P., & Tsımonıs, G. (2012). Using ES-QUAL to measure internet service quality of e-commerce web sites in Greece. *International Journal of Quality and Service Sciences*, 4(1), 86. https://doi.org/10.1108/17566691211219751

Škoda, M., Flegl, M., & Lozano, C. (2021). Fuzzy approach for group decision-making in crisis situations. *Business: Theory and Practice*, 22(1), 180-189. https://doi.org/10.3846/btp.2021.12148

Sun, C. C., & Lin, G. T. (2009). Using fuzzy TOPSIS method for evaluating the competitive advantages of shopping websites. *Expert Systems with Applications*, 36(9), 11764-11771. https://doi.org/10.1016/j.eswa.2009.04.017

Tóta, Á. (2016). Sustainable clothing consumption-Exploratory study about second hand clothes. (Master's Thesis, Unıversıty of Porto, Faculty of Economy). Retrieved from https://repositorio-aberto.up.pt/bitstream/10216/81997/2/37724.pdf

Vatansever, K., & Akgul, Y. (2014). Applying fuzzy analytic hierarchy process for evaluating service quality of private shopping website quality: A case study in Turkey. *Journal of Business Economics and Finance*, 3(3), 283-301. https://dergipark.org.tr/en/pub/jbef/issue/32410/360451

Wang, M., & Huarng, S. A. (2002). An empirical study of internet store customer post-shopping satisfaction. *Special Issues of Information Systems*, 3(5), 632-638. Retrieved from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=cf6a494fee7809319c2041e2bbc0b3bf76cd2de7

Yaralıoğlu, K., & Köksal, H. (2003). Analytic hierarchy process as a managerial decision tool in the evaluation of new product ideas. *Ege Akademik Bakış Dergisi*, 3(1), 119-137. Retrieved from https://dergipark.org.tr/en/pub/eab/issue/39834/472288

Yüksel, H. 2007. İnternet sitelerinin kalite boyutlarının değerlendirilmesi. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 1, 517-536. Retrieved from https://earsiv.anadolu.edu.tr/xmlui/handle/11421/407

How cite this article

Büyüksaatçı Kiriş, S. (2024). Evaluation of the Quality of Second-Hand Clothes Trading Websites using the VIKOR and MOORA Methods. *Acta Infologica, 8*(2), 157-175. https://doi.org/10.26650/acin.1462978

İSTANBUL UNIVERSITY PRESS

Acta Infologica
acin 2024, 8 (2), 176–187
DOI: 10.26650/acin.1462076

Research Article

Comparative Analysis of Transaction Performance in Different Virtualization Environments

Ali İhsan Candemir[1], Nilgün İncereis[2]

[1]İstanbul Okan University, Computer Engineering, İstanbul, Türkiye
[2]Bartın University, Distance Education Application and Research Center, Bartın, Türkiye

Corresponding author : Ali İhsan Candemir
E-mail : alcandemir@stu.okan.edu.tr

ABSTRACT

Virtualization technologies are increasing in importance day by day. The selection of virtualization software is an important factor to realize efficient use of physical hardware. In this study, the Ubuntu 22.04 LTS operating system was run on VirtualBox, VMware, and Docker, which are widely used virtualization software. Then, the Tiobench, Compress-7zip, C-ray, Smallpt, Tachyon, and OSBench tests were performed on the Ubuntu operating system with the Phoronix Test Suite software, and the test results were analyzed and compared. The results demonstrate that Docker outperformed the other virtualization technologies, although not in every test. Owing to its rapid deployment and efficient use of resources, Docker is suitable for applications that require agility and scalability. However, traditional virtualization technologies, such as VMware, may be more suitable for applications that require high security and extensive resource management. In addition, this study provides information that can guide users in selecting virtualization software.

Keywords: Virtualization, VirtualBox, VMware, Docker, Virtualization Technologies.

Submitted : 01.04.2024
Revision Requested : 31.05.2024
Last Revision Received : 25.09.2024
Accepted : 04.11.2024
Published Online : 26.11.2024

1. INTRODUCTION

Virtualization technologies include a set of technologies and methods that offer the ability to run multiple operating systems simultaneously by creating virtual environments on physical hardware. These technologies offer advantages such as using resources more efficiently, providing flexibility, and isolating workloads (Schlosser et al., 2011). Virtualization is widely used in many fields, such as cloud computing, server consolidation, application testing, and development (Wei et al., 2019). Virtualization technologies enable more efficient use of physical hardware resources by dynamically sharing them between virtual machines. These technologies are hardware independent and therefore easily portable between different physical environments. They offer security advantages by isolating different virtual machines from each other. Due to these advantages, virtualization technologies continue to be widely used today (Robbi et al., 2019). Processing performance is a critical measure of the speed and efficiency of a system or application's operations. Accurately measuring process performance is important for system optimization, resource allocation, and user experience. Fast processing improves user experience and increases user satisfaction. A good process performance leads to more efficient use of resources. Efficient processes can reduce costs by reducing energy and resource consumption. These items emphasize the importance of processing performance (Mongia et al., 2020).

This study describes virtualization environments at hardware, OS, and application level in section 2, and describes the performance metrics and evaluation criteria in section 3. Then, section 4 describes the materials and methods of this study, and section 5 describes the findings. Finally, section 6 presents the discussion and conclusions of this study.

2. VIRTUALIZATION ENVIRONMENTS

Virtualization environments use various technologies and configurations to support diverse workloads and distribute computing resources efficiently. The three levels of virtualization environments are described below.

2.1. Hardware Level Virtualization

Hardware level virtualization is a type of virtualization that enables the creation of virtual machines on physical hardware. This is implemented using a hypervisor or virtualization manager. Hardware-level virtualization allows each virtual machine to behave as if it had an independent operating system. Each virtual machine can run its own operating system and applications; thus, multiple virtual environments can be achieved on a single physical server. This method offers the advantages of effectively optimizing resource usage and providing isolation (Natawiguna et al., 2016; Perez et al., 2008).

2.2. Operating System Level Virtualization

An operating system-level virtualization scheme allows an operating system instance to host multiple isolated environments. In this type of virtualization, virtual environments (e.g., containers) running on a host operating system are run in isolation. Each container can host its own applications and dependencies. OS-level virtualization is known for its lightweight and fast startup/shutdown processes. Containers share the same core operating system but offer an isolated environment (Perez et al., 2008) (Natawiguna et al., 2016; Patil, 2021).

Containers encompass only the instructions and code required for the application to run; thus, they are lightweight compared to alternative approaches, such as virtual machines (VMs) (Brady et al., 2020).

You can find detailed information about why Docker, which makes container technology useful, is preferred in the study of Ayaz et al., 2016. In addition, Docker is widely used in web applications. (Yılmaz et al., 2016)

2.3. Application-Level Virtualization

Application-level virtualization is a form of virtualization that enables an application to run in an isolated environment within itself. This is typically implemented by running an application in a standalone container or running it isolated in a virtual environment. Application-level virtualization allows applications to run in their own independent operating environments, which offers the advantages of effectively using system resources and ensuring application independence (Perez et al., 2008; Natawiguna et al., 2016).

3. PERFORMANCE METRICS AND EVALUATION CRITERIA

To effectively evaluate virtualization environments, it is important to measure and analyze performance metrics. These metrics are used to evaluate different aspects of virtual infrastructure and the impact of virtual machines on physical resources. The most important performance metrics are CPU utilization, memory management, input/output operations, and network performance.

3.1. CPU use

CPU use refers to the system or application load placed on the processor. Performance metrics were used to determine the length of intensive CPU running. These metrics are used to evaluate the overall effectiveness of CPU utilization, identify performance bottlenecks, and determine whether resources are being used efficiently (Ferrari et al., 1972; Syauqi et al., 2023).

CPU-intensive tests such as c-ray and small, performed using the Phoronix test suite are important for evaluating processor use in a system. The overall efficiency of the CPU was evaluated by measuring the effectiveness of the processor performance in terms of computational tasks.

3.2. Memory Management

Memory management is a critical performance metric that evaluates how an application or system uses and manages memory resources. In this section, we focus on memory utilization, memory leaks, cache memory activity, and memory cleaning strategies. Good memory management can improve system stability and performance (Ferrari, 1972; Syauqi et al., 2023; Sudha, 2013).

Memory intensive operations such as Tiobench and compress7-zip, provide important data for memory management. These tests evaluate the efficiency of memory usage and are particularly useful for studying memory management during on-disk compression and decompression operations.

3.3. Input/Output Operations

Input/output (I/O) operations refer to the ability of an application to interact with a disk, network, or other external resources. Factors such as database access, file read/write speed, and asynchrony of I/O operations were evaluated.

bench testing is an effective way to measure I/O performance. The proposed method analyzes the efficiency of I/O operations by evaluating the performance of file system operations. This is especially important for understanding and optimizing disk performance (Ferrari, 1972; Sudha, 2013).

3.4. Network Performance

Network performance is an important metric that measures the effectiveness of an application or system in transmitting data over a network. In this section, the network bandwidth, latency, packet loss rate, reliability, and network traffic management factors are evaluated. Good network performance positively impacts user experience and ensures that applications run quickly and reliably (Sudha, 2013).

The tachyon test provided by the Phoronix test suite evaluates the network communication performance. This test is useful for measuring the system's network performance in terms of network traffic management, bandwidth use, and reliability.

4. MATERIAL AND METHOD

When analyzing performance, it is important to set up a test environment that can simulate real-life conditions. This increases the validity and reliability of the obtained results and provides decision makers with tangible data.

4.1. Virtualization Platforms for Testing

The virtualization platforms we chose to perform the tests were VirtualBox, VMware and Docker. The virtualization platforms were installed on the same machine, and the Ubuntu 22.04 LTS operating system was run on each machine.

The hardware specifications of the computer used in the tests affect the figures in the test results. However, the performance ratios of virtual machines to each other did not change even when they were tested on different hardware.

The hardware information of the computer on which the virtual machines were installed to perform the tests is described as follows.

Hardware Information
Processor: Intel Core i7-6500U @ 2.50 GHz
Motherboard: Lenovo 20EVS01G00
Memory: 2x8GB SODIMM DDR3 Synchronous 1600 MHz (0,6 ns)
Disk: 500 GB SanDisk SDSSDH3

4.2. Workloads and scenarios

The tests performed using the Phoronix Test Suite on the Ubuntu 22.04 LTS operating system were as follows:

Tiobench: The Tiobench test is used to measure file input/output performance. This test is intended to determine the speed of input/output operations in a file system. In other words, it can be used to measure and compare disk performance. Tiobench was used to test factors such as disk read/write speeds, file processing speeds, and delays in data access (Gurjar et al., 2019).

Compress-7zip: The 7-Zip is an archiving program that provides high compression rates and supports various file formats. The Compress-7zip test in the Phoronix Test Suite measures how fast 7-Zip compresses and archives different file types and sizes (Gupta et al., 2017).

C-ray: This test measures processor performance by simulating 3D modeling operations. It is used specifically to evaluate processor computational capabilities. c-ray shows how effectively processor power is used to render visual effects and 3D modeling using ray tracing. This test was used to measure and compare the performance of the processor cores. It is preferred to see the performance differences between different processor models, core counts, and clock speeds (Benchmark, 2024).

Smallpt: This benchmark test is a 3D ray tracing dataset. This test was used to measure the performance of the ray tracing techniques used to render 3D graphics. Ray tracing is used to create realistic visuals. It is used to simulate how light is refracted and reflected. Smallpt creates 3D scenes using these techniques in a simple manner and measures the rendering time of these scenes (Benchmark, 2024). CPU intensive tests such as c-ray and small, which you run with Phoronix-test-suite, are an important way to evaluate processor utilization in the system. The overall efficiency of the CPU was evaluated by measuring the effectiveness of the processor performance in terms of computational tasks.

Tachyon: It is another ray tracing benchmark test is designed to measure the performance of ray tracing algorithms used to render 3D graphics (Benchmark, 2024).

OSBench: This test aims to measure how fast an operating system can perform basic operating system calls and operations, such as file system access. It is used to evaluate operating system functions, such as file processing, memory management, threading, and other basic system calls (Benchmark, 2024).

For example, the implementation stages of the 7zip compression test on Docker:

First, we pull the Ubuntu 22.04 LTS image from the Docker Hub and run a container. Figure 1 shows the installation and operation of the Docker container.

```
→  ~ docker pull ubuntu:22.04
docker run -it --name phoronix-ubuntu ubuntu:22.04
```

Figure 1. Docker container installation and running

Then, we install the required packages. Figure 2 shows the installation of the required packages.

```
root@5b78ee19f44c:/# apt update
apt install -y software-properties-common wget
add-apt-repository ppa:phoronix-test-suite/public
apt update
apt install -y phoronix-test-suite
```

Figure 2. Installation of required packages

Then, we install the 7zip compression test in the PRONONix test suite Figure 3 shows the installation of 7zip.

```
root@5b78ee19f44c:/# phoronix-test-suite install pts/compress-7zip
```

Figure 3. Install 7zip

As illustrated in Figure 4, the 7zip compression test is conducted.

```
root@5b78ee19f44c:/# phoronix-test-suite run pts/compress-7zip
```

Figure 4. Run 7zip compression test

The results are presented in Figure 5.

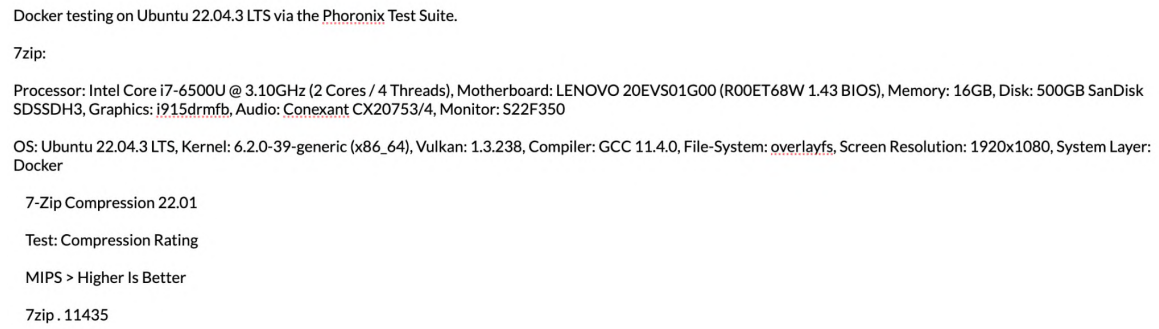


Figure 5. Results

4.3. Data collection and analysis

The test results are presented in tables for clarity and are not presented in the Phoronix Test Suite. In case of inconsistencies in the results, the tests were repeated. Certain variation selections were made in order not to extend the test times. The total test time was measured as 6 h.

Column charts were used to compare and analyze the data made into tables.
Then, the data were analyzed, and comments were made about the usage scenarios.

5. RESULTS

This section presents the results of experimental tests performed in different virtualization environments and their comparative analysis.

5.1. Performance Results of Each Virtualization Environment

The test results are given in the tables below.

5.1.1. Tiobench Test

The tests we apply with the Tiobench test are random write and random-read test.
Test:Size per thread: 64 MB, thread count: 8
The Tiobench test results are presented in Table 1.

Table 1. Tiobench test results

	VirtualBox	VMware	Docker
Random_read (MB/ms)	34,705	37,574	40,638
Random_write (MB/ms)	0032	1159	0255

5.1.2. Compress-7zip Test

7-zip compression and decompression tests were performed. Table 2 lists the results of the Compress-7zip test.

Table 2. Compress-7zip test results

	VirtualBox	VMware	Docker
Compression (MIPS) (million instructions per second)	5557	6495	11435
Decompression (MIPS) (million instructions per second)	5903	5187	8721

5.1.3. C-ray Test

The C-ray and standard ray tracing tests were performed. The results are given in the table below. Test 3 shows the results of the C-ray test.

Table 3. C-ray test results

	VirtualBox	VMware	Docker
Test (ms)	0482	0481	0455

5.1.4. Smallpt Test

The smallpt and standard ray tracing tests were performed. The results are given in the table below. The Smallpt test results are presented in Table 4.

Table 4. Smallpt test results

	VirtualBox	VMware	Docker
Test (ms)	0113	0115	0066

5.1.5. Tachyon Test

The tachyon test and standard ray tracing test were performed. The results are given in the table below. Table 5 presents the Tachyon test results.

Table 5. Tachyon test results

	VirtualBox	VMware	Docker
Test (ms)	2234	1493	1017

5.1.6. OSBench Test

The OSBench test was applied to the thread creation, process creation, and memory allocation tests. Table 6 presents the OSBench test results.

Table 6. OSBench test results

	VirtualBox	VMware	Docker
Create Threads (microsecond/event)	36	57,1	34,7
Create Process (microsecond/event)	85,9	126,2	70,7
Memory allocation (nanosecond/event)	161,9	159,3	136,4

5.2. Comparative Analysis and Findings

The results of each test performed using the Phoronix Test Suite are presented separately using bar graphs for clarity. There were two types of tests we applied in Tiobench tests; random_write and random_read. Because of the random_read test, we found that Docker had the highest read data per second speed among virtual software with a value of 40638 MB/s. In the random_write test, it is noteworthy that the values were very different from each other. Re-tests were performed, and similar results were obtained. This demonstrates that VMware's data-writing speed is superior to other virtualization environments.

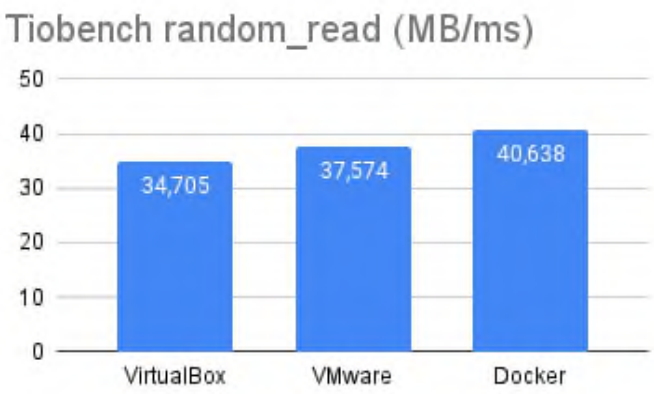


Figure 6. Tiobench random-read test result bar graph.

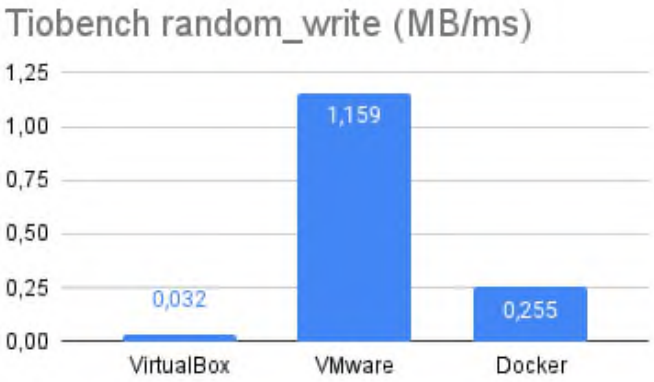


Figure 7. Tiobench random write test result bar graph.

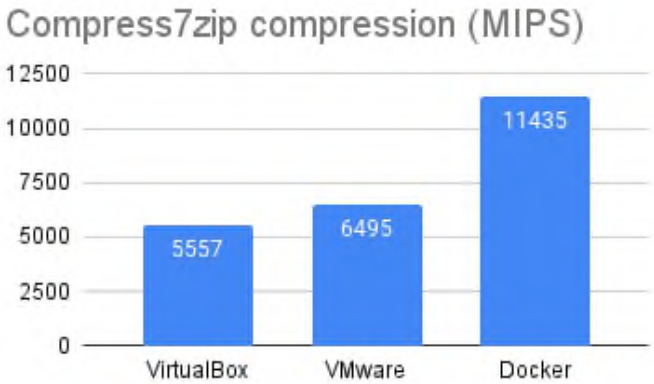


Figure 8. Compress-7zip compression test results are shown as a bar graph.

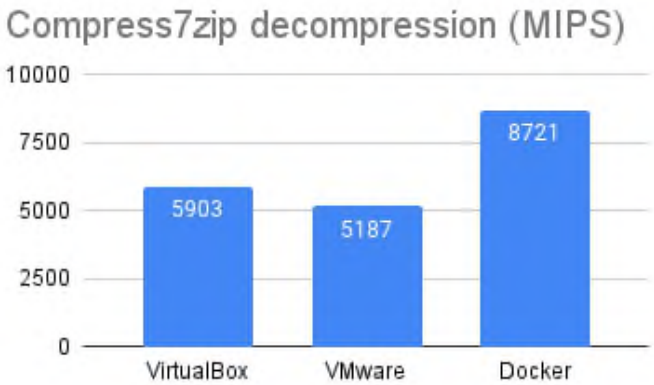


Figure 9. Compress-7zip decompression test results are shown as a bar graph.

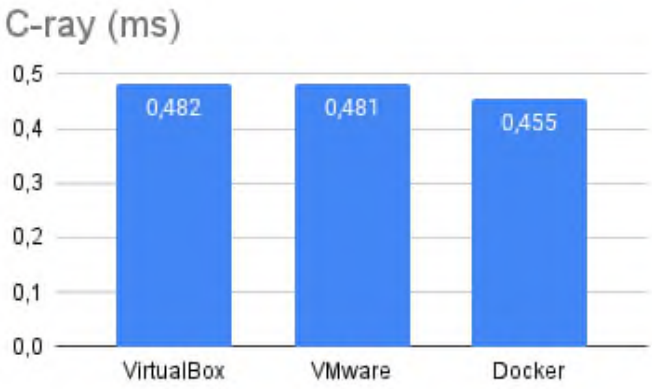


Figure 10. C-ray test results are shown as a bar graph.

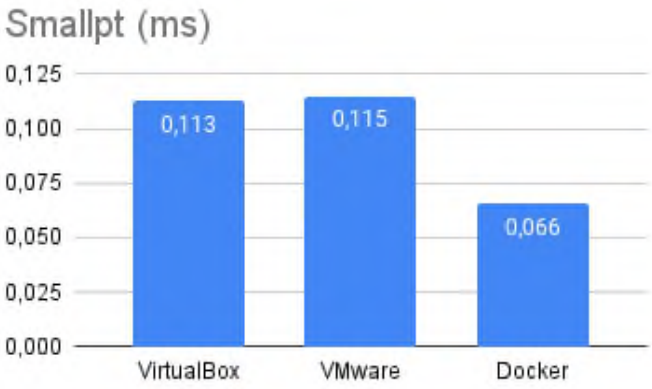


Figure 11. Bar graph of smallpt test results.

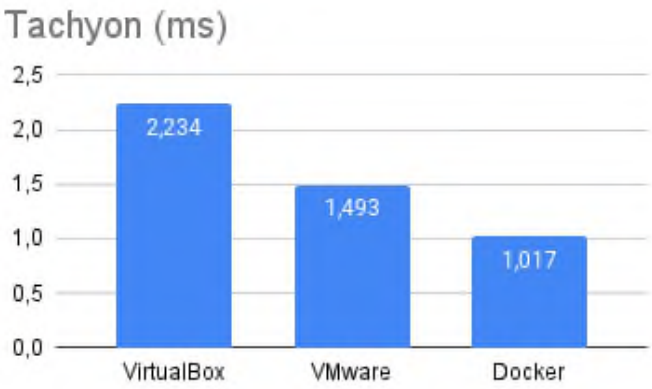


Figure 12. Tachyon test result bar graph.

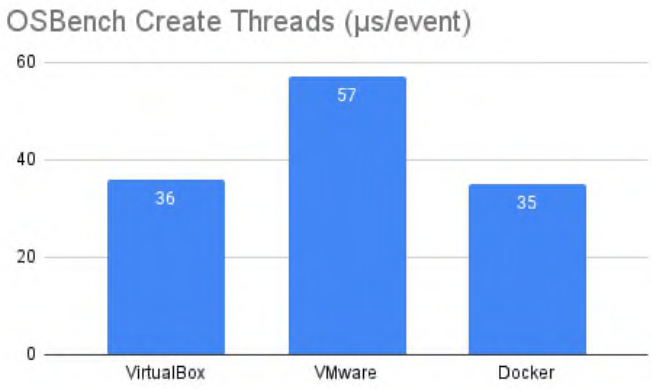


Figure 13. OSBench creates a thread test result bar graph.

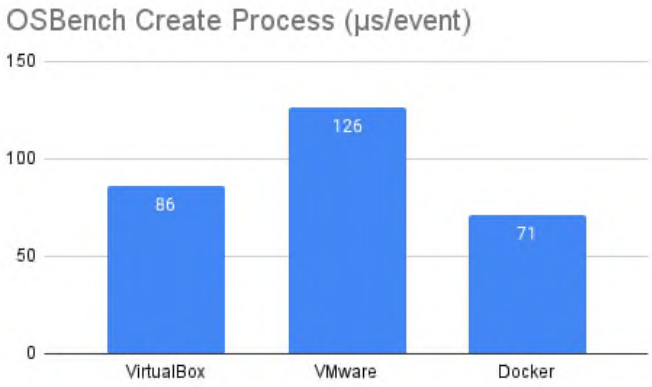


Figure 14. OSBench creates a process test result bar graph.

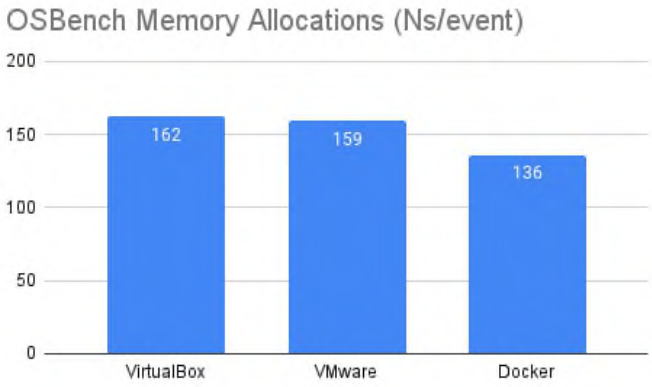


Figure 15. OSBench memory allocation test result bar graph.

In the Tiobench test, the results of which are shown in Figures 6 and 7, it can be seen that Docker demonstrated high performance in the reading test and VMWare in the writing test. This test provides us with information about the disk performance of the virtual machines. In the Compress7zip tests, the results of which are shown in Figures 9 and 10, it can be seen that Docker obtained the highest MIPS (Million instructions per second) values. This demonstrates that Docker outperformed other virtualization environments in tests measuring disk performance. When we look at the C-ray results (Figure 10), which mainly test the processor performance, we see that Docker is superior; however, the results are very close to each other. In the Smallpt (Figure 11) and Tachyon (Figure 12) test results, which focus on processor performance, it is seen that Docker's performance is quite high. However, VMware and VirtualBox demonstrated inconsistent results. This is because these tests measure not only the processor performance. In OSBench's Create Threads (Figure 13) and Create Process (Figure 14) tests; parameters such as processor performance, OS efficiency, memory management, input/output performance, and multi-core usage, were measured. The performance rankings are Docker, VirtualBox, and VMware. Another test from OSBench, Memory Allocations, measures memory-intensive performance such as memory allocation speed, memory-release speed, and memory fragmentation. Although the results (Figure 15) are similar to those of other OSBench tests, VMware appears to perform better than VirtualBox in this test. Docker also demonstrated the highest performance in this test.

Docker takes a different approach from other virtualization technologies. Instead of using hypervisors, as in other virtualization environments, it provides containerization at the operating system level (Pahl et al., 2017; Wang, 2022). This allows Docker containers to be lighter and consume fewer resources (Kaiser et al., 2022; Reis et al., 2022). The fact that MIPS values are superior to those of other virtualization environments suggests that Docker may incur lower overhead (added processing power). This structure of Docker indicates that it will outperform other virtualization technologies in all tests. The results show that this is not the case. We found that Docker outperformed other virtualization environments in certain tests and workloads. In particular, in the Compress-7zip test, we observed that Docker's compression and decompression performance was quite high compared to others. In addition, in the Tiobench test, notably Docker performs quite well in the random read and write operations. However, Docker was not superior in every test. In the Tiobench random write test, Docker outperformed the other virtualization environments.

These results demonstrate that for certain workloads and tests, Docker may have performance advantages; however, these advantages are not always and not in every test scenario. Thus, the selection of a virtualization environment should

depend on the workload, performance requirements, and case. Each can offer different advantages and disadvantages. It is important to choose a virtualization solution that best suits specific requirements (Potdar et al., 2020). VirtualBox and VMware are the most widely used virtualization technologies. Considering the test results, the performance of both virtualization technologies appeared to be close to each other. On average, VMware provides slightly higher performance. When selecting the virtualization technology to use, the use scenario should be considered. Because the usage scenario involves different workloads, performance varies for different workloads.

6. DISCUSSION AND CONCLUSION

Docker's overall success rate in tests was approximately 92%.

Based on the Tiobench test results, Docker achieved the highest read speed of 40,638 MB per second in the random_read test, while it exhibited an average performance of 255 MB in the random_write test. VMwareachievedd the highest write speed (1159 MB. These results demonstrate that Docker exhibits better disk performance, especially in read operations, whereas VMware exhibits better performance in write operations.

In the Compress-7zip test results, Docker achieved the highest values with 11.435 million instructions per second (MIPS) in the compression test and 8.721 MIPS in the decompression test. This demonstrates that Docker is more efficient than other virtualization technologies in terms of disk performance.

Docker obtained the fastest results in the processor-intensive ray tracing test with 455 µs. VMware and VirtualBox gave results of 481 µs and 482 µs respectively. It can be observed that Docker is faster than other processor-based operations, albeit by a small margin.

Dockeroutperformed then other virtualization environments by completing the test in less time than the Smallpt test with 66 µs and Tachyon test with 1017 µs. This demonstrates that Docker is more efficient, especially in CPU-intensive tasks, and provides faster results, especially in 3D graphic operations.

Considering the OSBench test results, Docker obtained the fastest results with 71 µs in the process creation test. VMware and VirtualBox delivered results of 126 and 86 µs, respectively. In addition, Docker demonstrated the fastest performance at 136 ns in the memory allocation test. These findings demonstrate that Docker is more efficient in operational processes such as memory management and process creation.

Docker outperformed other virtualization environments in various performance metrics. In particular, Docker has been observed to be faster and more efficient in tests such as disk reading, compression, ray tracing, and memory management. However, it may have disadvantages, such as security risks and resource isolation. VMware is more suitable for applications that require security and comprehensive resource management and stands out particularly for disk write performance and security requirements. VirtualBox, on the other hand, offered a more balanced performance but fell behind other technologies, especially in terms of disk performance and processor-intensive operations. Users should choose between these virtualization technologies based on their particular needs and application scenarios.

With the increasing use of Docker, security threats targeting Docker containers from malicious hacker teams such as TeamTNT, WatchDog, Kinsing, and Rocke, have increased (Unit42, 2022).

Our findings align with those of Reis and Oliveira (2022), who demonstrated that Docker outperforms traditional virtual machines in terms of start-up time and resource usage. Pahl and Lee (2017) also highlighted that while Docker containers are advantageous for lightweight and scalable deployments, they may not always offer the same level of isolation and security as hypervisor-based solutions. These comparisons underscore the necessity of selecting appropriate virtualization technologies that meet specific workload requirements and performance expectations.

For researchers focusing on virtualization technologies, it is recommended to consider both containerization and traditional virtualization methods depending on the specific case. Docker, with its rapid deployment and efficient resource usage, is ideal for applications requiring agility and scalability. However, for applications needing high security and extensive resource management, traditional virtualization technologies like VMware may be more suitable. Future research should also explore the integration of these technologies to exploit the strengths of both approaches.

This study has certain limitations. First, the tests were conducted using a single type of hardware configuration, which may not represent all possible scenarios. In addition, the focus was on a limited set of performance metrics, thereby eliminating other potentially significant factors, such as long-term stability and maintenance costs. Future studies should consider a broader range of hardware environments and performance indicators.

In future, we plan to consider other factors in the virtualization technology selection process, including security tests. Afterwards, we plan to extend this test, which we conducted using the Ubuntu 22.04 operating system, with different operating systems and other virtualization technologies (Hyper-V, KVM, Parallels).

Peer Review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- A.İ.C., N.İ.; Data Acquisition- A.İ.C., N.İ.; Data Analysis/Interpretation- A.İ.C., N.İ.; Drafting Manuscript- A.İ.C., N.İ.; Critical Revision of Manuscript- A.İ.C., N.İ.; Final Approval and Accountability- A.İ.C., N.İ.; Technical or Material Support- A.İ.C., N.İ.; Supervision- A.İ.C., N.İ.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

ORCID IDs of the authors

Ali İhsan Candemir 0009-0001-0167-7879
Nilgün İncereis 0000-0001-5508-8159

REFERENCES

Ayaz, Ö., Aydın, G. (2016), Uygulama Sanallaştırmada Yeni Bir Yaklaşım: Docker. Retrieved from https://ab.org.tr/ab15/bildiri/312.

Brady, K., Moon, S., Nguyen, T., Coffman, J. (2020), *Docker Container Security in Cloud Computing*, 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), doi: 10.1109/CCWC47524.2020.9031195

Benchmark Tests: https://openbenchmarking.org/tests, 01.01. 2024.

Ferrari, D. (1972). *Workload characterization and selection in computer performance measurement*. Computer 5: 18-24.

Gupta, A., Bansal, A. and Khanduja, V. (2017), *Modern lossless compression techniques: Review, comparison and analysis,* 2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, pp. 1-8.

Gurjar, D. and S. S. Kumbhar, (2019) *A Review on Performance Analysis of ZFS & BTRFS*, 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, pp. 0073-0076, doi: 10.1109/ICCSP.2019.8698103.

Kaiser, S., Haq, M.S., Tosun, A.S., & Korkmaz, T. (2022). Container technologies *for ARM Architecture: A Comprehensive Survey of the State-of-the-Art*, IEEE Access, 10, 84853-84881.

Mongia, V., Sharma, A. (2020 July-August), *Performance Comparison of Virtual Machine Selection Policies in Cloud Environment,* International Journal of Advanced Trends in Computer Science and Engineering, Volume 9, No.4.

Natawiguna, A., & InggrianiLiem, M. M. (2016), *Virtualization Methods for Securing Online Exam*, 2016 International Conference on Data and Software Engineering (ICoDSE), Denpasar, Indonesia, pp. 1-7.

Pahl, C., Brogi, A., Soldani, J., & Jamshidi, P. (2017). Cloud container technologies: *A State-of-the-art review. IEEE* Transactions on Cloud Computing, 7, 677-692.

Patil, S. (2021 May), *Study of container technology with docker.* International Journal of Advanced Research in Science Communication and Technology.

Perez, R., vanDoorn, L., & Sailer, R. (2008 November). *Virtualization and hardware-based security*. IEEE Security and Privacy Magazine 6(5):24 – 31

Potdar, A.M., Naraya, D. G., Kengond, S., Mulla, M.M. (2020), *Performance Evaluation of Docker Container and Virtual Machine*, Procedia Computer Science, 171, 1419-1428.

Reis, D., Piedade, B., Correia, F. F., Dias, J. P., & Aguiar, A., (2022). *Developing Docker and Docker-Compose Specifications: A Developers' Survey.* IEEE Access, vol. 10, pp. 2318-2329.

Robbi, F. A. A., A. B, Prasetijo, and E. D., Widianto ED (2019 January), *Perbandingan Kinerja Block Storage Ceph dan ZFS di Lingkungan Virtual*, Jurnal Teknologi dan Sistem Komputer, vol. 7, no. 1, pp. 7-11.

Schlosser, D., Duelli, M., & Goll, S. (2011). Performance comparison of hardware virtualization *Plaforms*, International Conference on Research in Networking, 393-405.

Sudha, M. (2013), *Performance Analysis of Kernel-Based Virtual Machine*, International Journal of Computer Science and Information Technology.

Syauqi, M. R., Sabrina, N. P., Santikarama, I. (2023), *K-Means Clustering with KNN and mean computation on CPU Benchmark Compilation Data*, Journal of Applied Informatics and Computing, 7(2), pp. 231–239. https://doi.org/10.30871/jaic.v7i2.6491

Unit42 (2022, April 12). I Am Your Defense against Cloud Threats: The Latest Unit 42 Cloud Threat Research. Available online: https://unit42.paloaltonetworks.com/iam-cloud-threat-research (accessed on 20 July 2024).

Wang, W., (2022), *Research on Using Docker Container Technology to Realize Rapid Deployment Environment on Virtual Machine*, 2022 8th Annual International Conference on Network and Information Systems for Computers (ICNISC), Hangzhou, China, pp. 541-544.

Wei, S., Zhang, K., & Tu, B. (2019 June). *HyperBench: A benchmark suite for virtualization capabilities*. Proc. ACM Meas. Anal. Comput. Syst. 3, 2, Article 24, 22 pages.

Yılmaz, E. C., Oktaş, R., & Karabulut, B. (2016). DOCKER İLE WEB UYGULAMASI GELİŞTİRME. Küresel Mühendislik Çalışmaları Dergisi, 3(2), 116-121.

How cite this article

Candemir, A. İ., & İncereis, N. (2024). Comparative Analysis of Transaction Performance in Different Virtualization Environments. *Acta Infologica, 8*(2), 176-187. https://doi.org/10.26650/acin.1462076

İSTANBUL UNIVERSITY PRESS

Acta Infologica
acin 2024, 8 (2), 188–198
DOI: 10.26650/acin.1557985

Research Article

A Machine Learning Approach for Quantifying Academic Misconduct

Almasi S. Maguya[1]

[1]Mzumbe University, Faculty of Science and Technology, Morogoro, Tanzania

Corresponding author : Almasi S. Maguya
E-mail : asmaguya@mzumbe.ac.tz

ABSTRACT

Evidence from the literature continues to reveal the problem of academic misconduct, particularly cheating, among university students. To deal with this problem effectively, a clear understanding of its magnitude is necessary for planning and resource allocation. This paper proposes a machine learning algorithm to quantify the magnitude of academic misconduct among undergraduate students. In this study, cluster analysis was employed with outlier detection and removal. The algorithm was trained on a dataset comprising 678 short texts. Results indicated that over 80% of students engage in the practice of academic misconduct. This shows that academic misconduct among undergraduate students poses a serious risk to the quality of graduates. This paper proposes a machine learning algorithm to quantify academic misconduct. The proposed algorithm is based on a modified k-means clustering algorithm that automatically detects and removes outliers. Universities can adopt the proposed method to combat the growing problem of academic misconduct among undergraduate students. The proposed approach for quantifying the magnitude of academic misconduct is more reliable and cost-effective than traditional (survey-based) methods.

Keywords: Unsupervised machine learning, machine learning, academic dishonesty, clustering, outlier removal

Submitted	: 29.09.2024
Revision Requested	: 06.11.2024
Last Revision Received	: 11.11.2024
Accepted	: 11.11.2024
Published Online	: 10.12.2024

1. INTRODUCTION

Academic dishonesty among students is a growing concern in academia. This is particularly true in higher education. While there are many forms of academic dishonesty, such as plagiarism, cheating—in its many forms—is considered one of the most widespread forms of academic dishonesty (Rettinger & Kramer, 2009; Anitha & Sundaram, 2021). Cheating, in the academic context, can be considered the act of gaining an unfair advantage by unauthorized copying or stealing someones' ideas. This unfair advantage can be gained in a number of ways, including copying answers from someone, taking an examination on behalf of another person, and bringing unauthorized materials into the examination room (Pino & Smith, 2003).

The ramifications of academic dishonesty are far-reaching and affect not only students but also educational institutions. Zhao et al. (2023), for example, found that cheating has a negative effect on learning among middle school children, whereby children who performed well as a result of being allowed to cheat in one test could not perform as well in a similar test when they were denied the opportunity to cheat. Arguably, this finding applies to students in higher learning institutions as well. A similar study by Malik et al. (2023) showed that students tend to perform better in online examinations than in physical ones simply because there is a good opportunity for cheating in online examinations.

Due to its importance in academia, the problem of cheating has received considerable attention. Specifically, many researchers have investigated the factors that motivate students to engage in cheating behavior. For example, in their study on students' perceptions and motivations for cheating, Waltzer and Dahl (2023) found that while most students did not realize their acts as cheating, those who did considered it acceptable, citing task feasibility and assignment goals as excuses.

Peer effects and procrastination from activities such as excessive television time were also found to contribute to cheating behavior (Pino & Smith, 2003; Fontaine, Frenette, & Hébert, 2020). In the former case, if a student embeds themselves in a company that engages in cheating behavior, he or she will become more susceptible to engaging in such behavior. Regarding the latter, if a student spends too much time on leisure activities, such as watching television and spending time on social media, the temptation to engage in cheating behavior to compensate for the wasted time will be high. Moreover, excessive pressure due to the desire to perform better in examinations and deliver homework and assignments on time has also been cited as motivating factors (Grenness, 2023).

Modern technologies are also an important motivating factor for cheating. For example, the proliferation of generative artificial intelligence (AI) tools such as ChatGPT and others has made it even simpler and tempting for students to engage in cheating behavior (Lancaster & Co tarlan, 2021; Lund et al., 2023; Uzun, 2023; Peres, Schreier, Schweidel, & Sorescu, 2023). This proliferation, coupled with easy access to mobile computing devices such as smartphones and tablets, raises new concerns to the problem. With the help of these tools, for example, students can generate essays in a matter of seconds. Similarly, the adoption of online learning by many academic institutions across the globe has contributed to this problem (Jenkins, Golding, Le Grand, Levi, & Pals, 2023; Newton & Essex, 2023). Malik et al. (2023) highlighted that cheating is generally more tempting and easier on online examinations than physical examinations. A comprehensive study by Noorbehbahani et al. (2022) provides a systematic review of the many facets of this problem.

The consequences of cheating are well documented in the literature. For example, cheating behavior practiced during academic life tends to propagate to professional life; that is, students who engage in cheating behavior during their studies tend to exhibit the same behavior in their professional life (Nonis & Swift, 2001; Carpenter, Harding, Finelli, & Passow, 2004). This observation means that the integrity of academic institutions is at stake because of students who are cheating. It is not surprising, therefore, that considerable research effort is being invested in combating this problem.

Over the years, different approaches have been used to combat the problem of academic dishonesty, and cheating in particular. Traditionally, the problem of cheating has been combated through the establishment of a number measures. These measures include a mix of punitive and non-punitive measures. There is no evidence on the suitability of these measures, but mixed opinions among researchers on their effectiveness indicate that neither is universally accepted (Simon et al., 2003; Gallant & Drinan, 2006).

Other approaches used to address this problem include determining the magnitude of the problem (quantifying it) and detecting its presence. The former has been mainly conducted through survey-based studies, whereas the use of modern technology, especially machine learning (ML), has been reported in the latter.

In this paper, we have proposed an ML method to quantify cheating behavior. Like other ML-based methods, the proposed method uses historical data to identify patterns. In this case, the patterns of interest are those related to cheating behavior. The data can be initially structured or unstructured. Specifically, this study addresses the question

"How can machine learning be used to quantify cheating behavior among university students?". The proposed ML method to quantify cheating behavior does not suffer from the bias and sampling errors inherent in surveys.

2. TRADITIONAL METHODS FOR COMBATING CHEATING

To combat cheating effectively, it is important to determine the magnitude of the problem, that is, to estimate the proportion of students engaging in the behavior. This is an important problem that has been studied extensively for many years. The method of choice for many past studies has been surveys. Studies like these have been conducted at many universities across the world where a representative sample of the target student population is interrogated (see, for example, Awdry (2021); DiPaulo (2022a); Newton and Essex (2023)).

One limitation of these methods is the possibility of respondents understating their cheating behavior, which could lead to bias. This phenomenon can be attributed to the lack of awareness and understanding among students about the concept of academic dishonesty and its consequences (Locquiao & Ives, 2020; Chala, 2021). For this reason, most of these studies report the possibility of understating cheating behavior as one of their limitations (Chala, 2021; Orok et al., 2023). Another limitation of the survey methods, which has also been reported in previous studies, is sampling error. This error occurs when a study fails to select a sample that is representative of the population under study, making it impossible to generalize the results.

3. MACHINE LEARNING METHODS FOR COMBATING CHEATING

The past few years have witnessed a notable increase in the use of technology in combating cheating behaviors among students. In particular, the application of machine learning (ML) techniques in this area has received significant attention. This increasing adoption of ML techniques to combat cheating behavior can be attributed to the success of the technology in other domains as well as the abundance of academic data that has been accumulating for many years (Clare, Walker, & Hobson, 2017). These data are available in a variety of formats: unstructured (such as student essays and other works of writing) and structured (such as examination scores). Moreover, deficiencies in traditional assessment practices have also prompted the use of ML techniques in this area (Swiecki et al., 2022).

Consequently, many scholars have approached the cheating problem using ML techniques. Renzella, Cain, and Schneider (2022), for example, used ML for student identity in oral examinations. While oral assessment offers several advantages, such as fostering student interpersonal and communication skills, it is prone to impersonation. This observation is particularly true in large classes where no instructor may know the identity of many students. The student identity validation approach proposed in this study uses ML to determine whether two audio samples resemble speech from the same person. This approach can be used, for example, to facilitate oral discussions between students and tutors while providing the ability to detect contract cheating cases.

Some situations—such as conducting examinations in large classes—call for real-time detection of cheating behavior. However, detecting cheating behavior in real-time is a significant challenge. In such situations, setting, copying answers from fellow students, or unauthorized sources can easily go unnoticed. While combating this problem using traditional methods has proven challenging, it has been demonstrated that post-examination analysis of examination data using ML techniques can be used to detect cheating cases. Meng and Ma (2023), for example, proposed a statistically defensible ML approach for labeling true test cheaters using examination data. The proposed method identifies irregular statistical patterns in examination data. Similarly, many other scholars have employed ML in their studies that deal with detecting cheating behavior (for example, Khabbachi, Zouhair, Mahboub, and El-ghouch (2023); Akiful, Roy, Abdullah, Priota, and Onim (2022); Kaddoura and Gumaei (2022); Bernius, Krusche, and Bruegge (2022); Kamalov, Sulieman, and Santandreu Calonge (2021)).

However, most existing studies focus on detecting cheating as part of a wider effort to address this problem. While detecting cheating behavior is important in addressing the problem, understanding the *magnitude* of the problem is a necessary step when planning for appropriate mitigation measures, such as resource allocation. For example, while the use of proctoring systems has shown some success, the decision to adopt such systems should be taken only when the magnitude of the problem is known in advance and justifies their adoption, as these systems tend to be costly (Nigam, Pasricha, Singh, & Churi, 2021).

For example, at Mzumbe University, where this study was conducted, final-year students in information technology (IT) programs are required to complete a substantial software development project in their final year of studies. Over the years, it has been learned that some students engage in cheating behavior while undertaking these projects. Cheating mainly occurs in the form of repeated work performed by another person in previous years (either as is or with minor modifications). Despite this understanding, it has been challenging to determine the magnitude of the problem due to

the shear number of projects completed each year and the fact that these projects are supervised by different supervisors. Given this challenge, choosing an appropriate control mechanism to address the problem was challenging.

4. MATERIALS AND METHODS

The aim of this study was to estimate the extent of cheating (hereafter referred to as the cheating index or C-index) among final-year university students using unsupervised ML. Specifically, the study employed cluster analysis with outlier detection and removal. The underlying assumption is that similar titles (those repeated as is or with modifications) tend to cluster together, while new titles tend to segregate themselves from the rest as outliers. With this assumption in mind, the extent of cheating (as a proportion of the original number of titles) can be obtained by detecting and removing outliers from each cluster and then finding a percentage.

In this section, we describe the test data used in the study as well as details on how the data were processed and analyzed to answer the main research question posed above.

4.1. Test data

This study was conducted at Mzumbe University in Tanzania. The data used in this study pertain to final-year student project titles. These project titles are typically short texts with an average length of 35 characters. In total, 678 titles spanning a period of 5 years (2018–2022) were used. This number represents an average of 135 titles per year.

The titles were collected from final-year students' project reports. The students were enrolled in three ICT-based bachelor's programs: Information Technology and Systems (ITS), Information and Communication Technology with Management (ICT-M), and Information and Communication Technology with Business (ICT-B).

Students enrolled in these programs are required to complete a software development project during their final year of studies as a partial fulfillment of their program. This project is usually a web or mobile application completed in one semester. Examples of the project titles used in this study are as follows:

Tanzania online business license application system
Computerisation of transfer for public school teachers
Accident detection and notification mobile application

After extraction from the students' final year project reports, the titles were stored in a UTF-8 plain text file (one title per line) for subsequent processing and analysis.

4.2. Data Analysis and C-index Computation

To analyze the test data and compute the C-index, we employed a workflow that consisted of a sequence of steps: data preprocessing, feature extraction, determination of a suitable number of clusters to use in a clustering algorithm, cluster analysis with outlier removal, and finally computation of the C-index (see Fig. 1). Each step produces an output that serves as input to the next step. The workflow was implemented in Python.

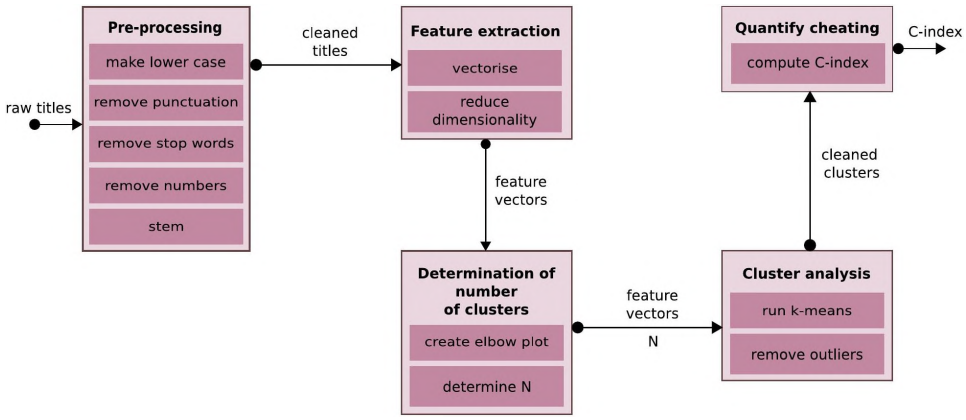


Figure 1. Proposed workflow for estimating C-index. The arrows indicate the sequencing of steps, and N represents the appropriate number of clusters for use in cluster analysis.

4.2.1. Data Pre-processing

The aim of the data pre-processing step was to prepare the data for the subsequent steps. First, each title in the data file was converted to lowercase letters to make the titles uniform and simplify other preprocessing steps. Second, all punctuation symbols were removed because they carry little or no useful information. Third, common terms, such as *system* and app, were removed for the same reason. The fourth and fifth steps removed numbers (digits) and normalized the words in each title by *stemming*, respectively. In addition, the titles were checked for misspellings and typos using a spellchecker. This step generated cleaned titles (documents) for the next step.

4.2.2. Feature Extraction

In this step, the titles are transformed into numerical feature vectors using the term frequency inverse document frequency (TF-IDF) algorithm. This text representation algorithm was selected because it is simple, considers the relative importance of individual terms, and takes care of stop words automatically. A drawback of the proposed model is that it typically produces extremely sparse feature vectors. The frequency of term t in document d is expressed as follows:

$$tf(t,d) = log(f_{t,d})$$

where $f_{t.d}$ is the frequency of t in d. The inverse document frequency of term t in a set D of N documents is given by

$$idf(t,D) = log\left(\frac{N}{n_t}\right)$$

where n_t is the number of documents in which t appears in. TF-IDF is computed as follows:

$$tf\text{-}idf(t,d,D) = tf(t,d) \cdot idf(t,D).$$

Both the term frequency and the inverse document frequency are usually scaled logarithmically to prevent bias in longer documents and in terms that appear more frequently relative to others (Aggarwal, 2022).

After the documents were vectorized, the dimensions of the resulting feature vectors were reduced by projecting them to a lower-dimensional feature space namely $\mathbb{R}^3$. This projection was performed to simplify handling and facilitate data visualization. The resulting TF-IDF feature vectors were extremely sparse; thus, this projection was performed using the singular value decomposition (SVD) technique, which, unlike regular principal component analysis (PCA), can handle sparse matrices. This step produces feature vectors that are used to determine the suitable number of clusters to use in the cluster analysis step.

4.2.3. Determining the Number of Clusters

The main input to the k-means clustering algorithm is a predetermined optimal number of clusters to be used. In practice, this number can be determined using an elbow plot. Therefore, the number of clusters to be used was determined using the elbow plot shown in Fig. 2. Although the elbow appeared at four clusters in the figure, five clusters provided more plausible results. Thus, the number of clusters was five ($N = 5$).

4.2.4. Cluster Analysis with Outlier Removal

In this step, two actions were performed. First, the feature vectors were clustered using the k-means algorithm. Second, each cluster was subjected to an outlier removal process. The k-means algorithm is widely used in machine learning and pattern recognition tasks to partition (cluster) data due to its simplicity and ease of implementation. Mathematically, the algorithm aims to partition given data $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_n\}$ into N groups (clusters) $C_1, C_2, \ldots, C_N$ such that the sum of the within-cluster squared distance between each point and the centroid of the cluster is minimized, that is

$$\underset{C}{arg\,min} \sum_{n=1}^{N} \sum_{\mathbf{x} \in C_n} ||\mathbf{x} - \mathbf{c}_n||^2$$

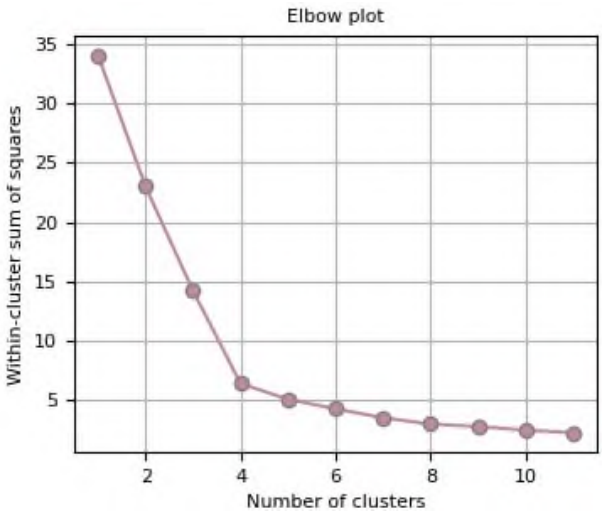


Figure 2. Elbow plot showing the optimal number of clusters for the k-means algorithm.

where $\mathbf{c}_n$ is the centroid of the n^{th} cluster. We ran k-means using the value of N determined above. The clusters obtained were expected to contain some outliers; thus, each cluster was subjected to an outlier removal process. The steps involved in the outlier removal process are summarized in Fig. 3.

The algorithm examines each cluster independently. For each point in a cluster, its k nearest neighbors are determined. Next, the distance between the point and its furthest neighbor (d_{max}) and the median interpoint distance (d_{max}) between the point and all its neighbors are computed (see Fig. 4). If the magnitude of the difference between d_{max} and d_{max} (σ) exceeds a predetermined threshold θ, the point is designated an outlier and marked for removal by adding it to a set of outlier points. Finally, all points in the outlier set are removed from the cluster.

The points are not removed immediately because doing this would distort the structure of the cluster under consideration and would confuse the algorithm. The values of the parameters k and t were determined experimentally; we found that the values k = 10 and θ = 0.018 served their purpose reasonably well.

input : Feature vectors $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset \mathbb{R}^3$
output: N clusters $\{P_1, \ldots, P_N\}$ with codebook $C = \{\mathbf{c}_1, \ldots, \mathbf{c}_N\}$ and outliers removed, where $P_i = \{\mathbf{x} : ||\mathbf{x} - \mathbf{c}_i|| \leq ||\mathbf{x} - \mathbf{c}_j|| \; \forall j \neq i\}$

Initialise: N, θ, k;
$Outliers \leftarrow \emptyset$;
$\{P_1, \ldots, P_N\} \leftarrow$ k-means(X, N);
for $P \in \{P_1, \ldots, P_N\}$ **do**
  for $\mathbf{x} \in P$ **do**
    $kNN(\mathbf{x}, k) \leftarrow S_x \subseteq X \; s.t. \; |S_x| = k$ and $\forall \mathbf{x}' \in X \setminus S_x$, $||\mathbf{x} - \mathbf{x}'|| \geq \max_{\mathbf{x}'' \in S_x} ||\mathbf{x} - \mathbf{x}''||$;
    $d_{max} \leftarrow max\{||\mathbf{x} - \mathbf{q}_i||\}, i = 1, \ldots, k$ and $\mathbf{q}_i \in S_x$;
    $d_{med} \leftarrow median\{||\mathbf{q}_i - \mathbf{q}_j||\}, i = j = 1, \ldots, k, i \neq j$ *and* $\mathbf{q}_i, \mathbf{q}_j \in S_x$;
    $\sigma = |d_{max} - d_{med}|$;
    if $\sigma > \theta$ **then**
      $Outliers \leftarrow Outliers \cup \{\mathbf{x}\}$;
    end
  end
  $P \leftarrow P \setminus Outliers$;
end

Figure 3. Proposed k-means clustering with outlier removal.

4.2.5. Computing C-index

After removing the outliers from each cluster, the cheating index is computed as the proportion of points remaining after removing the outliers:

$$C\text{-}index = \frac{\text{Number of remaining feature vectors}}{\text{Orginal number of feature vectors}} \times 100\%$$

5. RESULTS

The aim of this study was to use ML to estimate the extent of cheating among final-year university students. In this section we present the results of the analysis, both before and after the removal of the outliers. To realize this, we use visualization and various common clustering performance metrics.

The clustering results of the project titles prior to the removal of outliers are shown in Fig. 5. Similarly, Fig. 6 shows the clustering results of the titles after removing the outliers.

The silhouette score (a measure of cohesion among members of a cluster) of the clusters before the removal of outliers was 0.0082. After removing the outliers, the silhouette score improved to 0.7475, which reflects cluster homogeneity. Similarly, the within-cluster sum of squared errors (WCSSE) decreased significantly after the removal of outliers (Table 1). Using these results, the computed value of the cheating index was 81.3

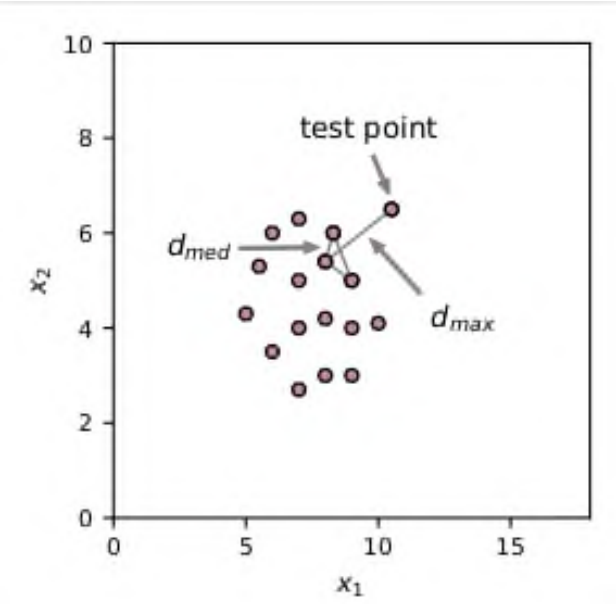


Figure 4. Detecting an outlier using distances between a test point and its nearest neighbors. Here, three neighboring points (k = 3) are used.

Table 1. *Within-cluster sum of squared errors for five clusters before and after outlier removal*

	C1	C2	C3	C4	C5
Before	0.7459	0.6669	1.3891	1.9139	0.3428
After	0.5838	0.2019	0.8301	0.0310	0.0000

6. DISCUSSION

Fig. 5 shows three clusters which are highly cohesive (C1, C2, and C3) in which most feature vectors are contained. In contrast, the remaining clusters (C4 and C5) were highly incoherent. The dispersed vectors in C4 and C5 indicate that project titles corresponding to those vectors have a higher degree of uniqueness than those in C1–C3. Points in these clusters represent titles that differ from the rest, which translates to the absence of cheating behavior.

When the points in C4 and C5 were removed (Fig. 6), the remaining clusters consisted of vectors corresponding to repeated titles (presence of high degree of cheating behavior). The dramatic drop in WCSSE (Table I) and the increase in the silhouette score from 0.0082 to 0.7475 indicate tight coherence among the clusters, which translates to a high degree of resemblance between the project titles represented by the vectors in those clusters. Here, the number of remaining clusters (3) indicates that over the five years that the test data span, students have been repeating project titles from three pools of titles represented by points in those clusters. The high silhouette score and small WCSSE values indicate that these titles were either reused as is or with only minor alterations (such as replacing words with synonyms or rearranging words in the title).

The high value for the cheating index (81.3%) was due to the fact that only a few vectors were identified as outliers and thus were removed (c.f. Fig. 5 (a)–(b) and Fig. 6 (a)–(b)). Because there is a one-to-one correspondence between the titles and students who worked on those titles,

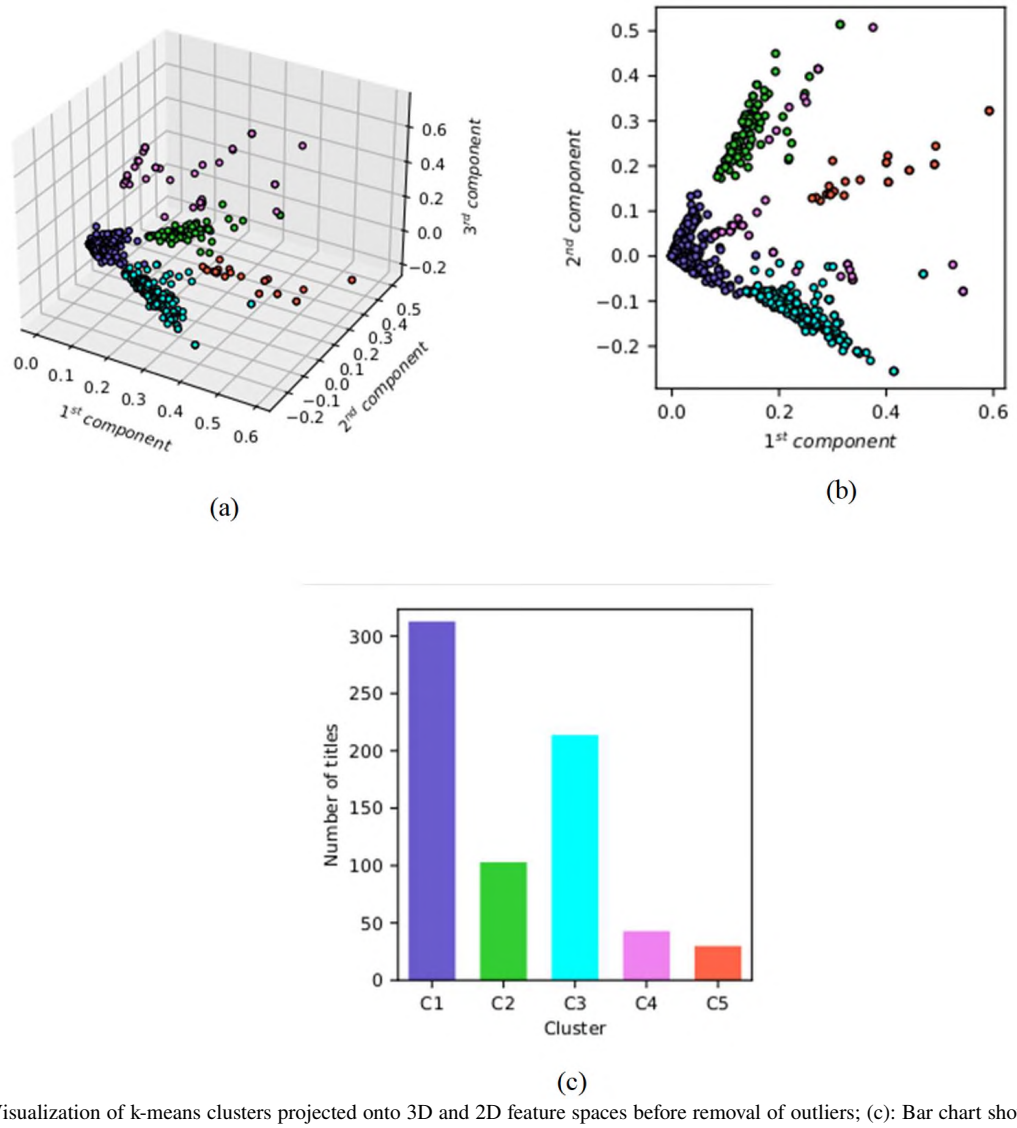


Figure 5. (a)–(b): Visualization of k-means clusters projected onto 3D and 2D feature spaces before removal of outliers; (c): Bar chart showing the color-coded number of points in each cluster.

this value can be regarded as the percentage of students who engaged in cheating behavior during the study period (2018–2022). This finding on the magnitude of cheating behavior is highly consistent with other studies that used traditional methods to determine the prevalence of academic misconduct among university students (Salehi & Gholampour, 2021; Wang & Xu, 2021; Awdry, 2021; DiPaulo, 2022b). Although both methods yield similar results, methods based on ML offer greater flexibility and are less laborious.

Although the present study does not address the reasons for this high rate of cheating, other studies may explain this observation (Anitha & Sundaram, 2021; Waltzer & Dahl, 2023). In addition, the temptation to reuse source code from previous years and not engage in the taxing endeavor of writing code from scratch can also be attributed to this finding.

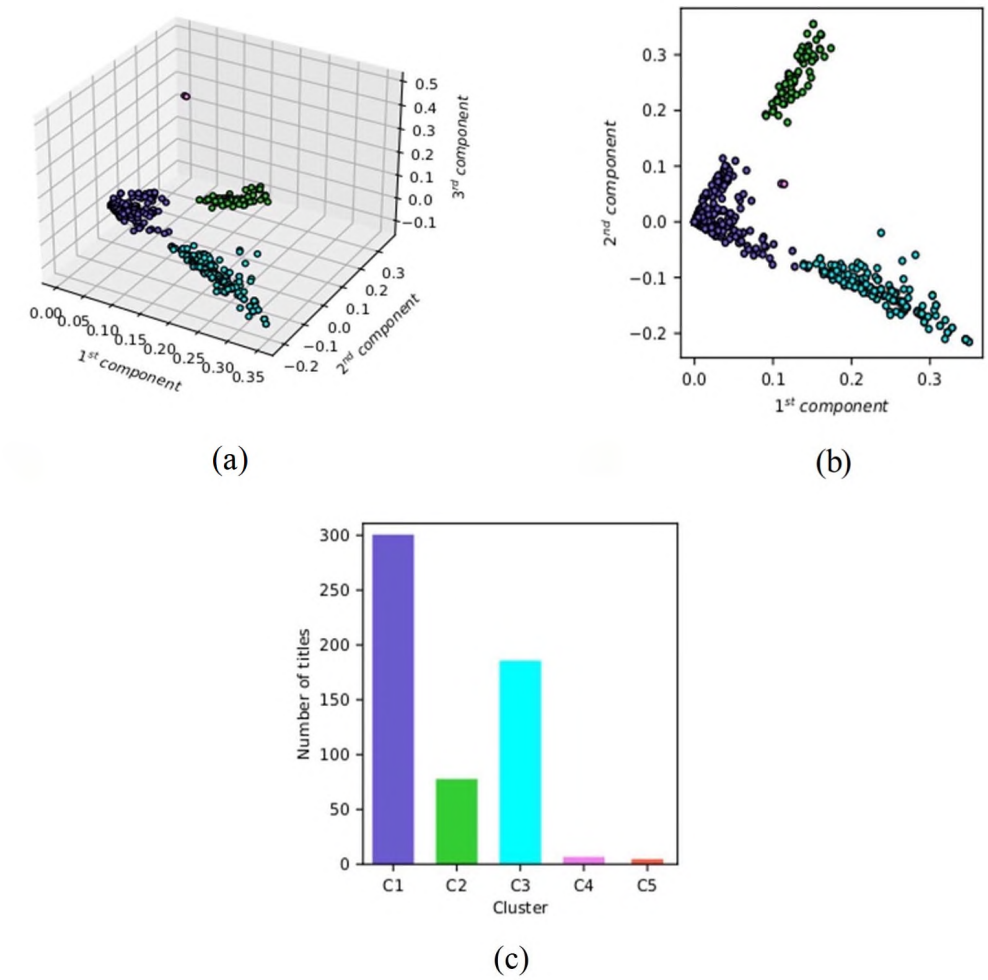


Figure 6. (a)–(b): Visualization of k-means clusters projected onto 3D and 2D feature spaces after removal of outliers; (c): Bar chart showing the color-coded number of vectors in each cluster.

The present study focused on quantifying the extent of cheating as opposed to most existing studies (which focus of detection and prevention of cheating). The results of this study can be used as a foundation for future studies. For example, some of these methods (e.g., Ranger, Schmidt, and Wolgast (2022); Chang and Chang (2023)) rely on supervised learning—particularly classification—which requires labeled data to work. The proposed method can be used to automatically create reliable training datasets for training classification models.

7. CONCLUSION

Understanding the extent of academic dishonesty is a key step in combating this problem; however, the literature is sparse on the use of machine learning (ML) techniques to quantify the magnitude of the problem. In this paper, an ML approach to quantify the magnitude of cheating behavior among final-year university students is proposed. The proposed method is based on cluster analysis coupled with outlier detection and removal. Results show that over 80% of the students engaged in cheating behavior during the period that the test data spans. This finding is consistent with existing studies that use traditional methods (surveys) to determine the extent of cheating behavior among university students. Because large amounts of academic data are readily available, the proposed method offers more flexibility and is more cost-effective than traditional methods. In addition, the proposed method can be used to create reliable datasets for studies that use ML to detect and prevent academic dishonesty.

Peer Review: Externally peer-reviewed.

Conflict of Interest: The author have no conflict of interest to declare.

Grant Support: The author declared that this study has received no financial support.

ORCID ID of the author

Almasi S. Maguya 0000-0002-1345-121X

REFERENCES

Aggarwal, C. (2022). *Machine learning for text.* Springer International Publishing.

Akiful, H. A., Roy, K., Abdullah, N., Priota, N. Z., & Onim, S. H. (2022). Performance Analysis of Machine Learning Models for Cheating Detection in Online Examinations. In 2022 *25th international conference on computer and information technology* (ICCIT) (pp. 342–347). doi: 10.1109/ICCIT57492.2022.10055801

Anitha, P., & Sundaram, S. (2021). Prevalence, types and reasons for academic dishonesty among college students. *Journal of Studies in Social Sciences and Humanities, 7*(1), 1–14.

Awdry, R. (2021). Assignment outsourcing: Moving beyond contract cheating. *Assessment & Evaluation in Higher Education, 46* (2), 220–235. doi: 10.1080/02602938.2020.1765311

Bernius, J. P., Krusche, S., & Bruegge, B. (2022). Machine learning based feedback on textual student answers in large courses. *Computers and Education: Artificial Intelligence, 3*, 100081. doi: https://doi.org/10.1016/j.caeai.2022.100081

Carpenter, D. D., Harding, T. S., Finelli, C. J., & Passow, H. J. (2004). Does academic dishonesty relate to unethical behavior in professional practice? An exploratory study. *Science and engineering ethics, 10*, 311–324.

Chala, W. D. (2021). Perceived seriousness of academic cheating behaviors among undergraduate students: an Ethiopian experience. *International Journal for Educational Integrity, 17*(1), 2.

Chang, S.-C., & Chang, K. L. (2023). Cheating Detection of Test Collusion: A Study on Machine Learning Techniques and Feature Representation. *Educational Measurement: Issues and Practice, 42* (2), 62–73. doi: https://doi.org/10.1111/emip.12538

Clare, J., Walker, S., & Hobson, J. (2017). Can we detect contract cheating using existing assessment data? Applying crime prevention theory to an academic integrity issue. *International Journal for Educational Integrity, 13*(1), 1–15.

DiPaulo, D. (2022a). Do preservice teachers cheat in college, too? A quantitative study of academic integrity among preservice teachers. *International Journal for Educational Integrity, 18*(1), 2.

DiPaulo, D. (2022b). Do preservice teachers cheat in college, too? A quantitative study of academic integrity among preservice teachers. *International Journal for Educational Integrity, 18*(1), 2.

Fontaine, S., Frenette, E., & Hébert, M.-H. (2020). Exam cheating among Quebec's preservice teachers: the influencing factors. *International Journal for Educational Integrity, 16*(1), 1–18.

Gallant, T. B., & Drinan, P. (2006). Organizational theory and student cheating: Explanation, responses, and strategies. *The Journal of Higher Education, 77* (5), 839–860.

Grenness, T. (2023). "If You Don't Cheat, You Lose": An Explorative Study of Business Students' Perceptions of Cheating Behavior. *Scandinavian Journal of Educational Research, 67*(7), 1122–1136. doi: 10.1080/00313831.2022.2116479

Jenkins, B. D., Golding, J. M., Le Grand, A. M., Levi, M. M., & Pals, A. M. (2023). When opportunity knocks: College students' cheating amid the COVID-19 pandemic. *Teaching of Psychology, 50*(4), 407–419.

Kaddoura, S., & Gumaei, A. (2022). Towards effective and efficient online exam systems using deep learning-based cheating detection approach. *Intelligent Systems with Applications, 16* , 200153. doi: https://doi.org/10.1016/j.iswa.2022.200153

Kamalov, F., Sulieman, H., & Santandreu Calonge, D. (2021). Machine learning based approach to exam cheating detection. *Plos one, 16*(8), e0254340.

Khabbachi, I., Zouhair, A., Mahboub, A., & Elghouch, N. (2023). Reduce Cheating in e-Exams Using Machine Learning: State of the Art. In M. Lazaar, E. M. En-Naimi, A. Zouhair, M. Al Achhab, & O. Mahboub (Eds.), *Proceedings of the 6th international conference on big data and internet of things* (pp. 225–238). Springer International Publishing.

Lancaster, T., & Cotarlan, C. (2021). Contract cheating among STEM students through file sharing websites: A COVID-19 pandemic perspective. *International Journal for Educational Integrity, 17*(1), 1–16.

Locquiao, J., & Ives, B. (2020). First-year university students' knowledge of academic misconduct and the association between goals for attending university and receptiveness to intervention. *International Journal for Educational Integrity, 16*(1), 5.

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology, 74*(5), 570–581.

Malik, A. A., Hassan, M., Rizwan, M., Mushtaque, I., Lak, T. A., & Hussain, M. (2023). Impact of academic cheating and perceived online learning effectiveness on academic performance during the COVID-19 pandemic among Pakistani students. *Frontiers in Psychology, 14*(2), 1124095.

Meng, H., & Ma, Y. (2023). Machine Learning-Based Profiling in Test Cheating Detection. *Educational Measurement: Issues and Practice, 42*(1), 59–75. doi: https://doi.org/10.1111/emip.12541

Newton, P. M., & Essex, K. (2023). How common is cheating in online exams and did it increase during the COVID-19 pandemic? A Systematic Review. *Journal of Academic Ethics*, 1–21.

Nigam, A., Pasricha, R., Singh, T., & Churi, P. (2021). A systematic review on AI-based proctoring systems: Past, present and future. *Education and Information Technologies, 26* (5), 6421–6445.

Nonis, S., & Swift, C. O. (2001). An examination of the relationship between academic dishonesty and workplace dishonesty: A multicampus investigation. *Journal of Education for business, 77*(2), 69–77.

Noorbehbahani, F., Mohammadi, A., & Aminazadeh, M. (2022). A systematic review of research on cheating in online exams from 2010 to 2021. *Education and Information Technologies, 27*(6), 8413–8460.

Orok, E., Adeniyi, F., Williams, T., Dosunmu, O., Ikpe, F., Orakwe, C., & Kukoyi, O. (2023). Causes and mitigation of academic dishonesty among healthcare students in a Nigerian university. *International Journal for Educational Integrity, 19*(1), 13.

Peres, R., Schreier, M., Schweidel, D., & Sorescu, A. (2023). On ChatGPT and beyond: How generative artificial intelligence may affect research, teaching, and practice. *International Journal of Research in Marketing*. doi: https://doi.org/10.1016/j.ijresmar.2023.03.001

Pino, N. W., & Smith, W. L. (2003). College students and academic dishonesty. *College Student Journal, 37*(4), 490–500.

Ranger, J., Schmidt, N., & Wolgast, A. (2022). Detecting Cheating in Large-Scale Assessment: The Transfer of Detectors to New Tests. *Educational and Psychological Measurement, 0*(0), 00131644221132723. doi: 10.1177/00131644221132723

Renzella, J., Cain, A., & Schneider, J.-G. (2022). Verifying student identity in oral assessments with deep speaker. *Computers and Education: Artificial Intelligence, 3* , 100044. doi:10.1016/j.caeai.2021.100044

Rettinger, D., & Kramer, Y. (2009). Situational and Personal Causes of Student Cheating. *Research in Higher Education, 50* , 293–313. doi: https://doi.org/10.1007/s11162-008 -9116-5

Salehi, M., & Gholampour, S. (2021). Cheating on exams: Investigating Reasons, Attitudes, and the Role of Demographic Variables. *SAGE Open, 11*(2), 21582440211004156. doi: 10.1177/21582440211004156

Simon, C. A., Carr, J. R., McCullough, S. M., Morgan, S. J., Oleson, T., & Ressel, M. (2003). The other side of academic dishonesty: The relationship between faculty scepticism, gender and strategies for managing student academic dishonesty cases. *Assessment & Evaluation in Higher Education, 28*(2), 193–207.

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan S. . . Ga˘sevi´c, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence, 3* , 100075. doi:10.1016/j.caeai.2022.100075

Uzun, L. (2023). ChatGPT and academic integrity concerns: Detecting artificial intelligence generated content. *Language Education and Technology, 3*(1), 45–54.

Waltzer, T., & Dahl, A. (2023). Why do students cheat? Perceptions, evaluations, and motivations. *Ethics & Behavior , 33* (2), 130–150. doi: 10.1080/10508422.2022.2026775

Wang, Y., & Xu, Z. (2021). Statistical Analysis for Contract Cheating in Chinese Universities. *Mathematics, 9* (14). doi: 10.3390/math9141684

Zhao, L., Peng, J., Dong, L. D., Compton, B. J., Zhong, Z., Li Y. . . Lee, K. (2023). Academic cheating interferes with learning among middle school students. *Journal of Experimental Child Psychology, 226*(2), 10556.

How cite this article

Maguya, A. S. (2024). A Machine Learning Approach for Quantifying Academic Misconduct. *Acta Infologica, 8*(2), 188-198. https://doi.org/10.26650/acin.1557985

İSTANBUL UNIVERSITY PRESS

Acta Infologica
acin 2024, 8 (2), 199–212
DOI: 10.26650/acin.1558583

Research Article

A New Kibria-Lukman-Type Estimator for Poisson Regression Models

Cemal Çiçek[1], Kadri Ulaş Akay[1]

[1]İstanbul University, Faculty of Science, Department of Mathematics, İstanbul, Türkiye

Corresponding author : Cemal Çiçek
E-mail : cicekc@istanbul.edu.tr

ABSTRACT

One of the most important models for the analysis of count data is the Poisson Regression Model (PRM). The parameter estimates of the PRM are obtained by the Maximum Likelihood Estimator (MLE). However, MLE is adversely affected in the presence of multicollinearity, which is known as the approximately linear relationship between the explanatory variables. Many shrinkage estimators have been proposed to reduce the effects of multicollinearity in PRMs. As an alternative to other biased estimators that are already in use in PRMs, we presented a novel estimator in this paper that is based on the Kibria-Lukman estimator. The superiority of the proposed new biased estimator over existing biased estimators is given by the asymptotic matrix mean square error. Furthermore, two separate Monte Carlo simulation studies are conducted to investigate the performance of the proposed biased estimators. Finally, real data is used to examine the superiority of the proposed estimator.

Keywords: Mean squared error, multicollinearity, poisson liu estimator, poisson regression, poisson ridge estimator

Submitted : 30.09.2024
Revision Requested : 07.10.2024
Last Revision Received : 06.12.2024
Accepted : 09.12.2024
Published Online : 24.12.2024

1. INTRODUCTION

The Poisson Regression Model (PRM) is a basic model used to analyze count data (Hilbe, 2014). Let y_i be the response variable and follow a Poisson distribution with mean μ_i, the probability mass function is defined as

$$f(y_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, i = 1, 2, .., n \quad y_i = 0, 1, 2, ... \tag{1.1}$$

where the canonical log-link function and a linear combination of explanatory variables are used to describe μ_i as follows: $\mu_i = \exp(x_i'\beta)$ where x_i' is the ith row of model matrix X, which is an $n \times (p+1)$ matrix with p explanatory variables and β is a $(p+1)\times$ n vector of parameters.

The most popular estimation approach for estimating parameters in PRMs is the Maximum Likelihood method. The following is the log-likelihood function for PRM:

$$l(\beta) = \sum_{i=1}^{n} y_i x_i'\beta - \exp(x_i'\beta) - \log(y_i!). \tag{1.2}$$

The log-likelihood function is maximized to obtain the Maximum Likelihood Estimator (MLE) of β , which yields the following equations:

$$S(\beta) = \frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^{n} \left[y_i - \exp\left(x_i'\beta\right)\right] x_i = X'(y-\mu) = 0 \tag{1.3}$$

where μ is an $n \times 1$ dimensional vector with elements are $\mu_i = \exp\left(x_i'\beta\right), i = 1, 2, ..., n$. Since Equation (1.3) is nonlinear in β , the following iteratively reweighted least squares (IRLS) algorithm is used to find the solution of $S(\beta)$:

$$\hat{\beta}_{MLE} = (X'\hat{W}X)^{-1}X'\hat{W}Z \tag{1.4}$$

where Z is a vector with the ith element $z_i = \log(\hat{\mu}_i) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}$ and $\hat{W} = diag[\hat{\mu}_i]$, $i = 1, 2, ..., n$. The iterations end when the difference between successive estimates converges or is less than a specified small value, which is usulally 10^{-8} (Dunn & Smyth, 2018). The logic of the IRLS algorithm for PRM with a canonical link function is summarized in Table 1 (Hardin & Hilbe, 2018).

Table 1. IRLS estimation algorithm for PRM with canonical link function.

$Dev = 0$
$\mu_i = \left(y_i + \text{mean}(y)\right)/2$, #where y is the response vector whose components are y_i, $i = 1, 2, \ldots, n$. Initialization of μ
$\eta = \ln(\mu)$ #initialization of η
While abs(ΔDev > tolerans) {
$W = \text{diag}(\mu_i)$ $i = 1, 2, \ldots, n$ #where W is the weighted matrix.
$z_i = \log(\mu_i) + (y_i - \mu_i)/\mu_i$ $i = 1, 2, \ldots, n$
$\beta = (X'WX)^{-1} X'WZ$ #where Z is a vector with the ith element $z_i, i = 1, 2, ..., n$.
$\eta = X'\beta$
$\mu = \exp(\eta)$
old$Dev = Dev$
$Dev = 2\Sigma\{y \ln(y/\mu) - (y-\mu)\}$
$\Delta Dev = Dev -$ oldDev } #Iterate until the change in deviance, log-likelihood, or estimated parameter values between two iterations is below a specified level of tolerance, or threshold.

An important drawback of the MLE is that estimated parameter values become unstable when multicollinearity occurs (Kibria et al. 2013; Türkan & Özel, 2016; Rashad & Algamal, 2019; Amin et al. 2022; Jadhav, 2022; Alkhateeb &

Algamal, 2020; Månsson & Kibria, 2020; Lukman et al., 2021; Akay & Ertan, 2022; Ertan & Akay, 2023). The estimates of model parameters in PRMs, as in linear regression models, are affected by the multicollinearity problem, which results from the approximately linear relationship between explanatory variables. The variance of MLE increases to such a degree that the estimates of model parameters become unstable due to multicollinearity between the explanatory variables Månsson & Shukur (2011), Månsson et al. (2012), Kibria et al. (2015), Asar & Genç (2018), Çetinkaya & Kaçıranlar (2019), Qasim et al. (2020b), Alheety et al. (2021).

Instead of using the MLE, alternative biased estimators are recommended to alleviate the negative impacts of multicollinearity. For instance, Månsson & Shukur (2011) defined the Poisson Ridge Estimator (PRE) as follows:

$$\hat{\beta}_{PRE} = \left(X'\hat{W}X + kI\right)^{-1} X'\hat{W}X\hat{\beta}_{MLE}, k > 0, \tag{1.5}$$

where k is a biasing parameter. The Ridge estimator (RE) proposed by Hoerl & Kennard (1970) for the linear regression model is generalized by the PRE.

The Poisson Liu Estimator (PLE) is proposed by Månsson et al. (2012), Amin et al. (2021), and Qasim et al. (2020a) as

$$\hat{\beta}_{PLE} = (X^{'}\hat{W}X + I)^{-1}(X^{'}\hat{W}X + dI)\hat{\beta}_{MLE}, 0 < d < 1, \tag{1.6}$$

where d is a biasing parameter. The Liu estimator (LE), which Liu (1993) proposed for the linear regression model, is extended by PLE.

As an alternative to PRE and PLE, two biased estimators have been merged in recent years to produce innovative estimators with two biasing parameters. The Poisson–Liu-type estimator (PLTE) for PRM is defined in this context by Algamal (2018) as follows:

$$\hat{\beta}_{PLTE} = \left(X'\hat{W}X + kI\right)^{-1} \left(X'\hat{W}X - dI\right) \hat{\beta}_{MLE}, k > 0, d \in R \tag{1.7}$$

where k and d are the biasing parameters, respectively. The PLTE is a generalization of the Liu-type estimator introduced by Liu (2003). Moreover, Asar & Genç (2018) and Çetinkaya & Kaçıranlar (2019) proposed another estimator with two biasing parameters, which was defined by Özkale & Kaçıranlar (2007) for linear regression models. The Poisson Two-Parameter Estimator (PTPE) is defined as follows:

$$\hat{\beta}_{PTPE} = \left(X'\hat{W}X + kI\right)^{-1} \left(X'\hat{W}X + kdI\right) \hat{\beta}_{MLE}, \ k > 0, 0 < d < 1, \tag{1.8}$$

where k and d are the biasing parameters, respectively.

However, as an alternative to estimators with two biasing parameters, Akay & Ertan (2022) proposed the improved Liu-type Estimator (ILTE). The following definitions of ILTE include MLE, PRE, PLE, PLTE, and PTPE:

$$\hat{\beta}_{ILTE} = (X^{'}\hat{W}X + kI)^{-1}(X^{'}\hat{W}X + f(k)I)\hat{\beta}^{*}, \ k > 0, \tag{1.9}$$

where $\hat{\beta}^{*}$ is an estimator of β and $f(k)$ is a continuous function of k. The ILTE is a generalization of the Liu-type estimator defined by Kurnaz & Akay (2015) for linear regression models.

Let $f(k) = -k$ and $\hat{\beta}^{*} = \hat{\beta}_{MLE}$ as a special case of the estimator given by (1.9). In the literature, this estimator is known as Kibria-Lukman type estimator. Aladeitan et al. (2021) defined the Kibria-Lukman-type estimator for PRM as follows:

$$\hat{\beta}_{PKLE} = \left(X'\hat{W}X + kI\right)^{-1} \left(X'\hat{W}X - kI\right) \hat{\beta}_{MLE}, k > 0, \tag{1.10}$$

where k is a biasing parameter.

Numerous biased estimators for linear regression models have been adapted for use with PRMs in the literature. In recent investigations, researchers have concentrated on the Kibria-Lukman type estimator (Aladeitan et al., 2021; Dawoud et al. 2022; Lukman et al., 2023; Akay et al., 2023; Alrweili, 2024). Therefore, in addition to the estimators given above, in this paper, we focus on the application to PRMs of a new estimator based on the PKLE estimator given by (1.10). Additionally, as an alternative to the PRE and PLE, our goal in this study is to examine the performance of this new estimator with a single biasing parameter.

The remainder of this paper is organized as follows. A new biased estimator is defined, and some of its characteristics are described in Section 2. The conditions under which the proposed new estimator outperforms ILTE in terms of the matrix mean squared error are illustrated in Section 3. In Section 4, several estimators are proposed to determine the biasing parameter. In Section 5, two separate Monte Carlo simulation studies are conducted to evaluate the performance of the proposed estimator compared to other estimators. A real-world data application is presented in Section 6 to demonstrate how well the suggested biased estimators function. Finally, the conclusions of the study are given in Section 7.

2. A NEW KIBRIA-LUKMAN-TYPE ESTIMATOR

For PRMs, we can generalize the Kibria-Lukman estimator given in (1.10) as follows:

$$\hat{\beta} = (X^{'}WX + kI)^{-1}(X^{'}WX - kI)\hat{\beta}^{*} \tag{2.1}$$

where k is the biasing parameter and $\hat{\beta}^{*}$ is any estimator of β. As an approach to the case of nested estimators, we consider the estimator obtained when $\hat{\beta}^{*} = \hat{\beta}_{PRE}$ as follows:

$$\hat{\beta}_{PKLTEI} = (X^{'}WX + kI)^{-1}(X^{'}WX - kI)(X^{'}WX + kI)^{-1}X^{'}WX\hat{\beta}_{MLE} \tag{2.2}$$

where k is a biasing parameter. If $\hat{\beta}^{*} = \hat{\beta}_{PKLE}$, the estimator obtained is as follows:

$$\hat{\beta}_{PKLTEII} = (X^{'}WX + kI)^{-1}(X^{'}WX - kI)(X^{'}WX + kI)^{-1}(X^{'}WX - kI)\hat{\beta}_{MLE} \tag{2.3}$$

where k is a biasing parameter.

Using the estimator provided in (2.2) and (2.3), we can now determine the asymptotic scalar mean squared error (SMSE) and the asymptotic matrix mean squared error (MMSE). We indicate $\alpha = Q'\beta$, $\Lambda = diag\left(\lambda_1, ..., \lambda_{p+1}\right) = Q'\left(X'\hat{W}X\right)Q$, where $\lambda_1 \geq \lambda_2 \geq ...\lambda_{p+1} > 0$ are the ordered eigenvalues of $X'\hat{W}X$, the eigenvectors of $X'\hat{W}X$ are represented by the columns of Q and the ith element of $Q'\beta$ is denoted as α_j, $j = 1, 2, ..., p+1$.

The asymptotic SMSE and the asymptotic MMSE of $\hat{\beta} = A\hat{\beta}_{MLE}$ are defined as follows:

$$MSEM(\hat{\beta}) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)^{'} = A(\hat{\beta}_{MLE} - \beta)(\hat{\beta}_{MLE} - \beta)^{'}A^{'} + (A\beta - \beta)(A\beta - \beta)^{'} \tag{2.4}$$

$$SMSE(\hat{\beta}) = E(\hat{\beta} - \beta)^{'}(\hat{\beta} - \beta) = (\hat{\beta}_{MLE} - \beta)^{'}A^{'}A(\hat{\beta}_{MLE} - \beta) + (A\beta - \beta)^{'}(A\beta - \beta).$$

The relationship between the MMSE and SMSE is $SMSE(\hat{\beta}) = tr(MMSE(\hat{\beta}))$. Because of the relation of $\alpha = Q^{'}\beta$; $\hat{\beta}_{MLE}, \hat{\beta}_{PRE}, \hat{\beta}_{PLE}, \hat{\beta}_{PLTE}, \hat{\beta}_{ILTE}$ and $\hat{\beta}_{PKLTE}$ possess identical SMSE values to $\hat{\alpha}_{MLE}, \hat{\alpha}_{PRE}, \hat{\alpha}_{PLE}, \hat{\alpha}_{PLTE}, \hat{\alpha}_{ILTE}$ and $\hat{\alpha}_{PKLT}$, respectively.

Using (1.4), (1.5), (1.6), (1.7), (1.9), (1.10), (2.2), and (2.3), we can calculate the MMSE of the considered estimators as follows:

$$MMSE\left(\hat{\beta}_{MLE}\right) = Q\Lambda^{-1}Q' \tag{2.5}$$

$$MMSE\left(\hat{\beta}_{PRE}\right) = Q\left((\Lambda + kI)^{-1}\Lambda(\Lambda + kI)^{-1} + k^2(\Lambda + kI)^{-1}\alpha\alpha'(\Lambda + kI)^{-1}\right)Q' \tag{2.6}$$

$$MMSE\left(\hat{\beta}_{PLE}\right) = Q\left((\Lambda + I)^{-1}(\Lambda + dI)\Lambda^{-1}(\Lambda + dI)(\Lambda + I)^{-1} + (d-1)^2(\Lambda + I)^{-1}\alpha\alpha'(\Lambda + I)^{-1}\right)Q' \tag{2.7}$$

$$MMSE\left(\hat{\beta}_{PLTE}\right) = Q\left((\Lambda + kI)^{-1}(\Lambda + dI)\Lambda^{-1}(\Lambda + dI)(\Lambda + kI)^{-1} + (d-k)^2(\Lambda + kI)^{-1}\alpha\alpha'(\Lambda + kI)^{-1}\right)Q' \tag{2.8}$$

$$MMSE\left(\hat{\beta}_{ILTE}\right) = Q\left((\Lambda+kI)^{-1}(\Lambda+f(k)I)\Lambda^{-1}(\Lambda+f(k)I)(\Lambda+kI)^{-1} + (f(k)-k)^2(\Lambda+kI)^{-1}\alpha\alpha'(\Lambda+kI)^{-1}\right)Q' \tag{2.9}$$

$$MMSE\left(\hat{\beta}_{PKLE}\right) = Q(\Lambda+kI)^{-1}(\Lambda-kI)\Lambda^{-1}(\Lambda-kI)(\Lambda+kI)^{-1}Q' + bias\left(\hat{\beta}_{PKLR}\right)bias\left(\hat{\beta}_{PKLR}\right)'$$
$$\text{where } bias\left(\hat{\beta}_{PKLE}\right) = \left((\Lambda+kI)^{-1}(\Lambda-kI)-I\right)Q\alpha \tag{2.10}$$

$$MMSE\left(\hat{\beta}_{PKLTEI}\right) = Q(\Lambda+kI)^{-1}(\Lambda-kI)(\Lambda+kI)^{-1}\Lambda(\Lambda+kI)^{-1}(\Lambda-kI)(\Lambda+kI)^{-1}Q'$$
$$+ bias\left(\hat{\beta}_{PKLTEI}\right)bias\left(\hat{\beta}_{PKLTEI}\right)' \text{where } bias\left(\hat{\beta}_{PKLTEI}\right) = \left((\Lambda+kI)^{-1}(\Lambda-kI)(\Lambda+kI)^{-1}\Lambda-I\right)Q\alpha \tag{2.11}$$

$$MMSE\left(\hat{\beta}_{PKLTEII}\right) = Q(\Lambda+kI)^{-1}(\Lambda-kI)(\Lambda+kI)^{-1}(\Lambda-kI)\Lambda^{-1}(\Lambda-kI)(\Lambda+kI)^{-1}(\Lambda-kI)(\Lambda+kI)^{-1}Q'$$
$$+ bias\left(\hat{\beta}_{PKLTEII}\right)bias\left(\hat{\beta}_{PKLTEII}\right)' \tag{2.12}$$

$$\text{where } bias\left(\hat{\beta}_{PKLTEII}\right) = \left((\Lambda+kI)^{-1}(\Lambda-kI)(\Lambda+kI)^{-1}(\Lambda-kI)-I\right)Q\alpha$$

Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be any two estimators of β. Then, $\hat{\beta}_2$ is superior to $\hat{\beta}_1$ with respect to the MMSE criterion if and only if $MMSE\left(\hat{\beta}_1\right) - MMSE\left(\hat{\beta}_2\right)$ is a positive-definite (pd) matrix. If $MMSE\left(\hat{\beta}_1\right) - MMSE\left(\hat{\beta}_2\right)$ is a nonnegative definite matrix, then $SMSE\left(\hat{\beta}_1\right) - SMSE\left(\hat{\beta}_2\right) \geq 0$. However, the opposite is not always true (Theobald, 1974).

The following theorem can be used to compare the MMSEs of biased estimators.

Theorem 2.1. (Farebrother, 1976). Let c be a nonzero vector and A be a positive-definite matrix, namely $A > 0$. Then, $A - cc'$ is a positive-definite matrix iff $c'A^{-1}c \leq 1$.

3. THE SUPERIORITY OF THE PKLTE IN PRMS

In this part, we use the MMSE criterion to compare the PKLTE II and ILTE. As a result, using several $f(k)$ functions allows a more comprehensive assessment of the estimator's performance.

Theorem.3.1: Let be $k > 0$ and $\left(\lambda_j + k\right)^2\left(\lambda_j + f(k)\right)^2 - \left(\lambda_j - k\right)^4 > 0$ where $j = 1, ..., p+1$. Then $MMSE\left(\hat{\beta}_{ILTE}\right) - MMSE\left(\hat{\beta}_{PKLTE\ II}\right) > 0$ iff

$$bias\left(\hat{\beta}_{PKLTE\ II}\right)' Q\left((\Lambda+kI)^{-1}(\Lambda+f(k)I)\Lambda^{-1}(\Lambda+kI)^{-1}(\Lambda+f(k)I)\right.$$
$$\left.-(\Lambda+kI)^{-1}(\Lambda-kI)(\Lambda+kI)^{-1}(\Lambda-kI)\Lambda^{-1}(\Lambda-kI)(\Lambda+kI)^{-1}(\Lambda-kI)(\Lambda+kI)^{-1}\right)^{-1} Q'bias\left(\hat{\beta}_{PKLTE\ II}\right) < 1 \tag{3.1}$$

where $bias\left(\hat{\beta}_{PKLTE\ II}\right) = \left((\Lambda+kI)^{-1}(\Lambda-kI)(\Lambda+kI)^{-1}\Lambda-I\right)Q\alpha$.

Proof. Using (2.9) and (2.12), we obtain

$$\begin{aligned}
&MMSE\left(\hat{\beta}_{ILTE}\right) - MMSE\left(\hat{\beta}_{PKLTE\ II}\right)\\
&= Q\left(AA' - BB'\right)Q' + bias\left(\hat{\beta}_{ILTE}\right)bias\left(\hat{\beta}_{ILTE}\right)' - bias\left(\hat{\beta}_{PKLTE\ II}\right)bias\left(\hat{\beta}_{PKLTE\ II}\right)'\\
&= Q\left((\Lambda+kI)^{-1}(\Lambda+f(k)I)\Lambda^{-1}(\Lambda+kI)^{-1}(\Lambda+f(k)I)\right.\\
&\left.-(\Lambda+kI)^{-1}(\Lambda-kI)(\Lambda+kI)^{-1}(\Lambda-kI)\Lambda^{-1}(\Lambda-kI)(\Lambda+kI)^{-1}(\Lambda-kI)(\Lambda+kI)^{-1}\right)^{-1} Q'\\
&+ bias\left(\hat{\beta}_{ILTE}\right)bias\left(\hat{\beta}_{ILTE}\right)' - bias\left(\hat{\beta}_{PKLTE\ II}\right)bias\left(\hat{\beta}_{PKLTE\ II}\right)'\\
&= Q\ diag\left\{\frac{\left(\lambda_j+f(k)\right)^2}{\left(\lambda_j+k\right)^2\lambda_j} - \frac{\left(\lambda_j-k\right)^4}{\lambda_j\left(\lambda_j+k\right)^4}\right\}_{j=1}^{p+1} Q' + bias\left(\hat{\beta}_{ILTE}\right)bias\left(\hat{\beta}_{ILTE}\right)' - bias\left(\hat{\beta}_{PKLTE\ II}\right)bias\left(\hat{\beta}_{PKLTE\ II}\right)'.
\end{aligned}$$

We can observe that $AA' - BB' > 0$ if and only if $\left(\lambda_j + k\right)^2\left(\lambda_j + f(k)\right)^2 - \left(\lambda_j - k\right)^4 > 0$ where $j = 1, 2, ..., p+1$. Therefore, $AA' - BB'$ is the pd matrix. The proof is completed by Theorem 2.1. completes the proof.

4. SELECTION OF BIASING PARAMETER

In general, the most important parameter affecting the estimator performance is the biasing parameter. However, many techniques can be used to determine an appropriate statistic for estimating the biasing parameter. In general, values that minimize the SMSE function with respect to the biasing parameter are usually recommended as estimates of the biasing parameter. Initially, to find the optimal k for PKLTE I, the function $h_1(k) = SMSE(\hat{\beta}_{PKLTEI})$ is given as follows:

$$SMSE(\hat{\beta}_{PKLTEI}) = \sum_{j=1}^{p+1} \frac{\lambda_j(\lambda_j - k)^2}{(\lambda_j + k)^4} + \sum_{j=1}^{p+1} \frac{k^2(k + 3\lambda_j)^2\alpha_j^2}{(\lambda_j + k)^4} \tag{4.1}$$

The derivative of $h_1(k)$ in relation to parameter k is given as follows:

$$h_1^{'}(k) = \sum_{j=1}^{p+1} \frac{2\lambda_j(k - 3\lambda_j)(\lambda_j - k - k\alpha_j^2(k + 3\lambda_j))}{(k + \lambda_j)^5} \tag{4.2}$$

When it is $h_1'(k) = 0$, we have the following:

$$\begin{aligned} k_{PKLTEI(1)} &= 3\lambda_j \\ k_{PKLTEI(2)} &= -\frac{1 + 3\alpha_j^2\lambda_j + \sqrt{\left(1 + \alpha_j^2\lambda_j\right)\left(1 + 9\alpha_j^2\lambda_j\right)}}{2\alpha_j^2} \\ k_{PKLTEI(3)} &= \frac{-1 - 3\alpha_j^2\lambda_j + \sqrt{\left(1 + \alpha_j^2\lambda_j\right)\left(1 + 9\alpha_j^2\lambda_j\right)}}{2\alpha_j^2} \end{aligned} \tag{4.3}$$

where $j = 1, 2, ..., p + 1$.
Similarly, the $SMSE(\hat{\beta}_{PKLTEII})$ function of the PKLTE II estimator is as follows:

$$SMSE(\hat{\beta}_{PKLTEII}) = \sum_{j=1}^{p+1} \frac{(\lambda_j - k)^4}{\lambda_j(\lambda_j + k)^4} + \sum_{j=1}^{p+1} \frac{16k^2\lambda_j^2\alpha_j^2}{(\lambda_j + k)^4} \tag{4.4}$$

To determine the optimal k, the derivative of $h_2(k) = SMSE(\hat{\beta}_{PKLTEII})$ with respect to k is given as follows:

$$h'_2(k) = \sum_{j=1}^{p+1} \frac{8\left(k - \lambda_j\right)\left(k^2 - 2k\lambda_j + \left(1 - 4k\alpha_j^2\right)\lambda_j^2\right)}{\left(k + \lambda_j\right)^5} \tag{4.5}$$

When it is accepted $h_2^{'}(k) = 0$, we have:

$$\begin{aligned} k_{PKLTEII(1)} &= \lambda_j \\ k_{PKLTEII(2)} &= \lambda_j + 2\alpha_j^2\lambda_j^2 - 2\sqrt{\alpha_j^2\lambda_j^3\left(1 + \alpha_j^2\lambda_j\right)} \\ k_{PKLTEII(3)} &= \lambda_j + 2\alpha_j^2\lambda_j^2 + 2\sqrt{\alpha_j^2\lambda_j^3\left(1 + \alpha_j^2\lambda_j\right)} \end{aligned} \tag{4.6}$$

where the biasing parameter k depends on $\hat{\alpha}_j^2$, $j1, 2, ..., p + 1$. To find the estimators of k, we substitute their unbiased estimator $\hat{\alpha}_j^2$ for them for practical purposes. Note that $h_1(k)$ and $h_2(k)$ are nonlinear functions of k. Numerical methods are used to minimize the values of these functions relative to k. To determine the approximate minimum values of $h_1(k)$ or $h_2(k)$, we can make some approximations based on the obtained roots. The biasing parameter k can be estimated

using the following estimators based on the simulation results:$\hat{k} = min(\lambda_j), \hat{k} = median(\lambda_j), \hat{k} = mean(\lambda_j)$ and $\hat{k} = quantile(\lambda_j, q)$ where q is the probability value used to generate sample quantiles and $j = 1, 2, ..., p + 1$.

5. THE MONTE CARLO SIMULATION STUDIES

In this section, we design two simulation designs to investigate the performance of PKLTE over other existing biased estimators in PRMs. In the first simulation design, we discuss the effects of sample size (n), the number of the explanatory variables (p) and the degree of the collinearity (ρ) on the behavior of the PRE, PLE, PLTE, PKLE, PKLTE I, and PKLTE II. In the second simulation scheme, we examine the behavior of the biasing parameter on the performances of PRE, PLE, PKLE, PKLTE I, and PKLTE II for each set of (n, ρ, p) . For both simulations, we generate the explanatory variables by following McDonald & Galarneau (1975), Asar & Genç (2018), & Akay & Ertan (2022):

$$x_{ij} = (1 - \rho^2)^{\frac{1}{2}} w_{ij} + \rho w_{ip+1}, i = 1, 2, .., n, j = 1, 2, ..., p \tag{5.1}$$

where ρ is defined such that the correlation between any two variables is given by ρ^2, and w_{ij} are independent standard normal pseudorandom numbers. Three correlation sets are examined, each of which corresponds to ρ=0.85,0.90, and 0.95. The number of explanatory variables selected is p=2,4, and 8. The sample sizes n were 100, 200, and 500. The parameter vector β is selected as the normalized eigenvector corresponding to the greatest eigenvalue of $X'X$ for every set of explanatory variables, so that $\beta'\beta = 1$. In addition, we set the intercept to zero.

In the simulation and application sections, the works of Månsson & Shukur (2011), Månsson et al. (2012), Kibria et al. (2015), Asar & Genc (2018), Alanaz & Algamal (2018), Qasim et al. (2020a), and Akay & Ertan (2022) are used for optimal estimates of biasing parameters for PRE, PLE, and PLTE.

For the biasing parameter k in PRE, we used the optimal estimate of k as$\hat{k}_{PRE} = max\left(\frac{1}{m_j}\right)$ where

$$m_j = \sqrt{\frac{\hat{\sigma}^2}{\hat{\alpha}_j^2}}, j = 1, 2, ..., p \text{ and } \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2}{(n - p - 1)} \text{which is given by Kibria et al. (2015).}$$

Based on the results of Qasim et al. (2020a), we use the optimal estimate of d in PLE as

$$\hat{d}_{PLE} = max\left(0, min\left(\frac{\hat{\alpha}_j^2 - 1}{max(\frac{1}{\lambda_j}) + \hat{\alpha}_{max}^2}\right)\right)$$

Three methods were considered to estimate the biasing parameters k and d of PLTE:

$$PLTEI : \hat{k}_{PLTE} = max\left(\frac{1}{m_j}\right) \text{where } m_j = \sqrt{\frac{\hat{\sigma}^2}{\hat{\alpha}_j^2}}, j = 1, 2, ..., p \text{ and } \hat{d}_{PLTE} = \frac{\sum_{j=1}^{p} \frac{1 - \hat{k}_{PLTE}\hat{\alpha}_j^2}{\left(\lambda_j + \hat{k}_{PLTE}\right)^2}}{\sum_{j=1}^{p} \frac{1 + \lambda_j\hat{\alpha}_j^2}{\lambda_j\left(\lambda_j + \hat{k}_{PLTE}\right)^2}}$$

$$PLTEII : \hat{k}_{PLTE} = \frac{\lambda_1 - 100\lambda_p}{99} \text{ and } \hat{d}_{PLTE} = \frac{\sum_{j=1}^{p} \frac{1 - \hat{k}_{PLTE}\hat{\alpha}_j^2}{\left(\lambda_j + \hat{k}_{PLTE}\right)^2}}{\sum_{j=1}^{p} \frac{1 + \lambda_j\hat{\alpha}_j^2}{\lambda_j\left(\lambda_j + \hat{k}_{PLTE}\right)^2}}$$

$$PLTEIII : \hat{d}_{PLTE} = \frac{1}{2} min\left\{\frac{\lambda_j}{1 + \lambda_j\hat{\alpha}_j^2}\right\} \text{ and } \hat{k}_{PLTE} = \frac{1}{p}\sum_{j=1}^{p} \frac{\lambda_j - \hat{d}_{PLTE}(1 + \lambda_j\hat{\alpha}_j^2)}{(\lambda_j\hat{\alpha}_j^2)}$$

The k values for PKLE, PKLTE I, and PKLTE II are estimated using $\hat{k} = quantile\left(\lambda_j, q = \frac{8p-16}{100}\right)$ where p is the number of variables.

A comparison of the proposed estimators is based on the performance of the estimated MSEs (EMSEs), which are computed for an estimator $\hat{\beta}$ of β as

$$EMSE\left(\hat{\beta}\right) = \frac{1}{N}\sum_{r=1}^{N}\sum_{j=1}^{p+1}\left(\hat{\beta}_{rj} - \beta_j\right)^2 \tag{5.2}$$

where $\hat{\beta}_{rj}$ denotes the estimate of the j-th parameter in r-th replication, β_j are the true parameter values and N is the number of replications. The experiment is repeated 2000 times by creating response variables for each n, p and ρ. Using the R programing language, we conducted our Monte Carlo simulation studies, and Table 2 presents the results.

Table 2. The EMSE values of the estimators for the model with p = 2, 4, and 8

ID	p	n	ρ	MLE	PRE	PLE	PLTE I	PLTE II	PLTE III	PKLE	PKLTE I	PKLTE II
1	2	100	0.85	5.7538	0.4173	0.6121	2.7869	2.9650	0.9499	0.4039***	0.3767**	**0.3650***
2	2	100	0.9	6.3544	0.4055	0.5847	3.0471	3.2076	0.9489	0.3858***	0.3590**	**0.3455***
3	2	100	0.95	10.8516	0.4072***	0.4867	5.0765	5.2186	0.8905	0.4121	0.3873**	**0.3670***
4	2	200	0.85	3.7320	0.4376	0.6913	1.8306	2.0112	0.9749	0.3808**	**0.3790***	0.4124***
5	2	200	0.9	6.6718	0.4123	0.5603	3.2277	3.3813	0.9257	0.3893***	0.3640**	**0.3503***
6	2	200	0.95	10.4451	0.3984***	0.4955	5.0736	5.2371	0.9384	0.4030	0.3747**	**0.3523***
7	2	500	0.85	4.0880	0.4278	0.6749	2.0199	2.2165	0.9888	0.3707**	**0.3643***	0.3923***
8	2	500	0.9	6.4054	0.4043***	0.5793	3.0480	3.2379	0.9691	0.4096	0.3762**	**0.3570***
9	2	500	0.95	10.3092	0.4273***	0.5140	4.8168	4.9829	0.9360	0.4371	0.4077**	**0.3841***
10	4	100	0.85	8.7874	0.4113	0.8322	3.1125	4.5803	0.9791	0.3021***	0.2655**	**0.2613***
11	4	100	0.9	11.8191	0.3594***	0.7546	4.1368	5.5644	1.0489	0.4161	0.2991**	**0.2405***
12	4	100	0.95	33.6826	0.3430**	0.5055	12.0338	13.9604	1.2690	0.6725	0.3530***	**0.2318***
13	4	200	0.85	9.9565	0.3975	0.8404	3.6038	4.8777	1.0626	0.3100***	0.2621**	**0.2490***
14	4	200	0.9	17.4294	0.3327***	0.6551	6.2684	7.4587	1.1962	0.3395	0.2566**	**0.2183***
15	4	200	0.95	29.3244	0.3489***	0.5191	10.3539	11.4370	1.2341	0.4108	0.2857**	**0.2325***
16	4	500	0.85	10.7389	0.3837	0.8153	3.8824	4.3418	1.0971	0.2480***	**0.2356***	0.2387**
17	4	500	0.9	14.9318	0.3411	0.6729	5.1903	5.8912	1.0949	0.2803***	0.2435**	**0.2267***
18	4	500	0.95	34.2084	0.3582***	0.4690	11.7316	12.3691	1.1561	0.3883	0.2838**	**0.2371***
19	8	100	0.85	24.9982	0.353***	1.4524	7.3804	10.0366	1.1071	0.4055	0.2044**	**0.1371***
20	8	100	0.9	33.3556	0.2801**	1.2469	9.9121	13.6827	1.1946	1.0685	0.2989***	**0.1852***
21	8	100	0.95	75.5253	0.2181**	0.7280	22.0949	24.0267	1.6149	1.5832	0.4653***	**0.1764***
22	8	200	0.85	21.1630	0.4074***	1.5106	6.3536	9.4580	1.0062	0.4180	0.2081**	**0.1535***
23	8	200	0.9	37.3296	0.2736***	1.1506	11.233	13.9802	1.2988	0.7033	0.2419**	**0.1458***
24	8	200	0.95	70.3848	0.2282**	0.7576	21.1294	22.5549	1.6376	0.7787	0.3043***	**0.1282***
25	8	500	0.85	24.2082	0.3560	1.3791	7.2097	8.9182	1.0212	0.2346***	0.1616**	**0.1424***
26	8	500	0.9	37.9996	0.2722***	1.0581	11.2893	12.6393	1.1571	0.3217	0.1766**	**0.1337***
27	8	500	0.95	76.1531	0.2428***	0.6816	22.4592	23.1806	1.4282	0.4454	0.2118**	**0.1325***

The estimators with the lowest EMSE values are indicated in the table by bolded numerals. The second and third smallest EMSE values are denoted by the signs (**) and (***), respectively.

The results obtained in Table 2 are listed below:

1) When p and ρ are kept constant, PRE, PLE, PKLE, and PKLTE II exhibit stable behavior as the number of observations increases. In contrast, PKLTE I shows a decreasing effect for large variables and high correlation values.

2) When the number of observations (n) and ρ are kept constant, the EMSE values of PRE, PKLTE I, and II decrease as the number of variables increases, whereas PLE and PKLE increase.

3) When n and p in the model are kept constant, PRE, PKLTE I, and PKLTE II are more robust than the other estimators as the correlation increases. In contrast, the EMSE of PKLE increased as the correlation increased, whereas PKLE decreased as the number of observations increased for large observation values.

As a result, when all cases are analyzed, the estimator with the smallest EMSE values is PKLTE II.

In the second simulation scheme, we investigated the performances of PRE, PLE, PKL, PKLTE I, and PKLTE II for each n, ρ, and p. The purpose of this simulation is to examine the performances of PRE, PLE, PKLE, PKLTE I, and PKLTE II at various values of k according to the EMSE values given in (5.2). The second simulation approach did not estimate the biasing parameter k. Only the EMSE values derived by increasing k values in the [0,1] range by 0.1 are compared. Depending on these n, ρ, and p values, the explanatory variables are generated according to equation (5.1). The simulation is conducted 2000 times for each k value. Figures 1, 2, and 3 graphically show the results.

The following results can be obtained based on Figures 1-3:

1) The EMSE values of PRE tend to decrease with increasing values of k. On the contrary, PLE tends to increase as the biasing parameter increases.

2) The EMSE values of PKLE, PKLTE I, and PKLTE II generally decreased faster than in PRE.

3) The EMSE values of the PKLTE I and PRE behave in almost the same way, whereas PKLE and PKLTE II show an increase in the EMSE values after a certain value of k.

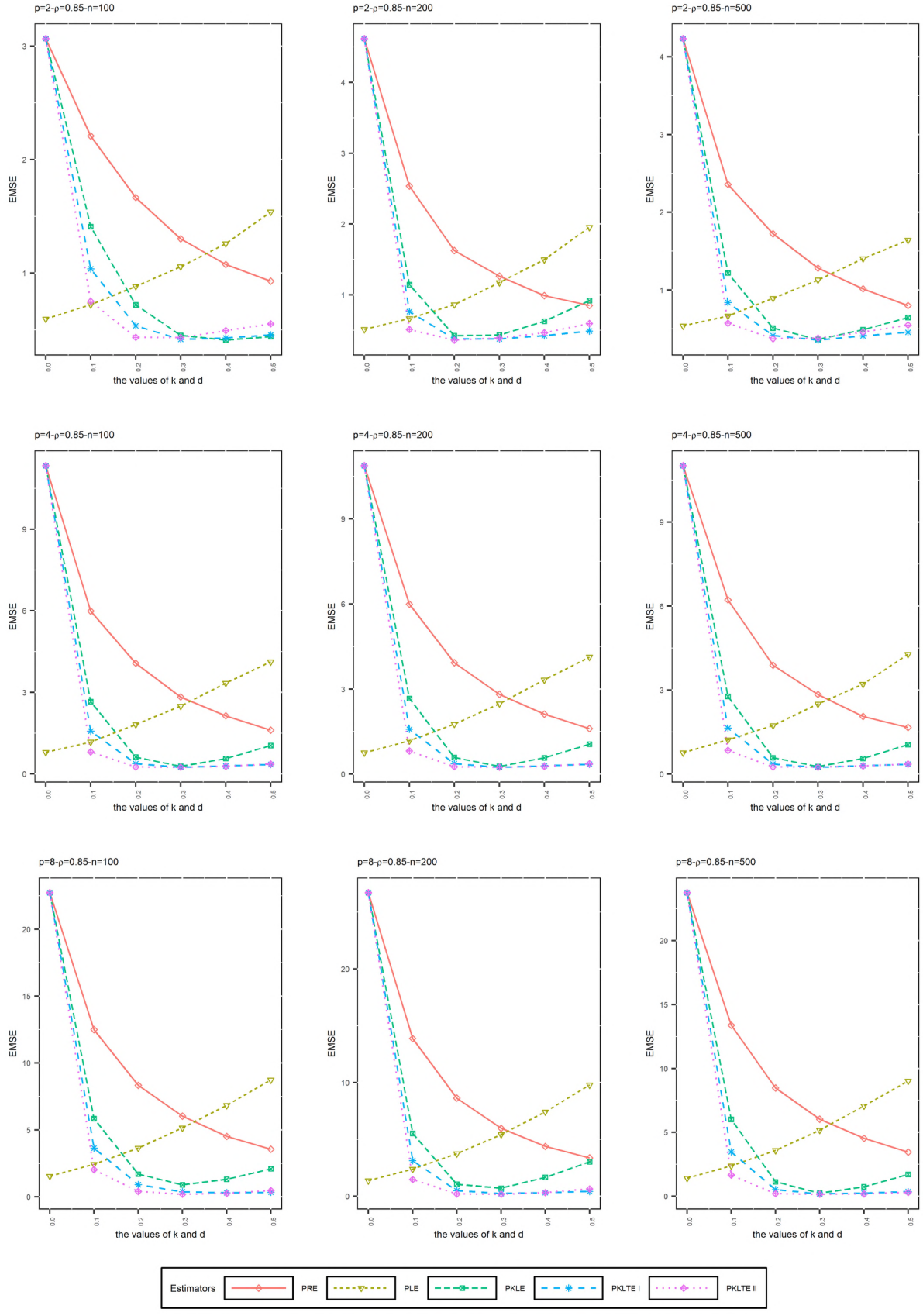


Figure 1. The EMSE values of PRE, PLE, PKLE, PKLTE I, and PKLTE II as a function k and d where ρ=0.85

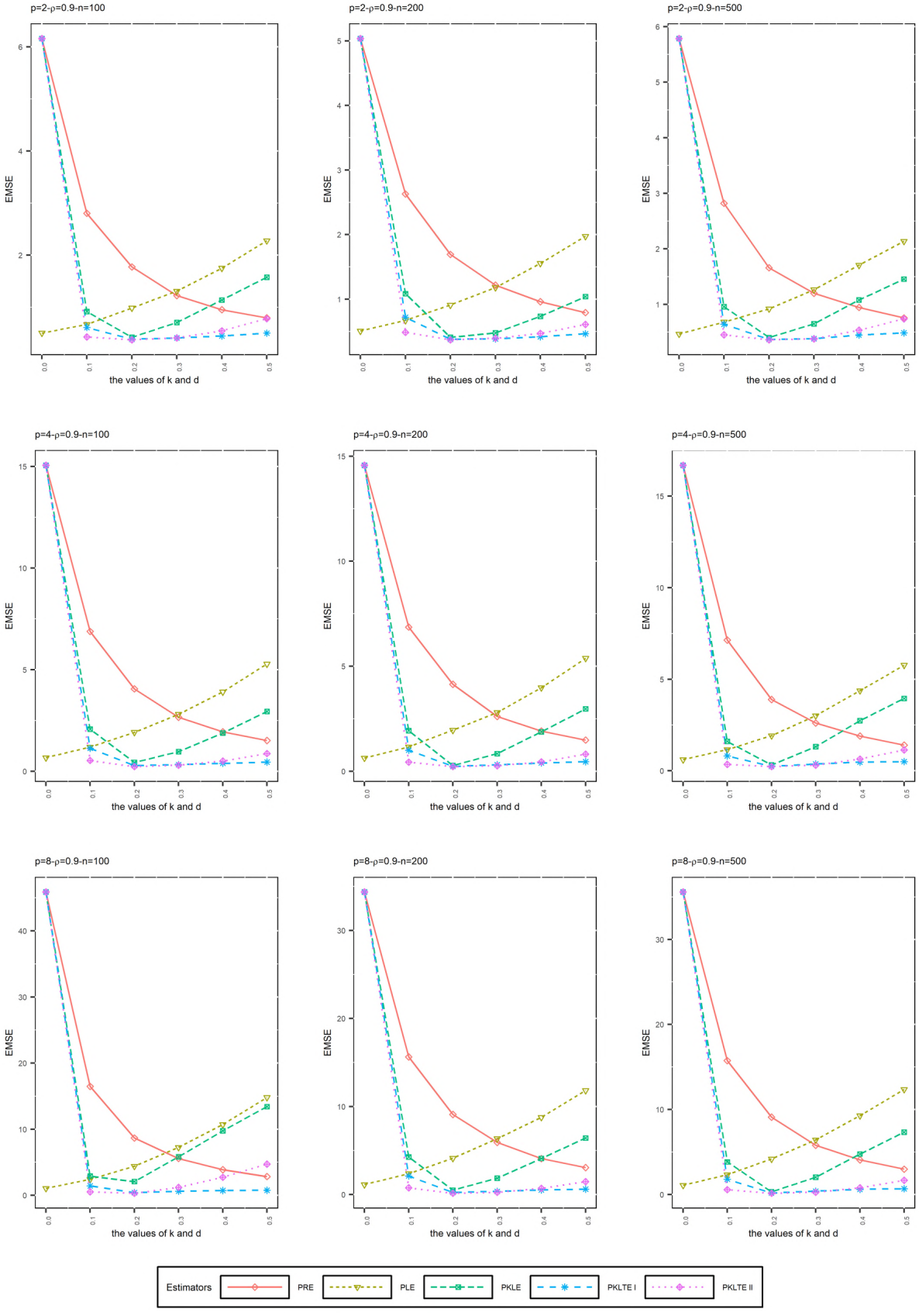


Figure 2. The EMSE values of PRE, PLE, PKLE, PKLTE I, and PKLTE II as functions of k and d where ρ=0.90

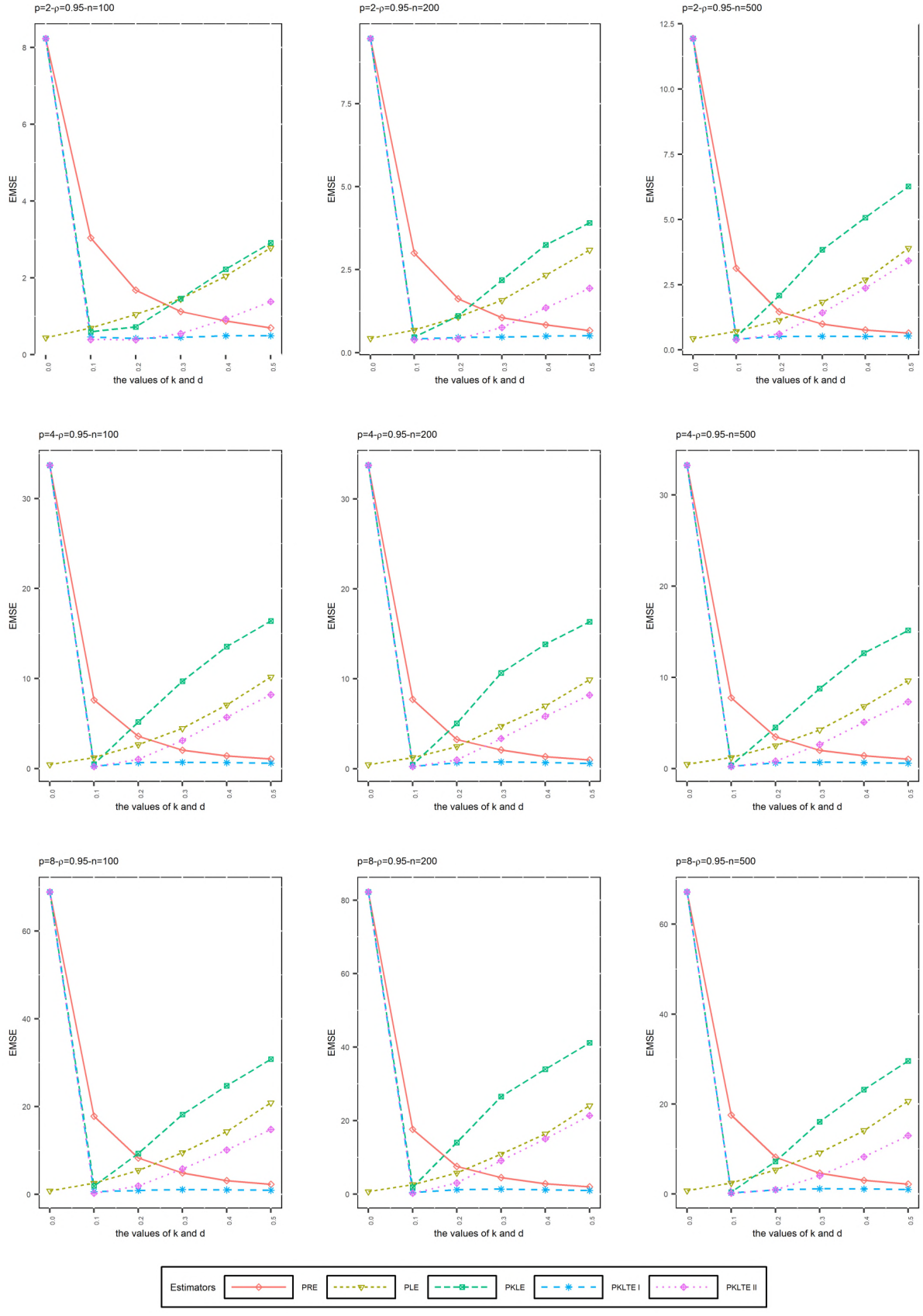


Figure 3. The EMSE values of PRE, PLE, PKLE, PKLTE I, and PKLTE II as functions of k and d where ρ=0.95.

6. AN EXAMPLE: THE AIRCRAFT DAMAGE DATA

This section examines the performance of PKLTE by considering aircraft damage data. Asar & Genc (2018), Myers et al. (2012), Lukman et al. (2021), Lukman et al. (2022), Amin et al. (2022), Akay & Ertan (2022), and Ertan & Akay (2023) also used these data. There are three explanatory variables and thirty observations in this data set. The kind of aircraft is indicated by the dichotomous explanatory variable (x_1) . The bomb load in tons and the total number of months of aircrew experience the explanatory variables (x_2) and (x_3), respectively. The number of locations where the aircraft was damaged is represented by the count variable y.

Asar & Genc (2018), Amin et al. (2022), and Akay & Ertan (2022) described the effects and solutions due to multicollinerity in the following model: $\mu = \exp(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)$. The eigenvalues of $X^{'}X$ are 2085.2251, 374.8961 and 4.3333. As a result, condition number 219.3654 suggests that there is a problem with multicollinearity among the explanatory variables. Additionally,$\lambda_1 = 283543.5$, $\lambda_2 = 789.85$, $\lambda_3 = 4.2887$ and $\lambda_4 = 1.2585$ are the eigenvalues of $X^{'}\hat{W}X$. The condition number is 474.653, which is significantly greater than 30, suggesting that multicollinearity continues to have an impact on MLE.

Table 3 summarizes the numerical results for comparing the PKLTEs with the other existing estimators. The average of the MLE values determined by the bootstrap sampling technique is considered as a true parameter to compute the SMSE values of the biased estimators. Table 3 shows that compared with the other estimators under consideration, PKLTEs produce good results in terms of variance and SMSE values.

Table 3. The parameter and SMSE values of the estimators

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\text{var}\left[\hat{\beta}\right]$	$SMSE\left[\hat{\beta}\right]$
$\hat{\beta}_{MLE}$	0.1262	1.5576	2.6710	-1.4157	3.8847	
$\hat{\beta}_{PRE}\ \left(\hat{k}_{PRE}=2.7044\right)$	0.4200	0.7970	1.2620	-0.5504	0.2329	0.2391
$\hat{\beta}_{PLE}\ \left(\hat{d}_{PLE}=0\right)$	0.2994	1.1554	1.8704	-0.8852	0.6714	0.6736
$\hat{\beta}_{PLTE\,\text{I}}\ \left(\hat{k}=2.7044, \hat{d}=-1.2946\right)$	0.2793	1.1611	1.9365	-0.9646	1.3472	1.3489
$\hat{\beta}_{PLTE\,\text{II}}\ \left(\hat{k}=0.1088, \hat{d}=0.3268\right)$	0.2329	1.3552	2.1716	-1.0640	0.6947	0.6961
$\hat{\beta}_{PLTE\,\text{III}}\ \left(\hat{k}=36.3139, \hat{d}=0.0484\right)$	0.3820	0.1383	0.2081	-0.0787	0.0093	0.0494
$\hat{\beta}_{PKLE}\ \left(\hat{k}=0.8681\right)$	0.4409	0.8377	1.2194	-0.4451	0.5924	0.5998
$\hat{\beta}_{PKLTE\,I}\ \left(\hat{k}=0.8681\right)$	0.5029	0.6212	0.9187	-0.3035	0.1200	0.1295
$\hat{\beta}_{PKLTE\,II}\ \left(\hat{k}=0.8681\right)$	0.5649	0.4047	0.6179	-0.1619	0.1109	0.1231

We now wish to examine, in terms of MMSE, the performance of the PKLTE II and ILTE that were derived from the selection of different $f(k)$ functions. We replace with the estimates obtained from the bootstrap sampling approach. Let us take $f(k) = 0.05k + 0.05$ for ILTE. In this instance, $cov(\hat{\beta}_{ILTE}) - cov(\hat{\beta}_{PKLTEII})$ is the pd matrix for $0 < k \leq 1.499$. In addition, the values of k satisfying the inequality in (3.1) are $0 < k \leq 1.245$. Consequently, $MMSE\left(\hat{\beta}_{ILTE}\right) - MMSE\left(\hat{\beta}_{PKLTEII}\right)$ is the pd matrix where $0 < k \leq 1.245$. Let us take $f(k) = 0.05k - 0.01$ for another comparison. In this case, $cov(\hat{\beta}_{ILTE}) - cov(\hat{\beta}_{PKLTEII})$ is pd matrix for $0 < k \leq 1.1679$. Additionally, $0 < k \leq 1.1676$ are the values of k that satisfy the inequality in (3.1). Thus, $MMSE(\hat{\beta}_{ILTE}) - MMSE(\hat{\beta}_{PKLTEII})$ is the pd matrix where $0 < k \leq 1.1676$.

7. CONCLUSION

In this paper, we propose a new biased estimator for PRMs called PKLTE as an alternative to MLE and other existing biased estimators in the presence of multicollinearity. The PKLTE is a general estimator that includes PKLE and its variations. We investigated the properties of PKLTE and proposed several estimators to estimate the biasing parameter. The performance of the proposed PKLTEs was evaluated using Monte Carlo simulations. The findings demonstrate

that in the scenario of low-moderate-high multicollinearity, the proposed PKLTE II performs better than the existing estimators. Additionally, a generic simulation study is provided to compare PRE, PLE, PKLE, PKLTE I, and PKLTE II. It can be observed that PKLTE I and PKLTE II exhibit a faster decrease in EMSE values than PRE when the biasing parameter k is varied in the range [0,1]. We can also say that PKLE and PKLTE II reach minimum EMSE values in this range. Furthermore, the considered estimators were applied to real data, and the results were found to be consistent with the simulation study. Therefore, based on the simulation and application results, PKLTE II is recommended when there is multicollinearity in PRMs.

Peer Review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- C.Ç., K.U.A.; Data Acquisition- C.Ç.; Data Analysis/Interpretation- K.U.A.; Drafting Manuscript- K.U.A.; Critical Revision of Manuscript- C.Ç.; Final Approval and Accountability- C.Ç., K.U.A.; Supervision- C.Ç., K.U.A.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

ORCID IDs of the authors

Cemal Çiçek 0000-0002-4855-9386
Kadri Ulaş Akay 0000-0002-8668-2879

REFERENCES

Akay, K. U., Ertan, E., & Erkoç, A. (2023). A New biased estimator and variations based on the Kibria Lukman Estimator. *Istanbul Journal of Mathematics, 1*(2), 74-85.

Akay, K. U., & Ertan, E., (2022). A new Liu-type estimator in Poisson regression models. *Hacet J Math Stat, 51* (5), 1484–1503.

Aladeitan, B. B., Adebimpe, O., Lukman, A. F., Oludoun, O., & Abiodun, O. E. (2021). Modified Kibria-Lukman (MKL) estimator for the Poisson Regression Model: application and simulation. *F1000Research, 10.*

Alanaz, M. M., & Algamal, Z. Y., (2018). Proposed methods in estimating the ridge regression parameter in Poisson regression model. *Electronic Journal of Applied Statistical Analysis, 11*(2), 506-515.

Algamal, Z. Y. (2018). Biased estimators in Poisson regression model in the presence of multicollinearity: A subject review. *Al-Qadisiyah Journal for Administrative and Economic Sciences, 20*(1), 37-43.

Alheety, M. I., Qasim, M., Månsson, K., & Kibria, B. G. (2021). Modified almost unbiased two-parameter estimator for the Poisson regression model with an application to accident data. *SORT*, 121-142.

Alkhateeb, A., & Algamal, Z. (2020). Jackknifed Liu-type estimator in Poisson regression model. *Journal of The Iranian Statistical Society, 19*(1), 21-37.

Amin, M., Akram, M. N., & Amanullah, M. (2022). On the James-Stein estimator for the Poisson regression model. *Communications in Statistics-Simulation and Computation*, 51(10), 5596-5608.

Amin, M., Akram, M. N., & Kibria, B. G. (2021). A new adjusted Liu estimator for the Poisson regression model. *Concurrency and Computation: Practice and Experience, 33*(20), e6340.

Alrweili, H. (2024). Kibria–Lukman Hybrid Estimator for Handling Multicollinearity in Poisson Regression Model: Method and Application. *International Journal of Mathematics and Mathematical Sciences, 2024*(1), 1053397.

Asar, Y., & Genç, A. (2018). A new two-parameter estimator for the poisson regression model. *Iranian Journal of Science and Technology, Transactions A: Science, 42*(2), 793-803.

Çetinkaya, M. K., & Kaçıranlar, S. (2019). Improved two-parameter estimators for the negative binomial and Poisson regression models. *Journal of Statistical Computation and Simulation, 89*(14), 2645-2660.

Dawoud, I., Abonazel, M. R., & Awwad, F. A. (2022). Generalized Kibria-Lukman estimator: Method, simulation, and application. *Frontiers in Applied Mathematics and Statistics, 8*, 880086.

Dunn, P. K., & Smyth, G. K. (2018). Generalized Linear Models With Examples in R. Springer, New York, NY.

Ertan, E., & Akay, K. U. (2023). A new class of Poisson-ridge-type estimator. *Scientific Reports*, 13(1), 4968.

Farebrother, R.W. (1976). Further results on the mean square error of ridge regression. *J R Stat Soc B*, (28), 248-250.

Hardin, J. W., & Hilbe, J. M. (2018). Generalized linear models and extensions. Fourth edition. Stata press.

Hilbe, J. M. (2014). Modeling Count Data; Cambridge University Press: Cambridge.

Hoerl, A.E., & Kennard, R.W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.

Jadhav, N. H. (2022). A new linearized ridge Poisson estimator in the presence of multicollinearity. *Journal of Applied Statistics, 4*9(8), 2016-2034.

Kibria, B. M. G., Månsson, K., & Shukur, G. (2013). Some ridge regression estimators for the zero-inflated Poisson model. *Journal of Applied Statistics, 40*(4), 721-735.

Kibria B. M. G., Månsson K., & Shukur, G. (2015). A simulation study of some biasing parameters for the ridge type estimation of Poisson regression. *Commun Stat Simul Comput*, 44(4), 943–957.

Kurnaz, F. S., & Akay, K. U. (2015). A new Liu-type estimator. *Stat Papers*, 56, 495–517.

Liu, K. (1993). A new class of biased estimate in linear regression. *Commun Stat Theory Methods* 22(2): 393-402.

Liu, K. (2003). Using Liu-type estimator to combat collinearity. *Commun Stat Theory Methods* 32(5):1009-1020.

Lukman, A. F., Adewuyi, E., Månsson, K., & Kibria, B. G. (2021). A new estimator for the multicollinear Poisson regression model: simulation and application. *Scientific Reports, 11*(1), 3732.

Lukman, A. F., Aladeitan, B., Ayinde, K., & Abonazel, M. R. (2022). Modified ridge-type for the Poisson regression model: simulation and application. *Journal of Applied Statistics*, 49(8), 2124–2136.

Lukman, A. F., Allohibi, J., Jegede, S. L., Adewuyi, E. T., Oke, S., & Alharbi, A. A. (2023). Kibria–Lukman-Type Estimator for Regularization and Variable Selection with Application to Cancer Data. *Mathematics, 11*(23), 4795.

Månsson, K., & Shukur, G. (2011). A Poisson ridge regression estimator. *Economic Modelling, 28*(4), 1475-1481.

Månsson, K., Kibria, B. G., Sjolander, P., & Shukur, G. (2012). Improved Liu estimators for the Poisson regression model. *International Journal of Statistics and Probability*, 1(1), 2-6.

Månsson, K., & Kibria, B. G. (2020). Estimating the Unrestricted and Restricted Liu Estimators for the Poisson Regression Model: Method and Application. *Computational Economics*, 1-16.

McDonald, G. C., & Galarneau, D. I. (1975) A Monte Carlo evaluation of some ridge-type estimators. *J Am Stat Assoc*, 70(350), 407–416.

Myers, R. H., Montgomery, D. C., Vining, G. G., & Robinson, T. J. (2012). Generalized linear models: with applications in engineering and the sciences, Wiley, New York.

Özkale, M. R., & Kaçıranlar, S. (2007). The restricted and unrestricted two-parameter estimators. C*ommun Stat Theory Methods, 36*, 2707–2725.

Qasim, M., Kibria, B. M. G., Månsson, K., & Sjölander, P. (2020a). A new Poisson–Liu Regression Estimator: method and application. *Journal of Applied Statistics, 47*(12), 2258-2271.

Qasim, M., Månsson, K., Amin, M., Kibria, B. G., & Sjölander, P. (2020b). Biased adjusted Poisson ridge estimators-method and application. *Iranian Journal of Science and Technology, Transactions A:* Science, 44(6), 1775-1789.

Rashad, N. K., & Algamal, Z. Y. (2019). A New Ridge Estimator for the Poisson Regression Model. *Iranian Journal of Science and Technology, Transactions A: Science, 43*(6), 2921-2928.

Theobald, C.M. (1974) Generalizations of mean square error applied to ridge regression. *J R Stat So B* 36: 103-106.

Türkan, S., & Özel, G. (2016). A new modified Jackknifed estimator for the Poisson regression model. *Journal of Applied Statistics, 43*(10), 1892-1905.

How cite this article

Çiçek, C., & Akay, K. U. (2024). A New Kibria-Lukman-Type Estimator for Poisson Regression Models. *Acta Infologica, 8*(2), 199-212. https://doi.org/10.26650/acin.1558583

İSTANBUL UNIVERSITY PRESS

Acta Infologica
acin 2024, 8 (2), 213–221
DOI: 10.26650/acin.1521835

Research Article

Advanced Phishing Detection: Leveraging t-SNE Feature Extraction and Machine Learning on a Comprehensive URL Dataset

Taha Etem[1], Mustafa Teke[2]

[1]Çankırı Karatekin University, Computer Engineering Department, Çankırı, Türkiye
[2]Çankırı Karatekin University, Department of Electrical-Electronics Engineering, Çankırı, Türkiye

Corresponding author : Taha Etem
E-mail : tahaetem@karatekin.edu.tr

ABSTRACT

Phishing attacks continue to pose a major challenge in today's digital world; thus, sophisticated detection techniques are required to address constantly changing tactics. In this paper, we have proposed an innovative method to identify phishing attempts using the extensive PhiUSIIL dataset. The proposed dataset comprises 134,850 legitimate URLs and 100,945 phishing URLs, providing a robust foundation for analysis. We applied the t-SNE technique for feature extraction, condensing the original 51 features into only 2, while preserving high detection accuracy. We evaluated several machine learning algorithms on both full and reduced datasets, including Logistic Regression, Naive Bayes, k-Nearest Neighbors (kNN), Decision Trees, and Random Forest. The Decision Tree algorithm showed the best performance on the original dataset, achieving 99.7% accuracy. Interestingly, the proposed kNN demonstrated remarkable results on feature-extracted data, achieving 99.2% accuracy. We observed significant improvements in Logistic Regression and Random Forest performance when using the feature-extracted dataset. The proposed method offers substantial benefits in terms of computational efficiency. The feature-extracted dataset requires less processing power; thus, it is well-suited for systems with limited resources. These findings pave the way for developing more powerful and flexible phishing detection systems that can identify and neutralize emerging threats in real-time scenarios.

Keywords: Machine learning, cybersecurity, feature extraction, data mining

Submitted : 24.07.2024
Revision Requested : 16.09.2024
Last Revision Received : 19.09.2024
Accepted : 11.12.2024
Published Online : 13.12.2024

1. INTRODUCTION

Phishing attacks are among the most common and harmful types of cybercrime. They pose significant risks to both individuals and organizations. These attacks usually trick victims into giving away sensitive information such as passwords, credit card numbers, and other personal data, through deceptive emails, websites, or messages. Due to the increasing sophistication and frequency of phishing attacks, effective and adaptive detection mechanisms are required. Traditional detection methods that rely on rule-based systems and block lists often fail to recognize new and evolving phishing techniques (Garera, Provos, Chew, & Rubin, 2007).

Recently, machine learning (ML) and deep learning (DL) have shown significant potential in improving phishing detection and prevention (Alhudhaif, Almaslukh, Aseeri, Guler & Polat, 2023; Buyrukoğlu & Savaş, 2023). These advanced computational techniques use large datasets to identify patterns and anomalies indicative of phishing activities. ML models, such as support vector machines (SVM) and random forests (RF), have been widely used to classify emails and URLs as either phishing or legitimate (Bergholz et al., 2010). More recently, DL models, particularly neural networks, have improved detection accuracy by capturing complex patterns in data that traditional ML models may miss (Adebowale, Lwin & Hossain, 2019; Türk, Lüy & Barışçı, 2020). One major advantage of ML and DL in phishing detection is their ability to generalize from training data to identify previously unseen phishing attempts. This ability is crucial given the constantly changing nature of phishing tactics. Techniques such as convolutional neural networks (CNN) and recurrent neural networks (RNN) have been applied to various aspects of phishing detection, including email content analysis, URL feature extraction, and web page layout analysis (Aburrous, Hossain, Dahal & Thabtah, 2010). However, applying ML and DL techniques to phishing detection faces several challenges (Etem & Teke, 2024). Issues like data imbalance, feature selection, model interpretability, and computational resource requirements must be carefully addressed to develop effective and efficient detection systems. In addition, adversaries continuously evolve tactics to evade detection, necessitating ongoing adaptation and refinement of these models (Jain & Gupta, 2022).

In a paper, Researchers provide a comprehensive overview of current state-of-the-art machine learning and deep learning phishing detection techniques. They discussed various techniques, highlighted their strengths and weaknesses, and explored future research directions in this critical area of cybersecurity using extreme learning machines (ELM) to detect phishing (Yang et al., 2021). Another review paper analyzed various ML and DL techniques used in phishing detection, focusing on the importance of combining multiple features and algorithms to enhance detection accuracy (Divakaran & Oest, 2022). A previous study compared the effectiveness of different ML algorithms in detecting phishing attacks, focusing on models such as SVM, RF, and neural networks, and evaluated their performance in terms of accuracy, precision, and recall (Jishnu & Arthi, 2023). A systematic review explores the use of DL techniques for phishing email detection, examining various models, including CNNs and RNNs, and their effectiveness in identifying phishing emails (Thakur, Ali, Obaidat & Kamruzzaman, 2023). Another study presented a model to detect phishing attacks using ML algorithms like RF and decision trees, emphasizing the importance of feature selection and engineering in terms of improving detection accuracy (Alam et al., 2020). In another study, the authors proposed a novel approach to phishing website detection using a combination of ML and DL models, highlighting the improvements in detection accuracy achieved through this multilayered approach (Bibi et al., 2024). Another study investigated the application of sequential DL models like Multi-Head Attention and Temporal Convolutional Networks in detecting phishing websites, and evaluated their performance in terms of accuracy and efficiency (Gopali, Namin, Abri & Jones, 2024). Another study proposed a DL model for phishing email detection, which was trained and tested on a comprehensive dataset, and demonstrated high accuracy in identifying phishing emails and discussed the implications of using DL for real-time phishing detection (Atawneh & Aljehani, 2023).

Datasets play a crucial role in training ML models to identify and counter sophisticated phishing attempts, making them a notable contribution to the field of cybersecurity. The PhiUSIIL dataset includes several data types that are essential for effective phishing detection, such as numerical, categorical, and text data (Prasad & Chandra, 2024). Numerical features include metrics such as URL lengths and email attribute frequencies, and categorical data include domain types and keyword presence. In addition, the dataset contains unstructured text data from emails and web pages, which require specialized techniques to detect phishing patterns. The diversity of data types allows the development of robust ML models that can accurately distinguish between legitimate and phishing communications. The PhiUSIIL dataset was used to train and evaluate various ML models, including SVMs, RFs, and Neural Networks, to enhance phishing detection capabilities. Security companies and software developers use this dataset to develop real-time phishing detection systems that can analyze incoming web traffic and emails to block phishing URLs before they reach end-users. The proposed dataset is invaluable for academic and industrial research, providing a benchmark for new algorithms and facilitating feature engineering and model optimization studies. Despite its benefits, the use of ML in

phishing detection faces significant challenges. The evolving tactics of cybercriminals, the need for high-quality and diverse datasets, and the requirement for substantial computational resources are major hurdles. Phishing detection systems must constantly adapt to new threats and ensure compliance with relevant privacy and ethical standards. The proposed PhiUSIIL dataset represents a significant step forward in the phishing fight, and it provides a powerful tool to develop more effective cybersecurity measures.

In examining all of these studies, it is evident that the most critical feature of phishing attacks is their ability to deceive people using constantly updated methods. Consequently, phishing attacks attempt to avoid detection by matching features in old datasets to fake websites. Continuous development of systems to detect phishing attacks and increase success rates is particularly important. To this end, we designed a phishing detection system based on the PhiUSIIL dataset. To achieve the best results and design a fast lightweight system, feature extraction was performed using the proven t-SNE method (Bibal, Delchevalerie & Frénay, 2023), and system evaluation was performed using different ML algorithms. The results show that the proposed method can play a significant role in detecting current phishing attacks.

2. PhiUSIIL PHISHING DATASET

Phishing involves creating unauthorized replicas of legitimate websites and emails, typically from financial institutions, to deceive individuals into divulging confidential information. These fraudulent communications often use legitimate company logos and slogans to appear credible, exploiting the HTML structure that allows easy copying of images or entire websites (Prasad & Chandra, 2024).

The PhiUSIIL dataset includes various data points that are essential for training and testing ML models to identify phishing attempts. This dataset includes distinct types of data such as numerical, categorical, and text data, each serving specific purposes in the analysis. Numerical data in the PhiUSIIL dataset include quantitative metrics such as the frequency of specific email attributes, URL length, and other measurable factors that could indicate phishing. These numerical features help quantify the critical aspects of detection algorithms. Categorical data include distinct categories or classes such as the type of domain used, the presence of certain keywords in the email content, and the classification of email sources as legitimate or suspicious. This categorical information is vital for creating classification models that distinguish between phishing and nonphishing activities. The dataset also includes unstructured text data, including email and web page contents, which require specialized extraction and analysis techniques. Text analysis helps identify common phrases or patterns used in phishing attempts, thereby enhancing the model's ability to detect subtle cues that might otherwise be overlooked.

The PhiUSIIL dataset is important in data-driven decision-making in cybersecurity. By training ML models on this dataset, researchers can build more robust phishing detection systems. The dataset's diverse range of features allows for comprehensive analysis and helps transform raw data into meaningful insights, which ultimately contributes to the more effective prevention of phishing attacks. The PhiUSIIL Phishing URL Dataset is an extensive collection of 134,850 legitimate URLs and 100,945 phishing URLs; thus, it is a substantial resource for developing and testing phishing detection algorithms. This dataset was meticulously curated by analyzing the source code of webpages and URLs to extract features such as CharContinuationRate, URLTitleMatchScore, URLCharProb, and TLDLegitimateProb. These features are essential in distinguishing legitimate and phishing URLs, and they offer numerous applications in research and practical fields.

3. MATERIALS AND METHODS

Phishing detection has become an essential component of cybersecurity strategies due to the evolving tactics employed by malicious actors, which deceive users and steal sensitive information. Traditional detection methods often fall short in countering sophisticated attacks, which necessitates advanced ML solutions. ML algorithms like Logistic Regression, Naïve Bayes, K-Nearest Neighbors, Decision Tree, and random forest, are used for phishing detection (Alsaç, Yenisey, Ganiz, Dagtekin & Ulusinan, 2023; Doğruel & Soner Kara, 2023; Efeoğlu, 2022; Tülay, 2023; Yaman & Tuncer, 2023).

Feature extraction involves identifying and extracting relevant features from raw data for machine learning and data analysis (Güler & Yücedağ, 2022). These features are then used to create an informative dataset for tasks such as classification, prediction, and clustering. The goal is to reduce data complexity or dimensionality while retaining as much relevant information as possible, thereby enhancing the performance and efficiency of machine learning algorithms. This process may include the creation of new features and data manipulation to separate meaningful and irrelevant features (Jiang, Shi, Liang & Liu, 2024).

t-SNE, or t-distributed Stochastic Neighbor Embedding, is a potent dimensionality reduction technique used for visualizing high-dimensional data. Developed by Laurens van der Maaten and Geoffrey Hinton in 2008, t-SNE has become a popular machine learning and data science tool for feature extraction and data visualization. The method functions by converting similarities between data points into joint probabilities and aims to minimize the Kullback-Leibler divergence between the joint probabilities of low-dimensional embedding and high-dimensional data. Unlike linear dimensionality reduction methods such as PCA, t-SNE is particularly adept at preserving local structures within the data, making it excellent for uncovering clusters and patterns that might be hidden in higher dimensions (Bibal et al., 2023). Pairwise similarities in high-dimensional space and Kullback-Leibler divergences can be found in Equation 1.

$$C = \Sigma_i KL(P_i||Q_i) = \Sigma_i \Sigma_j p_{i/j} log_2 \frac{p_{i/j}}{q_{i/j}} \tag{1}$$

In the sample space, P represents the conditional probability distribution, and in the latent space, Qi represents the conditional probability distribution. In t-SNE, entropy is used to construct a cost function (Kullback-Leibler divergence) that measures the difference between the probability distributions in high-dimensional space (P) and low-dimensional space (Q). The proposed algorithm attempts to minimize this difference by effectively preserving the local structure of the data in lower-dimensional space. The units can be described as "bits per data point" or "bits per pairwise similarity." Pairwise similarities in low-dimensional space is shown in Equation 2.

$$Perp(P_i) = 2^{H(P_i)} \tag{2}$$

$$H(x) = -\Sigma(p_i) * log(p_i) \tag{3}$$

Here, H introduces the Shannon entropy of the calculated P in Equation 3. t-SNE's nonlinear approach allows it to capture complex relationships in the data, often resulting in more intuitive and interpretable visualizations than linear methods. However, t-SNEs focus on preserving local structures, which means that they may not always maintain the global structure or distances between widely separated clusters. The output of the dataset after applying t-SNE is shown in Figure 1.

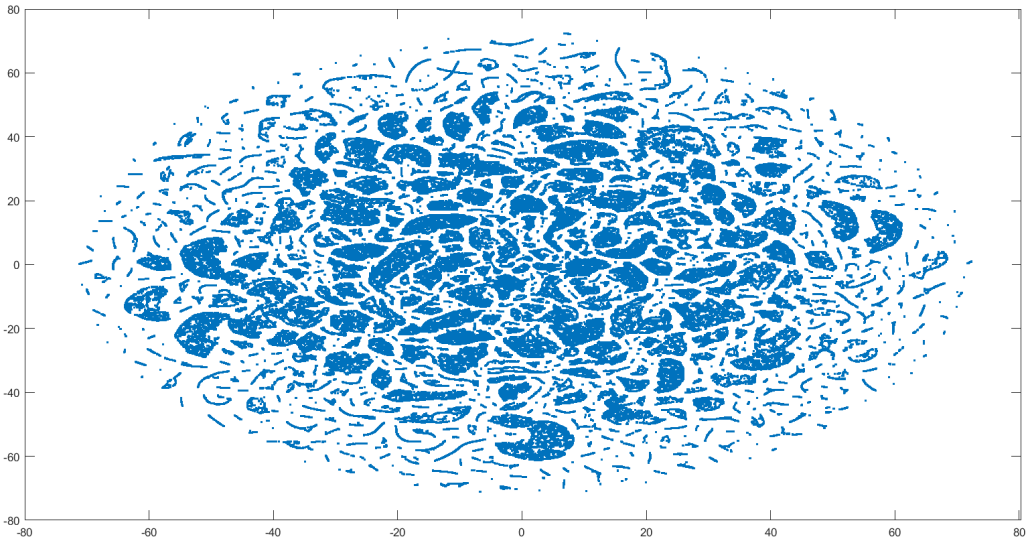


Figure 1. Outlook of Dataset after t-SNE

After all steps were applied to the MATLAB application, the dataset obtained and the code containing the applied methods were shared publicly via GitHub. The files can be accessed via the references ('GitHub - judger90/phishing_detection _tsne', n. d.).

4. RESULTS AND DISCUSSIONS

After removing the labels and text from the dataset, 51 features containing numerical values are obtained. In addition, 51 features determined by applying the t-SNE method converted into 2 features by feature extraction. After this stage,

the original dataset and the dataset to which t-SNE feature extraction was applied are evaluated separately with the help of machine learning methods, and the results shown in Table 1 are obtained.

Table 1. *Classification Accuracy of the Proposed Methods*

Machine learning	**Original Dataset**	**Feature Extracted Dataset**
Logistic Regression	% 42,1	% 57,2
Naïve Bayes	% 99,4	% 77,2
kNN	% 99,6	**% 99,2**
Decision Tree	**% 99,7**	% 83,0
Random Forest	% 57,2	% 81,0

As seen in the table, the best results were obtained using the kNN algorithm for the t-SNE feature extraction method and the decision tree algorithm for the original dataset. Other metrics, Precision, Recall, F1-Score and AUC for the best methods are shown in Table 2.

Table 2. *Performance Metrics of the Proposed Methods*

Machine learning	**kNN: Original Dataset**	**kNN–Feature Extracted Dataset**	**Decision Tree: Original Dataset**	**Decision Tree–Feature Extracted Dataset**
Accuracy	% 99,6	% 99,2	% 99,7	% 83,0
Precision	% 99,6	% 99,4	% 99,7	% 83,0
Recall	% 99,6	% 99,3	% 99,2	% 82,9
F1-Score	% 99,5	% 99,2	% 98,6	% 82,8
AUC	% 99,9	% 99,9	% 99,6	% 86,1

Accuracy measures the overall model accuracy. All methods performed very well. Precision indicates the proportion of positive identifications that were correct. Results closely mirror the accuracy, with kNN and Decision Tree on Original Dataset performing the best (99.6% and 99.7%). Recall represents the proportion of actual positives that were identified correctly. kNN methods on both datasets demonstrated high recall (99.6% and 99.3%). The F1-score is the harmonic mean of precision and recall, providing a single score that balances both metrics. kNN methods on both datasets maintained high F1-scores (99.5% and 99.2%). AUC (Area Under the ROC Curve) measures the model's ability to distinguish between classes. kNN methods on both datasets demonstrated near-perfect AUC (99.9%) and Decision Tree on Original Dataset is also excellent (99.6%).

Generally, the proposed kNN performed consistently well across both datasets. The decision tree method demonstrated a significant drop in performance when used with the Feature Extracted Dataset, that essential information may have been lost during the feature extraction process. The high AUC scores for all methods indicate excellent ability to distinguish between classes even when the other metrics are slightly lower.

Table 3. *Classification Accuracy of the Proposed Methods*

Machine learning	Prediction Speed (predict/second)	Training Time (seconds)	Model Size (bytes)	Selected Features
Logistic Regression: Original Dataset	486172,250	29,049	18217	51/51
Logistic Regression: Feature Extracted Dataset	2192167,345	40,134	11028	2/2
Naïve Bayes: Original Dataset	668,875	1802,257	5019572	51/51
Naïve Bayes– Feature Extracted Dataset	230,602	4335,852	15103075	2/2
kNN: Original Dataset	145,829	6873,399	101873948	51/51
kNN–Feature Extracted Dataset	83355,533	43,985	14090229	2/2
Decision Tree: Original Dataset	584450,559	33,817	10839	51/51
Decision Tree– Feature Extracted Dataset	2293163,544	9,204	8002	2/2
Random Forest: Original Dataset	238485,666	6953,444	261049	51/51
Random Forest– Feature Extracted Dataset	107519,752	356,531	9604	2/2

The prediction speed is shown in Table 3. The prediction speed indicates predictions can be made per second. Therefore, higher values indicate more efficient system designs. The feature-extracted dataset generally exhibited much higher prediction speeds than the original dataset. The logistic regression and decision tree models on feature extracted datasets demonstrated the highest prediction speeds (over two million predictions/second). kNN on the original dataset was the slowest (145,829 predictions/second).

The training time metric expresses the total time required for training in seconds. The feature extracted dataset often (but not always) results in faster training times. The decision tree on the feature extracted dataset had the fastest training time (9.204 seconds). Naïve Bayes and kNN methods on the original dataset had notably longer training times.

The model size metric expresses the total space held by the model in memory in bytes. Feature-extracted datasets generally result in smaller model sizes, with some exceptions. The logistic regression and decision tree models are consistently small. kNN on the original dataset had the largest model size (101,873,948 bytes).

The selected feature column shows the number of features used in the training process. All Original dataset models use all 51 available features, while feature-extracted dataset models use only 2 features, indicating significant dimensionality reduction.

Feature extraction significantly improves efficiency across most metrics:

1. the proposed t-SNE algorithm drastically increases the prediction speed.
2. t-SNE algorithm generally reduces the training time (except for Naïve Bayes).
3. t-SNE algorithms usually decrease the model size (except for Naïve Bayes).
4. t-SNE algorithm achieves high efficiency with only 2 features instead of 51 features.

The feature extraction process is highly effective in this case, as it captures the most valuable information in only two features.

The ROC curves obtained by these algorithms are shown in Figure 2.

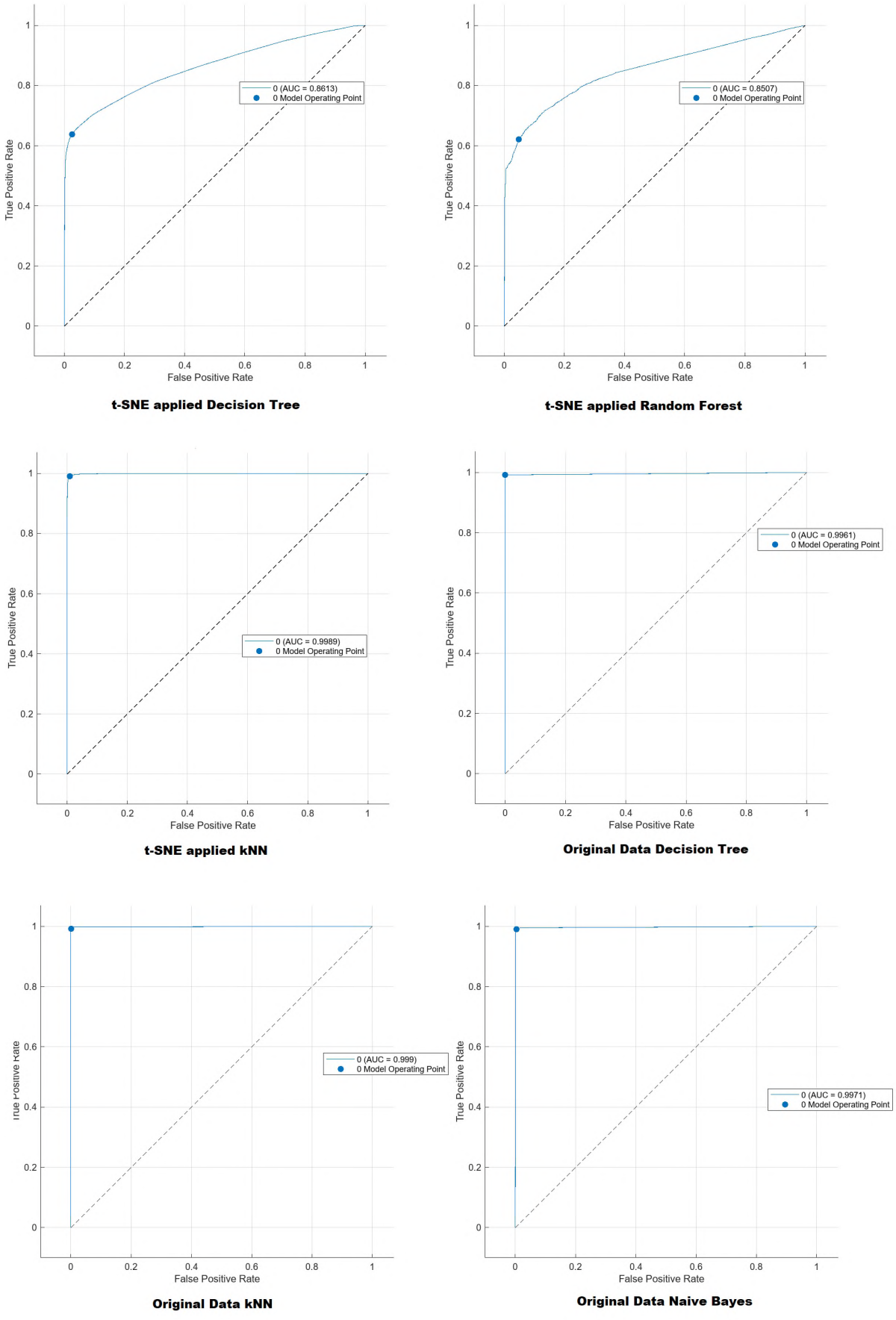


Figure 2. ROC Curves of the Machine Learning Algorithms

In general, the best result obtained by the decision tree in the original dataset. However, in the feature extraction method, the kNN algorithm again demonstrated high performance 99.2%. Another remarkable result was obtained with feature extraction (Table 1, where the success of the Logistic Regression classifier and random forest algorithms significantly increased. The fact that the algorithms use 51 features in the original dataset and only 2 features in the feature extraction dataset indicates a reduction in the amount of memory and training time.

5. CONCLUSIONS

The increasing number of phishing attacks each day supports the continuous conduct of up-to-date studies. In this study, it is aimed to detect websites containing phishing attacks using the PhiUSIIL dataset, which is an up-to-date phishing detection dataset. The dataset has demonstrated high performance in machine learning algorithms both in its original form and with feature extraction using the t-SNE method. The most important advantage of applying inference with the t-SNE method is that the method itself has a lightweight structure, operates fast, and offers high performance in machine learning algorithms with the help of only 2 features obtained. With these characteristics, the proposed method can be used successfully to detect phishing attacks and can also be applied to structures with low system requirements.

Peer Review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- T.E.; Data Acquisition- T.E.; Data Analysis/Interpretation- T.E., M.T.; Drafting Manuscript- T.E.; Critical Revision of Manuscript- M.T.; Final Approval and Accountability- T.E., M.T.; Technical or Material Support – M.T.; Supervision- M.T.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

ORCID IDs of the authors

Taha Etem 0000-0003-1419-5008
Mustafa Teke 0000-0002-7262-4918

REFERENCES

Aburrous, M., Hossain, M. A., Dahal, K., & Thabtah, F. (2010). Intelligent phishing detection system for e-banking using fuzzy data mining. *Expert Systems with Applications, 37*(12), 7913–7921. doi:10.1016/J.ESWA.2010.04.044

Adebowale, M. A., Lwin, K. T., & Hossain, M. A. (2019). Deep learning with convolutional neural network and long short-term memory for phishing detection. *2019 13th International Conference on Software, Knowledge, Information Management and Applications, SKIMA 2019*. doi:10.1109/SKIMA47702.2019.8982427

Alam, M. N., Sarma, D., Lima, F. F., Saha, I., Ulfath, R. E., & Hossain, S. (2020). Phishing Attacks Detection using Machine Learning Approach. *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 1173–1179. doi:10.1109/ICSSIT48917.2020.9214225

Alhudhaif, A., Almaslukh, B., Aseeri, A. O., Guler, O., & Polat, K. (2023). A novel nonlinear automated multi-class skin lesion detection system using soft-attention based convolutional neural networks. *Chaos, Solitons & Fractals, 170*, 113409. doi:10.1016/J.CHAOS.2023.113409

Alsaç, A., Yenisey, M. M., Ganiz, M., Dagtekin, M., & Ulusinan, T. (2023). The Efficiency of Regularization Method on Model Success in Issue Type Prediction Problem. *Acta Infologica, 7*(2), 360–383. doi:10.26650/ACIN.1394019

Atawneh, S., & Aljehani, H. (2023). Phishing Email Detection Model Using Deep Learning. *Electronics 2023, Vol. 12*, Page 4261, 12(20), 4261. doi:10.3390/ELECTRONICS12204261

Bergholz, A., De Beer, J., Glahn, S., Moens, M. F., Paaß, G., & Strobel, S. (2010). New filtering approaches for phishing email. *Journal of Computer Security, 18*(1), 7–35. doi:10.3233/JCS-2010-0371

Bibal, A., Delchevalerie, V., & Frénay, B. (2023). DT-SNE: t-SNE discrete visualizations as decision tree structures. *Neurocomputing, 529*, 101–112. doi:10.1016/J.NEUCOM.2023.01.073

Bibi, H., Shah, S. R., Baig, M. M., Sharif, M. I., Mehmood, M., Akhtar, Z., & Siddique, K. (2024). Phishing Website Detection Using Improved Multilayered Convolutional Neural Networks. *Journal of Computer Science, 20*(9), 1069–1079. doi:10.3844/JCSSP.2024.1069.1079

Buyrukoğlu, S., & Savaş, S. (2023). Stacked-Based Ensemble Machine Learning Model for Positioning Footballer. *Arabian Journal for Science and Engineering, 48*(2), 1371–1383. doi:10.1007/s13369-022-06857-8

Divakaran, D. M., & Oest, A. (2022). Phishing Detection Leveraging Machine Learning and Deep Learning: A Review. *IEEE Security and Privacy, 20*(5), 86–95. doi:10.1109/MSEC.2022.3175225

Doğruel, M., & Soner Kara, S. (2023). Determining the Happiness Class of Countries with Tree-Based Algorithms in Machine Learning. *Acta Infologica, 7*(2), 0–0. doi:10.26650/ACIN.1251650

Efeoğlu, E. (2022). Kablosuz Sinyal Gücünü Kullanarak İç Mekan Kullanıcı Lokalizasyonu için Karar Ağacı Algoritmalarının Karşılaştırılması. *Acta Infologica, 0*(0), 0–0. doi:10.26650/ACIN.1076352

Etem, T., & Teke, M. (2024). Enhanced deep learning based decision support system for kidney tumour detection. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations, 4*(2), 100174. doi:10.1016/J.TBENCH.2024.100174

Garera, S., Provos, N., Chew, M., & Rubin, A. D. (2007). A framework for detection and measurement of phishing attacks. *WORM'07 - Proceedings of the 2007 ACM Workshop on Recurring Malcode*, 1–8. doi:10.1145/1314389.1314391

GitHub - judger90/phishing_detection_tsne. (n.d.). Retrieved 19 September 2024, from https://github.com/judger90/phishing_detection_tsne

Gopali, S., Namin, A. S., Abri, F., & Jones, K. S. (2024). *The Performance of Sequential Deep Learning Models in Detecting Phishing Websites Using Contextual Features of URLs. In SAC '24: Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing* (pp. 1064–1066). Association for Computing Machinery (ACM). doi:10.1145/3605098.3636164

Güler, O., & Yücedağ, İ. (2022). Hand Gesture Recognition from 2D Images by Using Convolutional Capsule Neural Networks. *Arabian Journal for Science and Engineering, 47*(2), 1211–1225. doi:10.1007/S13369-021-05867-2/TABLES/8

Jain, A. K., & Gupta, B. B. (2022). A survey of phishing attack techniques, defence mechanisms and open research challenges. *Enterprise Information Systems, 16*(4), 527–565. doi:10.1080/17517575.2021.1896786

Jiang, D., Shi, X., Liang, Y., & Liu, H. (2024). Feature extraction technique based on Shapley value method and improved mRMR algorithm. *Measurement, 237*, 115190. doi:10.1016/J.MEASUREMENT.2024.115190

Jishnu, K. S., & Arthi, B. (2023). Review of the effectiveness of machine learning based phishing prevention systems. *AIP Conference Proceedings, 2917*(1). doi:10.1063/5.0175593/2919402

Prasad, A., & Chandra, S. (2024). PhiUSIIL: A diverse security profile empowered phishing URL detection framework based on similarity index and incremental learning. *Computers & Security, 136*, 103545. doi:10.1016/J.COSE.2023.103545

Thakur, K., Ali, M. L., Obaidat, M. A., & Kamruzzaman, A. (2023). A Systematic Review on Deep-Learning-Based Phishing Email Detection. *Electronics 2023*, Vol. 12, Page 4545, 12(21), 4545. doi:10.3390/ELECTRONICS12214545

Tülay, E. (2023). Detection of Orienting Response to Novel Sounds in Healthy Elderly Subjects: A Machine Learning Approach Using EEG Features. *Acta Infologica, 0*(0), 0–0. doi:10.26650/ACIN.1234106

Türk, F., Lüy, M., & Barışçı, N. (2020). Kidney and Renal Tumor Segmentation Using a Hybrid V-Net-Based Model. *Mathematics 2020, Vol. 8, Page 1772, 8*(10), 1772. doi:10.3390/MATH8101772

Yaman, O., & Tuncer, T. (2023). Plant Classification Method Using Histogram and Machine Learning for Smart Agriculture Applications. *Acta Infologica, 0*(0), 0–0. doi:10.26650/ACIN.1070261

Yang, L., Zhang, J., Wang, X., Li, Z., Li, Z., & He, Y. (2021). An improved ELM-based and data preprocessing integrated approach for phishing detection considering comprehensive features. *Expert Systems with Applications, 165*, 113863. doi:10.1016/J.ESWA.2020.113863

How cite this article

Etem, T., & Teke, M. (2024). Advanced Phishing Detection: Leveraging t-SNE Feature Extraction and Machine Learning on a Comprehensive URL Dataset. *Acta Infologica, 8*(2), 213-221. https://doi.org/10.26650/acin.1521835

İSTANBUL UNIVERSITY PRESS

Acta Infologica
acin 2024, 8 (2), 222–244
DOI: 10.26650/acin.1540603

Research Article

A Qualitative Study on Management Information Systems Faculty Members' Perceptions of Accreditation in Türkiye

Esra Cengiz Tırpan[1,2], Erman Coşkun[3], Nihal Sütütemiz[1]

[1]Sakarya University, Faculty of Business Administration, Department of Management Information Systems, Sakarya, Türkiye
[2]Bilecik Şeyh Edebali University, Faculty of Economics and Administrative Sciences, Department of Management Information Systems, Bilecik, Türkiye
[3]İzmir Bakırçay University, Faculty of Economics and Administrative Sciences, Department of Management Information Systems, İzmir, Türkiye

Corresponding author : Esra CENGİZ TIRPAN
E-mail : esracengiz01@gmail.com; esra.cengiz6@ogr.sakarya.edu.tr

*This study is partially produced from the first authors PhD dissertation, prepared under the supervision of the second and the third authors.

ABSTRACT

In recent years, the accreditation of universities and programs has gained special importance worldwide and in our country, and higher education institutions have set various accreditation targets. Accreditation, which documents the compliance and quality of an institution or program with specific standards, includes an audit process that evaluates and approves the compliance of the institution or program with national or international standards. Although a successful accreditation process depends on many factors, one of the obstacles is the lack of ownership and support by faculty members and even resistance to accreditation. The fact that accreditation processes are new in Management Information Systems (MIS) departments in Türkiye and that no MIS department with international program accreditation has been evaluated using program-specific criteria reveals the importance of measuring MIS faculty members' perceptions toward accreditation. The primary purpose of this study is to determine the current status evaluations of MIS bachelor programs in Türkiye, examine the perceptions and perspectives of faculty members regarding accreditation, and reveal the problems encountered in the process. Within the scope of the research, in-depth interviews were conducted with 11 faculty members working in the MIS department in Türkiye using a semi-structured interview form. The questions in the interview form were divided into three categories: perception and perspective on accreditation, Accreditation of MIS programs, problematic areas encountered during the accreditation process, and the benefits of accreditation. Comparisons between participants were presented, and a descriptive analysis of the categories was performed. The findings revealed that faculty members have a positive perspective on the accreditation process regarding quality education, continuous improvement, and process management. However, documentation, monitoring of processes, and the associated workload are seen as difficulties in accreditation. Therefore, university administrators must make necessary arrangements to handle paperwork and routine data collection. Staff members and faculty members should be able to focus on their primary activities, such as increasing the quality of teaching and producing high-impact research.

Keywords: Accreditation, higher education, management information systems, qualitative analysis

Submitted : 29.08.2024
Revision Requested : 22.10.2024
Last Revision Received : 06.11.2024
Accepted : 16.12.2024
Published Online : 23.12.2024

1. INTRODUCTION

The quality of the higher education system significantly impacts a society's economic and social welfare. Countries with weak higher education systems have an economy that is intensive to agriculture and physical production and a relatively low level of welfare. In contrast, countries with strong higher education have industries and service sectors based on data and information (Prendergast et al., 2001). The purpose of higher education has radically changed over the past 30 years because of globalization, technological advancements, economic dynamics, shifts in the business landscape, and various other factors. Beyond scientific studies and producing science, universities must graduate individuals that companies, the public, and society need based on market dynamics and expectations and offer them to the market. For most universities, the luxury of doing science for its sake has disappeared, and the necessity of doing science and applying it to society and companies has emerged. Considering these developments, accreditation can be defined as the determination of professional education standards and the evaluation of compliance with these standards. Thanks to accreditation, it is possible to assure the public and relevant administrations that accredited program graduates have the education and skills to enable them to work in the appropriate field. Accreditation, in its simplest definition, involves reviewing the quality of higher education institutions and programs (CHEA, 2024). Many accredited programs benefit from a well-designed evaluation and assessment process that collects meaningful data. These data offer insights into the achievement of program objectives and guide decisions on improvement. (Oudshoorn et al., 2018).

Accreditation has also been gaining importance in our country, and the increasing competition in higher education has made accreditation a necessity. In July 2024, the Council of Higher Education (CoHE) ended universities' evening education programs. The rapidly increasing number of universities in the last 30 years has melted the accumulated student potential due to the increase in educational opportunities abroad, and quality, not quantity, has now come to the fore for Turkish higher education. Due to the labor market's supply-demand imbalance and business demands, training qualified graduates is essential for universities' survival and continued operation. Accredited universities are approaching this goal and are becoming an elite group. For this reason, institutions and organizations that manage this issue, primarily the umbrella organization Turkish Higher Education Quality Council (THEQC), have been established and continue their work. The authorization and recognition of accreditation agencies in Türkiye is the responsibility of THEQC. MÜDEK, Türkiye's first national accreditation agency in higher education, was established in 2002 under the "Engineering Evaluation Board" name to evaluate engineering undergraduate programs (Tantekin Ersolmaz, 2018). Today, 24 accreditation bodies carry out evaluation processes in different fields, are registered by THEQC, and whose registration period continues (YÖKAK, 2024). In addition, universities and programs pursuing internationalization have carried out activities and continue to receive international accreditation.

Many elements and factors contribute to the success of accreditation. The main areas of focus are faculty members, students, senior management support, processes, curriculum, physical spaces, and financial resources. An institution that wants to be accredited should be able to demonstrate that it is ready to a certain extent for all these factors. In this context, this study aims to conduct a general accreditation readiness assessment for MIS Departments, whose numbers have increased rapidly in Türkiye in the last 10 years but have never received program accreditation at the international level, and to examine the perceptions and perspectives of faculty members, who are the most important and unfortunately the most problematic factors for accreditation. This study also aims to determine the accreditation criteria and current status evaluations for MIS programs and to reveal the problems encountered along with the benefits of accreditation. The research has the following sub-objectives within the scope of this aim;

1. General determination of the accreditation status of MIS departments in Türkiye
2. Accreditation options for MIS departments in Türkiye
3. Prominent elements, obstacles, and supporting elements in the process
4. Determination of faculty members' perceptions and perspectives on accreditation
5. Determination of faculty members' MIS program accreditation evaluations
6. Evaluation of the current status of the program in line with the ABET accreditation criteria
7. Discussion of problematic areas regarding accreditation and the benefits that accreditation offers

2. QUALITY AND ACCREDITATION IN HIGHER EDUCATION

Higher education institutions are vital for equipping students with the knowledge and skills necessary for a sustainable future. (Lozano et al., 2013). Although many higher education institutions contribute positively to sustainable development, a more profound and comprehensive transformation is necessary (Caeiro et al., 2020). Universities face ongoing challenges from rapid advancements in information and communication technologies, the growing diversity

of students, and the shift toward a knowledge society (Al Shobaki and Naser, 2017). Along with these changing challenges, stakeholders also demand better university results regarding research, teaching, employability, and social access (Frasquet et al., 2012; Angulo-Ruiz, 2016). Students prefer high-quality and qualified educational services to help them develop their careers (Mpinganjira, 2011). In today's competitive and innovative educational landscape, universities must focus more on meeting stakeholders' expectations. However, the globalization of professions and the mobility of professionals create a need for institutions to provide qualifications recognized in the international labor market and enable the comparability of academic standards (Hernes and Martin, 2008). The expansion, privatization, and globalization of higher education are increasing the need to ensure quality (Kumar et al., 2020). Quality assessment has long been a focus for higher education institutions and is increasingly crucial to their success (Van Vught and Westerheijden, 1994; Mursidi et al., 2020). The pressure to improve academic programs and increase graduate employability leads to a greater emphasis on quality in the higher education sector (Hijazi, 2016).

Quality in higher education is a multifaceted concept that is examined in various dimensions. According to Harvey and Williams (2010), quality in educational programs is conceptualized in different ways, such as meeting specified standards, being fit for purpose, or being transformative. Welzant et al. (2011) suggested four broad conceptualizations of quality in higher education (purposeful, transformative, exceptional, and quality as accountability) and several quality indicators to evaluate each of these broad conceptualizations. While the definitions in the literature focus primarily on processes, policies, or actions, they emphasize quality aspects related to accountability and continuous improvement (Welzant et al. (2011). Harvey and Knight (1996, p.97) presented the ultimate quality goal that higher education should pursue: "the need to invest in continuous improvement of the quality of the student experience through staff development, innovation in teaching and learning, research and scholarship."

Accreditation is a widely used mechanism to assess and ensure the quality of higher education institutions (Harvey, 2004; Sanyal and Martin, 2007; Eaton, 2015; Blanco-Ramirez, 2015). Accreditation is based on an external evaluation process to check the achievement of specific quality standards (Van Berkel and Wijnen, 2010). Accreditation in higher education can be considered the formal verification of quality assurance compliance by control bodies (Lundquist, 1997). Accreditation affects many aspects of academic reputation and internationalization, research and innovation, teaching quality, and employability (Kumar et al., 2020).

2.1. Accreditation and Management Information Systems Departments in Türkiye

Considering the field of informatics with the first working computer in 1947 and the first commercially sold computers in the 1950s, the emergence of Information Systems as a discipline for the application of information processing technologies to businesses and competitive advantage did not exist until the late 1960s (Ceccucci and White, 2008). MIS is a system that manages data, processes, and people in organizations. It mainly contributes to the managerial aspects of the organization (King 1978). As a discipline, MIS deals with computer science, management science, operations research, psychology, sociology, and economics research to develop information systems and produce solutions to real business problems (Laudon and Laudon 2014). The foundation of MIS is organization, management, and technology (Kroenke 2012; Laudon and Laudon 2014). Management information systems, which was included in the curriculum as a course in Türkiye, was later included as a department by the decision of the CoHE. Following the department's opening under the name of "Business Informatics" at Marmara University in 1991, Boğaziçi University established the first department under the name of MIS in 1995 (Coşkun and Alma, 2018). While education in this field is gathered under a standard name, Business Informatics, in German-speaking countries, there is no single standard name in English-speaking countries. In parallel with the different intensities of subjects related to information and communication technologies and business administration, it is seen that the teaching programs are generally named "Business Information

Technology," "Computer Information Technology," "Business Computing," "Business Informatics," "Information Systems," "Management of Information Systems," "Management of Business Systems," and these branches are divided into sub-branches within themselves (Akpınar, 2011). At the Higher Education Executive Board meeting dated 11.03.2020, the names and contents of undergraduate departments/programs with the same content but different names in higher education institutions were reviewed and simplified. As a result of this study, the departments previously named "Business Information Management," "Business Informatics," "Technology and Information Management," and "Management Science and Information Systems" were gathered under the name of "Management Information Systems" to be implemented as of the 2020 Higher Education Institutions Exam (YKS). MIS is an interdisciplinary field of study that exists at the intersection of many other disciplines, such as business, industrial engineering, computer engineering, informatics, mathematics, and statistics (Bensghir, 2002).

In higher education, accreditation is usually given at the institution and college levels, while some accrediting

organizations provide accreditation at the level of individual schools/programs (Hilton, 2003). MIS departments can be established within different faculties and schools. In the world, Information Systems (IS) programs in computing and engineering schools usually apply for ABET accreditation. In contrast, the most common accreditations for business schools are AACSB and EQUIS (Leidig and Salmela, 2021). Founded in 1932 as the Engineers' Council for Professional Development (ECPD), ABET is an engineering organization focused on education, accreditation, regulation, and professional development for engineering professionals and students in the United States. It changed its name to ABET in 1980. In 2001, ABET established the Computing Accreditation Commission with the assistance of the Computer Science Accreditation Board to accredit Computing Programs (ABET, 2024). Almost every engineering program in the United States is accredited by ABET (MacKinnon et al., 2012). The long-standing indecision of AACSB regarding IS content has motivated the creation of specific BS accreditation standards by ABET, the agency responsible for accrediting computer science and engineering programs (Hilton et al., 2004). ABET is an organization that accredits programs in computing, engineering, and engineering technology, as well as applied and natural sciences. The ABET Commission on Computing Accreditation (CAC) accredits computer science, information systems, information technology, cybersecurity, and other computing programs, primarily in the United States and more than 20 countries worldwide. For a program to be accredited by ABET, it must meet ABET criteria, which focus on what students learn, whether the curriculum, faculty, and facilities are appropriate, and whether the program meets quality standards to produce graduates who will enter and succeed in the global workforce as educated professionals (Oudshoorn et al., 2018).

2.2. Accreditation Status of Management Information Systems Departments in Türkiye

The idea of accreditation for MIS programs in Türkiye is still very new. Unfortunately, there is no MIS program accredited by an international accreditation agency that accredits information systems programs in Türkiye, and it is seen that the general awareness level of MIS on this issue is very low. However, as of October 31, 2022, Sakarya University Faculty of Business Administration and as of February 21, 2023, Özyeğin University Faculty of Business Administration received AACSB accreditation (Sakarya University, 2024; Özyeğin University, 2024). AACSB was established in 1916 as the "American Assembly of Collegiate Schools of Business" to provide accreditation to business schools in North America. It began its work by accepting the first standards in 1919 (AACSB, 2023). AACSB business accreditation is provided as institutional accreditation or business unit accreditation. In institutional accreditation, all business degrees are included in the scope of the AACSB accreditation review unless excluded (AACSB, 2023, p.12). According to current AACSB data, more than 1,900 member institutions and more than 1,000 accredited schools are located in more than 100 countries. Since MIS programs are also available within the scope of the Sakarya University Faculty of Business Administration and Özyeğin University Faculty of Business Administration, these MIS programs also have AACSB accreditation. Another international accreditation agency is the AQAS. The organization performs accreditation processes for program accreditation and system accreditation in all disciplines provided by higher education institutions (Zayachuk and Yamelynets, 2021). The basic idea of program accreditation by AQAS is to assess whether the program is conducted by the "European Standards and Guidelines for Quality Assurance in the European Higher Education Area" (ESG) while ensuring that the curriculum meets current academic standards. However, it has developed its criteria and indicators based on ESG, which allows focus on relevant indicators of teaching and learning processes (AQAS, 2024). The first international institution to become a member of AQAS is Istanbul Gelişim University (IGU) from Türkiye. In 2017, some IGU programs were accredited based on ESG. The IGU MIS undergraduate program was also accredited in 2018 within the framework of the criteria until September 2024 (Istanbul Gelişim University, 2024). When national accreditation is examined, the "Social, Humanities and Basic Sciences Accreditation and Rating Association" (STAR) was established in 2020 to present the evaluation and accreditation activities of "Business Administration, International Trade, Management Information Systems, Economics, Finance, Banking, Labor Economics, Public Administration, Political Science, and International Relations Undergraduate Programs in the Republic of Türkiye and the Turkish Republic of Northern Cyprus" (STAR, 2024). The STAR accreditation agency, authorized by THEQC, has accredited more than 70 undergraduate programs in 26 different institutions for 2 years, 3 years, 4 years, or 5 years. Atatürk University, Başkent University, Istanbul Beykent University, and Karadeniz Technical University MIS undergraduate programs have STAR accreditation on a program basis.

2.3. Faculty Members' Perceptions of Accreditation

Faculty members play an important role in the accreditation process to ensure academic integrity and self-regulation (Alstete, 2004). The accreditation process enables faculty to review program data and design while advocating for

effective practices (Gilbert, 2010). Alstete (2004) noted that most faculty members entering higher education do not anticipate participating in activities such as planning, organizing, or managing accreditation. This creates a conflict between faculty expectations and accreditation requirements. Faculty members fulfill various roles, such as serving on committees for curriculum development and evaluation, which are directly tied to the accreditation process (Rhodes, 2012). Accrediting organizations want the curriculum to be structured to have a lasting impact on students (Rhodes, 2012). Faculty members must create meaningful learning experiences while fulfilling accreditation requirements in the curriculum (Lewis, 2016). Many studies are in the literature to determine faculty members' perceptions of accreditation. Ceccucci and White (2008) conducted a study to determine the attitudes and interests of IS faculty members regarding ABET accreditation. They concluded that ABET accreditation is slowly gaining popularity, and the most important reason for targeting it is the perceived quality that it brings to a program. Wilson-Hail et al. (2019) aimed to examine faculty members' perceptions of the CAEP process when accrediting teacher education programs. According to the research results, faculty members believe that the national accreditation process is important for status and prestige, but they question whether it helps with the necessary systematic changes. In addition, faculty members perceive accreditation as a disadvantage when accreditation-related work is not included in the workload or hours. Hoare and Goad (2022) aimed to determine the factors that positively and negatively affect quality culture in accredited institutions in North America. Focus group interviews were conducted with 36 academic staff members, and a survey was conducted with 198 participants. According to the research results, there was a noticeable difference between those invited to participate in accreditation processes and those who contributed to the decision-making process. This lack of inclusive governance prevents institutions from responding appropriately to the needs of their communities. Bougherira and Elasmar (2023) investigated the impact of program accreditation conducted by the Saudi National Center for Academic Accreditation and Evaluation (NCAAA) on teaching and learning. Within the scope of the research, a survey was conducted with 469 faculty members, and interviews were conducted with 10 faculty members in 5 accredited programs. The research findings show that accreditation has a significant impact on teaching and learning. The greatest effect was observed in the planning and design of learning outcomesIto et al. (2024) conducted a qualitative study to examine the impact of accreditation on the quality of education—in-depth interviews conducted with faculty members of an accredited business school in Japan.The findings revealed that the participants agreed that accreditation positively affected teaching and encouraged active learning and the case method. However, they also stated that accreditation had a restrictive effect on assessment activities, promoting compliance rather than assessment of real learning. Mady et al. (2023) investigated the attitudes of faculty members of the Kuwait University College of Business Administration toward the AACSB accreditation system and how they perceived its benefits. The most positive attitudes in the research findings were that accreditation is an esteemed stamp of excellence and an indicator of a high-quality business program. Shaiban (2024) studied the effects of NCAAA program accreditation on perceived stress among dental college staff and compared levels before and after the process. The findings indicated that 44.67% of participants experienced high stress before NCAAA, while 31.33% reported low stress both before and after the accreditation. Their qualitative research conducted with faculty members working in education faculties in Türkiye, Alpaydın, and Topal (2022) revealed three main themes: "problems encountered by faculty members in the accreditation process," "positive aspects of the process," and "suggestions regarding the process." The highest frequencies in the codes formed under the themes are seen in the codes "extra paperwork and workload problem," "positive contribution to educational content," and "suggestion for communication with experienced institutions." Semerci et al. (2021) surveyed 181 faculty members in their screening model study to measure their perceptions of accreditation. The research results revealed that faculty members' perceptions were high. Still, there were some obstacles, such as "inappropriate standards," "inadequacy of accreditation boards," and "institutional deficiencies" in the higher education accreditation process in Türkiye.

3. METHOD

In this study, after determining the accreditation criteria and current status evaluations for MIS programs in Türkiye, the case study method, one of the qualitative research methods, was preferred to examine the perceptions and perspectives of faculty members regarding accreditation and to reveal the problems encountered in the process. Case studies are defined as the in-depth description and examination of a limited system (Merriam, 1998; Yıldırım and Şimşek, 2018). The accreditation process is considered a system that can be limited by specific rules and conditions. In this study, an in-depth examination is aimed at understanding this system to identify its benefits and problems regarding the functioning of the process. Multiple case study (Yin, 2003) are the basis of this study. In this direction, MIS departments at 11 universities in Türkiye were studied, and each department was evaluated in terms of accreditation. The holistic multiple-case design allows each case to be compared with each other (Yıldırım and Şimşek, 2018). In this research, the accreditation status of the departments examined was compared after data were obtained. Various categories and

codes were created from the received data to obtain inferences and explanations about the situation. Thus, within the scope of the research, common points and differences between departments regarding the accreditation process can be seen more clearly.

3.1. Sample Group

The scope of this research consists of foundation and state universities in the education process in MIS departments, which can be called a new department in Türkiye. The research sample was determined using maximum variation sampling, which is a purposeful sampling method. Various characteristics were focused on to provide maximum variation while creating the sample (Patton, 1987). In maximum variation sampling, the sample is determined based on the factors related to the problem and shows diversity, difference, variability, and similarity within themselves (Grix, 2010). Within the scope of the research, characteristics such as whether the MIS departments are in state or foundation universities, whether they are well-established or relatively newly established among the MIS departments in Türkiye, whether the department or the unit where the department is located has carried out accreditation work, its geographical location, and the language of education were determined. In this way, sample selection was used to provide diversity for the general purpose and questions of the research and to understand the situation from different perspectives. Based on the sampling strategy of the study, according to the 2022 ÖSYM data taken into account on the sample determination date, 82 MIS undergraduate programs in 71 different universities with MIS departments out of 208 universities in Türkiye were determined within the scope of the study. 40 of the 82 programs are in foundation universities, and 42 are in state universities. Most universities where the programs are located are in big cities, 33 of which are in Istanbul, 11 in Ankara, and 4 in Izmir. When examined in terms of language of instruction, it is seen that MIS undergraduate education is provided in 3 different languages: Turkish, German, and English. While there is 1 state university providing education in German, 7 of the 25 departments providing MIS undergraduate education in English are state universities, and 18 are foundation universities. MIS departments providing education in state and foundation universities have different backgrounds, are deep-rooted, and are relatively new.

When the accreditation status of the MIS departments is examined, 1 national and 1 international program accredited department is present at the time of sample determination. When accreditation is examined on a unit basis, it is seen that only 1 university has international accreditation. Considering these features within the scope of the research, the maximum diversity of the sample was targeted. According to Cropley (2002), the sample size in qualitative research should be determined by evaluating the research focus, data amount, and theoretical sampling principles. The researcher can decide that they have sufficient data at the point where the concepts and processes that emerge start to repeat (Guest et al., 2006; Yıldırım and Şimşek, 2018). In this study, after working with 11 participants, repetitive data were obtained, and the desire to interview other people was eliminated. The participant information is presented in Table 1.

Table 1. Information about the Departments Conducting the Case Study

Participant	The type of University	Region	Department Establishment History	Education in a Foreign Language	Accreditation Status/Readiness	Interviewee Title, Position
MISI	State	Marmara Region	Between 2011-2019	No	Yes	Prof. Dr., Department Head
MISII	State	Marmara Region	2010 and before this date	Yes	No	Prof. Dr., Dean
MISIII	State	Marmara Region	2010 and before this date	Yes	Yes	Assoc. Prof. Dr. Vice Dean
MISIV	State	Central Anatolia	Between 2011-2019	No	Yes	Assoc. Prof. Dr. Vice President of Department
MISV	State	Black Sea Region	Between 2011-2019	No	Yes	Prof. Dr., Department Head
MISVI	State	Marmara Region	Between 2011-2019	No	No	Assoc. Prof. Dr. Department Head
MISVII	State	Marmara Region	Between 2011-2019	No	Yes	Prof. Dr., Department Head
MISVIII	State	Central Anatolia	Between 2011-2019	Yes	Yes	Assistant Prof., Vice President of Department
MISIX	Foundation	Central Anatolia	2010 and before this date	Yes	Yes	Assoc. Prof. Dr. Department Head
MISX	Foundation	Marmara Region	2020 and beyond	No	Yes	Assistant Prof, Department Head
MISXI	Foundation	Marmara Region	Between 2011-2019	Yes	Yes	Research Assistant., Quality Commission Member

3.2. Data Collection Tool

The interview form used in this study was adapted from that developed by Coşkun et al. (2020). Before implementing the prepared interview form, a pilot study was conducted with 2 faculty members. The data obtained from the pilot interviews were examined, and the interview form was finalized. Before proceeding with the data collection process, ethics committee approval was obtained from the "Sakarya University Social and Human Sciences Ethics Committee."

3.3. Data Collection

After determining the universities where the case study would be conducted, interview requests were created with the determined individuals. The individuals who would participate in the study were contacted via their institutional email addresses. An explanation was given about the purpose and content of the study, and an interview request was made. Interviews with individuals who wanted to participate in the study were planned to be held in their own institutional offices on a specified day and time. While the interviews were conducted face-to-face with 8 participants who agreed to participate in the study, 4 participants were interviewed via the Google Meet video-conferencing application. Before the interview, the researcher briefly introduced himself and provided brief information about the study's purpose, scope, and content. Prior to the interviews, permission for audio recording was requested, and the interviews were recorded with the consent of 11 participants, except for 1 participant. Interviews for which permission for audio recording could not be obtained were not included in the analyses to not affect the validity of the research.

The average interview duration was 1.5 hours (88 minutes); the most extended interview was completed in 2 hours and 25 minutes, and the shortest interview was completed in 42 minutes. The interviews were held over approximately 3 months, starting in the Spring semester of March 2022-2023.

3.4. Data Analysis

In the rationalist paradigm, the terms typically used to determine the degree to which a study's findings are solely a function of the respondents and the conditions of the study rather than the researcher's biases, motivations, interests, perspectives, etc., are internal validity, external validity, reliability, and objectivity, respectively. (Guba and Lincoln, 1982). Guba and Lincoln (1982) proposed alternative concepts by suggesting that qualitative research should consider trustworthiness rather than validity and reliability. These concepts are grouped under four main headings: credibility, transferability, dependability, and confirmability. Some strategies for applying these criteria have been adopted by many researchers (Cho and Trent, 2006; Herr and Anderson, 2015; Merriam and Tisdell, 2016; Patton, 2015). In a study, one or more of these strategies should be used to check the accuracy of the findings (Creswell, 2003).

Credibility means that the research results match reality (Zohrabi, 2013). Detailed interviews were conducted with the participants to provide credibility criteria for the research. The interviews recorded with a voice recorder were transcribed using transcription software and were given as direct quotes. The obtained voice recordings were archived to be viewed again when necessary. Long-term interaction was provided with the transcribed interviews, and the following steps were followed.

- The audio recordings converted to text via software were listened to once and checked.
- Incorrect conversions in the text file were corrected.
- The text was simplified by deleting interview records that were not within the scope of the research.
- The documents were read 5-6 times to ensure familiarity with the text.

Transferability is the equivalent of the concept of generalization in quantitative research in qualitative research (Yıldırım and Seçkin, 2018). In quantitative research, generalization is achieved by showing that the data are collected from a sample representing the universe (Guba and Lincoln, 1982). There is no generalization purpose for qualitative research. Nevertheless, the research results should be explained in detail so that other researchers who examine the research can apply them in their study (Başkale, 2016). According to Erlandson et al. (1993), detailed descriptions and purposeful sampling methods are recommended to ensure the transferability of the results. A purposeful sampling method was selected for this research, and the results were described in direct quotation.

Guba and Lincoln (1982) suggested the concept of dependability in qualitative research instead of the concept of reliability, which indicates repeatability in quantitative research. Repeatability in qualitative research is impossible because of the constant change in events and phenomena. Accordingly, qualitative research does not pursue reliability as in quantitative research. Instead, qualitative research aims to accept variability and reflect this variability in the study in a consistent manner (Yıldırım and Şimşek, 2018). If the procedures of previous case studies are not well documented, this may cause doubts about the reliability of the study (Aytaçlı, 2012). According to Yin (2003), a study should be

prepared to record the procedures performed and guide the researcher to ensure the repeatability of the research. Within the scope of the study, the following methods were carried out based on the relevant literature to ensure dependability (LeCompte and Goetz, 1982; Miles and Huberman, 1994).

1. The research questions were clearly stated, and the steps followed were explained.
2. The research data were obtained using the interview form and presented with direct quotes.
3. The obtained data's ability to answer the research questions and the consistency of the results were constantly checked.
4. In the content analysis, code lists were created using a deductive approach based on the relevant literature. The researcher obtained a standard code by comparing the codes made at different times.
5. The codes obtained from the second researcher were compared; the codes with differences were examined, and a consensus was reached.
6. The codings were finalized based on the expert opinion.

In scientific research, results are expected to be objective and to avoid the subjective judgments of the researcher. However, complete objectivity is not possible in qualitative research, and the researcher has an influence. In qualitative research, the concept of confirmability, which is used instead of objectivity, is expected from the researcher to verify the results obtained with the data they collect constantly and to be able to provide a logical explanation to the reader (Yıldırım and Şimşek, 2018). To meet the confirmability criterion within the scope of this research, the researcher attempted to reflect the participants' opinions objectively with an objective attitude. The data collection tools used in the study, data collection process, analysis phase, and codebook were described in detail and presented to the experts for review.

The categories and subcategories were created based on the study of Coşkun et al. (2020). The categories and subcategories created in line with the purpose of the research are presented in Table 2.

Table 2. Category and Subcategory List

Category	Subcategory
Perceptions and perspectives on accreditation	The concept of accreditation
	The purpose of accreditation
	Accreditation is a fashion or quality
	The contribution of accreditation to continuous improvement
	Indicators of the best accreditation system
	Turkish Higher Education System
Accreditation of MIS programs	Information about the processes and steps to follow
	MIS accreditation preference criteria
	Information about ABET accreditation
	Assessment in line with ABET criteria (Deficient Criteria)
Problematic Areas and Benefits of Accreditation	Information about the cost of accreditation
	Institutional budget for accreditation
	The idea that accreditation is worth the high cost
	Obstacles to accreditation
	The effect of being a state or foundation university on the process
	Advantages of accreditation
	Difficulties and disadvantages of accreditation

4. RESEARCH FINDINGS

The MAXQDA 2022 package program was used to code the deciphered data and present the codes. The codes created within the scope of the research were counted once per document, and the code matrix scanner Excel output was prepared. The prepared tables are presented in each category.

Participants were asked to evaluate the awareness and knowledge levels of their faculty members, students, and administrators about accreditation and the general situation regarding the accreditation process on a 10-point scale. The median values of the responses of the 11 participants are presented in Table 3.

Participants stated that they have a high level of knowledge about the objective (8), process (9), and criteria (8) about accreditation. This indicates that the accreditation process is generally understood. The belief in the necessity of accreditation is relatively high (9). However, the belief that all accreditation criteria can be successfully implemented is lower (7). This situation reflects concerns about the difficulties of the process and the inadequacy of available resources. The average scores for the adequacy of resources (7) and expert staff (7) also support these concerns. Top management support was evaluated positively (8). This is an important factor for the successful implementation of the accreditation process. Curriculum adequacy (8) is considered to be at a reasonable level. In the questions aimed at assessing awareness

Table 3. Participants Rating

Questions	Median
Knowledge about the objective of accreditation	8
Knowledge about the accreditation process	9
Familiarity with accreditation criteria	8
Belief in all criteria to be realized	7
Belief in the necessity of accreditation	9
Sufficiency of resources for accreditation	7
Sufficiency of expert personnel for accreditation	7
Top management support	8
Curriculum adequacy	8
Independent management level	7
Qualified academics	8
Awareness of faculty members about accreditation	7
Knowledge of faculty members about accreditation	7
Awareness of students about accreditation	6
Students' knowledge about accreditation	5
Self-awareness about accreditation	9
Self-knowledge about accreditation	9
Awareness of managers about accreditation	8
Knowledge of managers about accreditation	8

and knowledge, it was evaluated that students had the lowest level of knowledge (5). The knowledge and awareness of the administrators (8,8) were assessed as higher than those of the faculty members in the department (7,7).

4.1. Descriptive Analysis of Categories

Category 1: Perceptions and Perspectives on Accreditation

Within the scope of the research purpose, prominent codes were determined in line with the questions asked to understand faculty members' perceptions and perspectives on accreditation.

The concept of "standardization" comes to the fore in response to the question of what accreditation means to you. The participant with code MISI interpreted this concept of standardization in terms of the quality process and expressed it as "*a quality process, a process of complying with standards.*" The participant with code MISIII similarly stated that "*this indicates the quality that comes with standardization*." Most participants interpreted what accreditation meant to them using expressions such as "providing standards, complying with standards, determining standards." The distribution of codes related to the purpose of the accreditation category by participants is presented in Table 4.

Table 4. Code Distribution of the Purpose of Accreditation

Code System	Number of Participants (out of 11)	Percentage of Participants
Ensuring standards	6	54.5
Process management	4	36.3
Ensuring quality	4	36.3
Educational studies	3	27.2
Improvement	3	27.2

When the category of the purpose of the accreditation is examined, the prominent codes are "ensuring standards, process management, ensuring quality, educational studies, and improvement." Educational studies include "harmonization of the curriculum, aiming to provide contemporary education." It is thought that there are problems regarding how much accreditation serves these purposes in Türkiye today because "most of the criteria are not fully implemented." Participants believe that accreditation serves its purpose in universities managed by accreditation standards and is beneficial. The dominant idea is that since accreditation is new in the MIS field, determining the standards is more important.

The distribution of codes regarding the perception of whether accreditation is fashionable or not is presented in Table 5. Although none of the participants directly approached the accreditation process with a perception of fashion, the

dominant thought was that it could be evaluated for advertising purposes, but it is still needed. Table 6 was created to understand the comparative situation better.

Table 5. Code Distribution Regarding the Fashion Perception of Accreditation

Code System	Number of Participants (out of 11)	Percentage of Participants
Necessity	8	72.7
Quality	7	63.6
Fashion by perspective	6	54.5
Fashion	0	0.0

3 participants thought that accreditation would be fashionable from their perspective, but at the same time, they stated that it was a necessity. 5 participants also indicated that they believed that accreditation was necessary and that they had a perception of quality.

Table 6. Comparison of Fashion Perceptions About Accreditation

Code System	MISI	MISII	MISIII	MISIV	MISV	MISVI	MISVII	MISVIII	MISIX	MISX	MISXI
Fashion by perspective	√	√	√	√	√		√				
Necessity		√		√	√	√		√	√	√	√
Quality	√		√			√		√	√	√	√
Fashion											

In general, it was emphasized that accreditation is sometimes perceived as a fashion in Türkiye, that long-term efforts and labor are required for accreditation, that the accreditation process is complex, and that this process requires good work and effort. Therefore, the dominant thought is that accreditation is important for an institution, even if it is done for advertising purposes rather than fashion. There is a thought that accreditation is not a temporary fashion. The participant with the code MISI drew attention to the fact that it is important even if it is done for advertising purposes rather than for fashion perception with the statement, "So if they get that accreditation from good places, if it complies with its rules, even if they do it for advertising purposes, it means they are doing something right, they are working well, they are working on it." The participant with code MISIV made a different assessment for state and foundation universities with the statement, "*Actually, we can't think of it as a fashion perception because we are a state university. . . But I think foundation universities have such a concern. In foundation universities, there is both financial expectation and because they are concerned about bringing international students to their institutions." The participant with code MISIX stated, "It is more like a necessity than a fashion, like someone else has it and I don't. Because a student comes. I do the promotion and say we have accreditation. Then they go to another university and say that university X has accreditation. . . , after a while, this will become a necessity." The idea that it has become necessary for universities with student demands is dominant. The participant with code MISV stated, "So it can be called accreditation, or it can be in another way. It can be a certificate or a certificate of competence. I believe in the necessity of these, but as I said at the beginning, it should not be a fashion if these are realized by believing in their necessity and internalizing them.*" This situation draws attention to the implementation process of accreditation.

Table 7 presents the code distribution of participants' perceptions of accreditation's contribution to continuous improvement.

Table 7. Code Distribution of Participants' Perceptions of Accreditation's Contribution to Continuous Improvement

Code System	Number of Participants (out of 11)	Percentage of Participants
Yes	9	81.8
No	2	18.1

While 2 participants stated that accreditation does not contribute to continuous improvement, all others stated that accreditation positively contributes to continuous improvement. The participant with code MISII stated that an institution can remain at a certain level as long as specific accreditation criteria are met. Therefore, it cannot directly contribute to continuous improvement. The participant with code MISVII also thought there is no constant improvement because accredited institutions are not checked every year, and there is no continuous improvement because the criteria can be provided similarly during the re-evaluation process once they are met. The participants' statements with codes MISII and MISVII are as follows.

MISII: "*Well, it doesn't mean it will happen because it wants certain things and certain criteria. Once they meet those criteria, they can stay there. . . So, I am accredited. Let me maintain myself at this level. What about sustainability? Yes, as long as it maintains that level, but will it lead to improvement?*"

MISVII: "*So they give 5-year accreditation. If you get it, you'll stay like that for 5 years. You're not doing anything else. It doesn't make any improvements. . . you already think I provided it 5 years ago, and I am in the same situation again. I will provide it again. You cannot make improvements.*"

Participant MISIII, who thought that accreditation contributed to continuous improvement, said, "*It is beneficial because accreditation changes constantly. When accreditation changes, you will try to catch up with the times... Accreditation challenges you, provides control... accepts you, and accredits you but says, look, you have a deficiency here. At least someone says that the king is naked.*" This statement drew attention to the fact that the criteria of accreditation agencies are also updated, that they expect these updates from the institutions they will accredit and that they control the institutions they will accredit, and emphasized that continuous improvement is achieved thanks to these updates. Participant MISV drew attention to the relationship between accreditation and continuous improvement with this statement, "*. . . if you are already entering an accreditation process, it is like riding a bicycle. . . It is a process that you should not stop. You should constantly move forward, like a bicycle ride. Therefore, they support each other. Continuous improvement ensures accreditation. Accreditation requires continuous improvement. They create a cycle within themselves.*" The participant with code MISX emphasized the continuity of the accreditation process and touched upon its positive contribution to continuous improvement with the statement, "*I think it has critical contributions in terms of continuous improvement. . . It is similar to TQM, that is, continuity in total quality management, that is, it is not like saying, 'Okay, we are accredited at once, everything is finished,' but in the form of continuous improvement.*"

When the answers to the question of what indicators of the best accreditation system are examined, the prominent codes are "accreditation agency recognition, accreditation agency experience, and stakeholder contribution." "The internationality of the accreditation agency, accreditation agency image, accreditation agency guidance" constitute the codes related to the accreditation agency that determine a good accreditation system. The code related to the "accreditation agency recognition" stated by 5 participants consists of expressions such as worldwide recognition, international recognition, and awareness. The distribution of codes for the indicators of the best accreditation system is presented in Table 8.

Table 8. Code Distribution of Criteria Determining the Indicators of the Best Accreditation System

Code System	Number of Participants (out of 11)	Percentage of Participants
Accreditation agency recognition	5	45.4
Accreditation agency experience	3	27.2
Stakeholder contribution	3	27.2
Accreditation agency image	2	18.1
Accreditation agency guidance	2	18.1
Meeting department requirements	2	18.1
Being international	2	18.1
Applicability of the criteria	1	9.0
Corporate culture	1	9.0
Leadership	1	9.0
Acceptance of graduates	1	9.0
Provided collaboration network	1	9.0
Quality-oriented standards	1	9.0
Independence of Universities	1	9.0
Attitude of top management	1	9.0

The codes that stand out in evaluating the Turkish higher education system's compliance with accreditation criteria are "capacity planning problems, limitations, differences between universities, and failure to fully implement the criteria." Table 9 presents the distribution of codes related to the Turkish higher education system category according to the participants.

The statements that make up the code "Capacity planning problems" generally include an insufficient number of teaching staff and the distribution of quotas not being in line with the physical conditions of institutions. The participant with the code MISV noted that the limited capacity could not meet the increased quotas with the statement, "*I request 40 students every year. They send me 70 students every year, and then the students rightfully complain and complain.*" The participant with code MISVII drew attention to the problems caused by the increased quotas in computer laboratories. The statement, "*Currently, our computer laboratory has a capacity of 45 people. CoHE gives us a quota of 65. With the horizontal and vertical transfer international students, my first-year student count is 150; there is such a distorted*

Table 9. Code Distribution Regarding Turkish Higher Education System

Code System	Number of Participants (out of 11)	Percentage of Participants
Capacity planning problems	6	54.5
Limitations	5	45.4
Failure to fully implement the criteria	4	36.3
Differences between universities	4	36.3
Insufficient resources	3	27.2
Not ready for accreditation	1	9.0
Level of awareness	2	18.1
Advantages of Bologna studies	1	9.0

quota structure. And these 150 people are not the remaining students. They are students taking a course for the first time." The participant with code MISVI stated, " *CoHE is also responsible for the teaching staff and our technical needs at work. . . I need teaching staff, the number of students per teaching staff is high.*", emphasizing that the number of students per teaching staff is high due to the increase in quotas. The participant with code MISIX, who works at a foundation university, stated that the accreditation criteria are compatible with the Turkish higher education system, considering a national accreditation agency.

Category 2: Management Information Systems Accreditation

Prominent codes were determined in line with the questions to understand faculty members' accreditation knowledge levels, MIS departments' accreditation agency preference evaluations, and their perspectives on ABET accreditation. It was observed that all participants had general information about the accreditation process. Table 10 presents the code distribution of accreditation criteria suitable for the MIS program.

Table 10. Code Distribution of Accreditation Criteria for MIS Programs

Code System	Number of Participants (out of 11)	Percentage of Participants
Interdisciplinary structure	6	54.5
Evaluation different from social sciences	4	36.3
General accreditation preference criteria are valid	3	27.7
Processes and criteria	3	27.2
Mastery of the MIS field	1	9.0
MIS awareness	2	18.1

When the answers to the accreditation criteria to be preferred in the accreditation of MIS programs are examined, the prominent codes are as follows: "interdisciplinary structure, evaluation different from social sciences, processes, and criteria." 3 participants stated that there are no different criteria to consider when evaluating MIS programs and that "general accreditation preference criteria are valid." Most participants evaluated the accreditation process by drawing attention to the interdisciplinary structure of MIS programs. The participant with code MISIV drew attention to the fact that MIS departments in Türkiye are located under different units with the statement, "*As you know, we are in the Faculty of Economics and Administrative Sciences, the Faculty of Applied Sciences, and the Faculty of Business Administration... However, this management information system may not provide all the features of the faculty it is in. There needs to be an accreditation place that sets the standard for this. . .* " The participant with code MISXI stated, "*. . . it is more logical to choose an interdisciplinary institution. Due to the structure of MIS, if we choose an institution that is completely focused on engineering, then we may be subject to a very harsh evaluation process, and in fact, we may receive some unnecessary criticism.*" Similarly, the interdisciplinary structure of MIS programs was emphasized by the participant with the code MISIX. "*As we said, we do not have a distinction such as completely social sciences or science, that is, since it is an interdisciplinary department. . . maybe different things can be done about application-based courses, etc. specifically for MIS.*" When the statements representing the code "evaluation differently from social sciences" were examined, it was stated that it would not be right to evaluate it only under the social sciences, again emphasizing its interdisciplinary structure. The participant with the code MISXI stated, "*. . . if an institution that is completely communicative and focused on social sciences is chosen, some points may be missed there.*" and drew attention to the problems that may arise because MIS programs are not evaluated one-sidedly and this one-sided perspective. The MISVII branch participant also stated the relationship between MIS departments and engineering with the statement, "*For example, MIS should not be considered under management. . . It is something that should be evaluated together with industrial engineering.*", again thinking that it would not be right to evaluate only the social sciences.

The first question about the ABET accreditation agency was intended to measure the participants' knowledge about it. In the answers of the majority of the participants, the name of the ABET accreditation agency was heard, but there was no information about the processes and criteria. Participants stated this situation with the expression, "*I heard*

about it, but I don't have much of an idea. I heard about it, but I don't know the details." 3 participants stated that they had no information about ABET before, using the expression, "*I heard about it for the first time, no I haven't.*" It was observed that participants with codes MISII and MISVII had information about the processes and criteria for ABET accreditation. The participants' distribution of the codes regarding ABET accreditation agency knowledge is presented in Table 11.

Table 11. Distribution of Codes Related to ABET Accreditation Agency Information

Code System	Number of Participants (out of 11)	Percentage of Participants
Partially (I heard, but I have no information)	6	54.5
Yes	2	18.1
No	3	27.2

Due to their low level of knowledge about the ABET accreditation agency, they were first given information about the ABET and then explained the information systems program accreditation criteria. After this information, they were asked to evaluate the MIS undergraduate program in which they were working according to the ABET accreditation criteria. Responses were received for each criterion, and coding was performed for criteria found to be missing. The codes created for the criteria found to be missing in the program evaluation according to the ABET criteria are presented in Table 12.

Table 12. Code Distribution for Missing Criteria according to the ABET Criteria

Code System	Number of Participants (out of 11)	Percentage of Participants
Institutional support	5	45.4
Facilities	5	45.4
Student outcomes	3	27.2
Faculty	3	27.2
Curriculum	0	0.0
Program educational objectives	0	0.0
Continuous improvement	0	0.0
Students	0	0.0

When Table 12 is examined, in the coded structure created in line with the 8 criteria presented by the ABET accreditation agency, the criteria the participants found deficient are "institutional support, facilities, student outcomes, and faculty." The participants stated that they had no deficiencies in the 3 uncoded situations (MISII, MISV, MISIX). For the difficulties experienced in the institutional support criterion, the participant with code MISVII was compared with a national accreditation agency with the statement, "*Departments and universities are not very comfortable with institutional support*". Moreover, stated that due to the centralized structure of the Turkish higher education system, the deficiency in this criterion could be understood in the evaluation of the national accreditation agency, but that there would be difficulties in international accreditation. The participant with code MISVII also emphasized the faculty professional development support heading in the institutional support criterion and said, "*. . . there is no professional development support. Maybe this has become clearer in the last few years, or I have become more aware. I do not know, but I believe academics are left alone in terms of professional development support. My personal development is entirely dependent on my efforts.*" The participant with code MISIV also noted the limited budget opportunities and said, "*We do not have a program budget or financial support. We only have a faculty budget for the program budget and financial support. Our faculty budget is also limited.*"

It was stated that the facilities criterion could not be met due to the generally increased quotas. The participant with code MISIV drew attention to the fact that the laboratories could be insufficient with the statement, "*We have 2 laboratories. . . 60 people each. Sometimes, students cannot fit into intensive courses. For this reason, we conduct those courses in groups.*" The participant with code MISVI noted that the laboratories could be insufficient with the statement, "*The facilities are always insufficient for the teachers. No situation you can say is sufficient because this is related to technology. Do we have a laboratory for the metaverse? No. Do I have a special laboratory for Unity? No, I don't.*" It was noted that they have a laboratory in the field of big data and that it works very well. However, other laboratories are needed for subjects that fall within the scope of MIS programs, and the importance of laboratories for students and technology development due to the structure of MIS programs was emphasized.

Category 3: Problematic Areas and Benefits of Accreditation

Prominent codes were determined based on the answers to the questions about problematic areas of accreditation and the benefits of accreditation.

3 of the interviewed participants stated that they did not have any information about accreditation costs. 1 participant,

without relevant information, worked at an institution that did not have accreditation studies based on their faculty or department. The other 2 participants worked in the MIS department of a foundation university. 4 participants who were informed about the cost evaluated the international and national accreditation by comparing them and stated that international accreditation is more costly. The participant with code MISI expressed this situation as "*Although it is more costly and takes longer, international accreditation has very important advantages.*" In the answers given to the question "Do you have a budget to cover this cost?", 10 participants stated that they had a budget, while 1 participant working at a foundation university stated that they did not have the necessary budget. When the answers to whether accreditation is worth the high costs are examined, the prominent codes are "worth the cost, varies by accreditation agency, varies by the institution to be accredited." None of the participants think that it is not worth the relevant costs.

Table 13 presents the distribution of codes by participants on the question of whether accreditation is worth the high costs.

Table 13. Code Distribution Regarding High Accreditation Costs

Code System	Number of Participants (out of 11)	Percentage of Participants
Worth the cost	5	45.4
Varies by accreditation agency	4	36.3
Varies by the institution to be accredited	2	18.1
Not worth	0	0.0

The participant with code MISXI, who has a value mindset without any restrictions, stated, "*Absolutely value. . . . so many things have changed in the institution and department that if a university wants to develop, it needs to go through these processes. In other words, it deserves the benefits it will provide.*" Although it causes high costs, it has been evaluated as valuable. There is a belief that it is especially necessary for developing universities. Participants with the code MISVI drew attention to the importance of human capital with the statement, ". . . *this is a much more important issue than money. In other words, human capital is significant. You are training people here. I am speaking as an educator. In other words, when it comes to education, everything else is unimportant. . . We said accreditation is needed. For education, for quality education.*", and emphasized the importance of the control mechanism that accreditation will provide and stated its importance in ensuring quality education. Under the code that "varies by accreditation agency", there are topics such as the benefits the accreditation will provide in the international arena, its role in attracting quality students, and the convenience it will provide in student exchanges. For example, the idea that national accreditation would not offer these benefits and would not be worth the high costs to be paid was expressed. The participant with code MISVII expressed this situation in the cost calculation and made for a national accreditation agency in the following way: "*It is not worth it at the moment. I do not see much value in accreditation.*" The participant with code MISIII emphasized the importance of international accreditation and stated that the returns of accreditation would be much more than the costs to be endured with the statement: "*For example, if you start accepting students from abroad, it would be good. If you start raising your level, these are all the returns, then you can earn much more.*" The code that "varies by the institution to be accredited" includes the idea that the department is not ready for this process, there are deficiencies in basic quality processes, and the budget that the institution will allocate for accreditation is not worth the returns of accreditation for the institution. The participant with code MISVIII stated, "*If I am still at the beginning of the road, I need to learn to walk a little earlier. I need to go through quality processes. I need to improve the process. Maybe we need to work with more national accreditation companies.*" and emphasized that the institution must fulfill basic quality requirements and that preparing for the process before bearing accreditation costs is important.

When the answers given to the question of what are the biggest obstacles to accreditation are examined, the codes that emerged are as follows: "budget, insufficient resources, lack of expert personnel, unconsciousness/unawareness, lack of top management support, will, limitations, established corporate culture, anxiety." Table 14 presents the distribution of codes regarding obstacles to participant accreditation.

Most participants see the "budget" as one of the biggest obstacles to accreditation. The participant with code MISVII drew attention to the dependency on the state in terms of the budget. The statement, ". . . *the state can only provide services to the extent of the budget it provides, and that is an obstacle.*" The participant with code MISIV similarly drew attention to the dependency on the state and stated that the limited budget should be allocated to priority areas, and expressed this situation: "If we evaluate it according to the university we are at, for example, I think our university's priority. . . the budget should be allocated for building construction. . . we do not have a building. . . the university is newly established. New faculties are being opened. New research centers are being opened."

The prominent code "insufficiency of resources" consists of a lack of academic and administrative staff and physical infrastructure deficiencies. The participant with code MISVI stated that "there is a concern whether we will be sufficient

Table 14. Code Distribution Regarding Obstacles to Accreditation

Code System	Number of Participants (out of 11)	Percentage of Participants
Budget	5	45.4
Insufficent resource	4	36.3
Unconsciousness/unawareness	3	27.2
Lack of expert personnel	3	27.2
Limitations	2	18.1
lack of top management support	2	18.1
Will	2	18.1
Established corporate culture	1	9.0
Anxiety	1	9.0

both in terms of human resources and technically, that is, whether we will be sufficient both as faculty members and technically" and expressed their situation of not being ready for the accreditation process as an institution and their concerns about issues that will create obstacles during the process. The participant with code MISII drew attention to the lack of administrative staff and said, "We have to handle it with our resources, our personnel resources," and mentioned that as a state university, they cannot make independent decisions in hiring new staff.

Another prominent code is the "unawareness/unconsciousness" code, which describes not having enough information about accreditation processes and their importance. The participant with code MISXI emphasized the importance of stakeholders and awareness in completing the process. The statement, "If there is no desire in the institution to improve such things or if the management does not create this desire, I think it is difficult. Because again, at some point, the people who will make everything happen are the people in the departments, that is, the faculty members there." The participant with code MISIV evaluated the budget as the first of the biggest obstacles to accreditation. They expressed the second biggest obstacle as "unawareness, that is, lack of awareness."

Another issue that will create an obstacle in the accreditation process is the code "lack of expert personnel," which indicates the lack of administrative or academic personnel who will deal with the process and have knowledge about the process. The participant with the code MISIX stated, "An expert staff will definitely... that is, someone who is knowledgeable about this subject" and drew attention to the workload brought by accreditation, indicating the need for the employment of expert personnel who will only deal with the accreditation process in both reducing the workload and controlling the processes. The participant with code MISI stated, "We were lucky. We had 3 dedicated, expert, and qualified personnel with high English proficiency, and there were 3 people in the dean's office. All three of them worked together and completed this process by working in harmony." Moreover, emphasized the importance of expert personnel and stated that completing the process would be difficult if they were lacking.

When the answers given to the question of how being a state university affect the process are examined, the codes of "resource limitation, flexibility in decision-making" contain negative evaluations, while the codes of "having qualified students, high competence and proficiency" express the positive contribution of being a public university. The distribution of codes regarding the effect of being a state university on the process according to the participants is presented in Table 15.

Table 15. Code Distribution Regarding the Effect of Being a State University on the Process

Code System	Number of Participants (out of 7)	Percentage of Participants
Resource limitations	5	71.4
Flexibility in decision-making	2	28.5
Having qualified students	1	14.2
High competence and proficiency	1	14.2

The participant with code MISIV stated, "*The financial resources of the state university are limited. The prices given by the accreditation agencies I have examined are $3,000 and $5,500 for just one department. It is not possible for us to give it.*" This draws attention to the obstacles that the budget will cause and indicates that the resources of the state university are limited. Similarly, the participant with code MISVI stated, "*Therefore, of course, we can also say something about financial resources and their management. We are a bit limited in those matters, state universities. They are all like that.*"

Regarding the code "Flexibility in decision making," the participant with code MISVI stated, "Of course, we cannot write our own goals. We cannot do that, especially about student conditions and the number of students. Our SWOT analyses are not very useful. We are very dependent, we are dependent on process management in state universities." This indicates that there is no authority to make decisions in line with the analyses conducted and no independent

decision-making authority. The participant coded MISVIII stated that ". . . *compared to foundation universities, state universities have more qualified students and that competence and proficiency are higher in state universities..*".

The 3 interviewed participants work in an MIS department at a foundation university. The "resource limitation" code that emerged regarding the impact of being a state university on the process also emerged similarly in the foundation university evaluation. The participant with code MISX expressed the following response: "*The staff and opportunities at state universities are very extensive, but this is not the case here.*" The codes "Taking quick action and quick access to top management" are the codes that express the advantages of being a foundation university. The participant with code MISIX expressed this situation as follows: ".. *our university is also a corporate university, but things do not work as slowly as in the state ones, it allows us to take quicker action.*" The distribution of the codes regarding the impact of being a foundation university on the process is presented in Table 16.

Table 16. Code Distribution Regarding the Effect of Being a Foundation University on the Process

Code System	Number of Participants (out of 3)	Percentage of Participants
Resource limitation	2	66.6
Quick action	1	33,3
Quick access to top management	1	33.3

When the answers to the advantages of accreditation are examined, the codes that emerge with equal frequency are "increasing quality, providing recognition, providing improvement, cooperation network."

Table 17 presents the distribution of codes regarding the advantages/benefits of accreditation according to the participants. One prominent code, "increasing quality," concerns the issues of increasing the quality of research, education, and training and attracting quality students and teaching staff.

Table 17. Code Distribution Regarding the Advantages and Benefits of Accreditation

Code System	Number of Participants (out of 11)	Percentage of Participants
Increasing quality	7	63.6
Providing standards	7	63.6
Providing recognition	7	63.6
Providing improvement	7	63.6
Collaboration network	7	63.6
Process management	5	45.4
Providing prestige	2	18.1
Documentation and accessibility	2	18.1
Student exchange	2	18.1
Increasing competition	1	9.0

The participant with code MISI stated the situation by saying, "Both the quality of education and research is increasing." The participant with code MISVIII emphasized the change in the quality of education by evaluating it with the change in the curriculum and expressed this situation: "... *here is the change in the curriculum; the quality of education is increasing*." *Another prominent code is the code "providing improvement.*" The participant with code MISIII drew attention to the accreditation criteria and emphasized that it provides an improvement in many areas, and expressed this situation as "*Accreditation, by the way, means living like a human being. In other words, there are certain criteria. It says don't living in such a cramped campus, for example.*" The participant with code MISV evaluated it with continuous improvement and said, "*Accreditation requires continuous improvement; they create a cycle within themselves.*" The participant with code MISI stated that the accreditation process brings improvement with the statement ... it may force us to develop and improve ourselves." The code "providing recognition," which was clearly expressed by the participants, was evaluated as national or international according to the accreditation agencies. Participant MISIII emphasized the importance of international accreditation and stated, "*There is also this in accreditation, for example, when you apply for a master's degree, you need to explain your transcript. The student needs to explain. Some documents are sent from here, and some are received from there, but when you are accredited, it is already clear what it is. Especially if they are accredited by the same institution, the other party provides something amazing. That's why I emphasized global.*" Another prominent code is the "providing standards" code, which refers to the provision of accepted and required criteria. Participant MISVIII drew attention to the effort to provide the necessary criteria and expressed this situation as, "*When we consider the sub-divisions of accreditation processes, there is a theme related to each process, and there is an effort to meet a relevant standard in each process.*"

When the answers to the question of the disadvantages of accreditation are examined, the prominent codes are

"workload, difficulty in documenting, following the processes, bearing the cost." Table 18 presents the distribution of the codes regarding the disadvantages of accreditation according to the participants.

Table 18. Code Distribution Regarding Disadvantages/Difficulties of Accreditation

Code System	Number of Participants (out of 11)	Percentage of Participants
Workload	6	54.5
Difficulty in documenting	4	36.3
Following the processes	3	27.2
Bearing the cost	2	18.1
Pushing educational institutions' resources	1	9.0
Seeing it as the ultimate goal	1	9.0
Cultural differences	1	9.0
Resistance to change	1	9.0
Eliminating differences	1	9.0
Lack of a compatible team	1	9.0
Long-lasting process	1	9.0

Most participants see "workload" as a significant disadvantage. The participant with code MISVII evaluated the time and effort required for accreditation with the statement, "*I can't say much in general, but it puts a burden on the department, faculty members spend time on accreditation instead of the time they should spend on students. These are disadvantages, and they cause excessive load.*" The participant with code MISII similarly emphasized the situation that the process will cause with the statement, "*Maybe it will cause boredom as workload.*" The participant with code MISX also stated the problem with the statement, "*Workload, there is no other disadvantage. . .* "

Another code that expresses the disadvantage of accreditation is the code that describes the situation related to the process and ensures that the work and processes are recorded and can be accessed whenever desired. Regarding this code, expressed as "difficulty in documentation," the participant with code MISIII emphasized the difficulty that the recording process will create by saying, "*Writing a lot of reports, writing reports, and trying to explain things.*" The participant with code MISIX stated "*Paperwork, writing work, I think documentation is the biggest thing*", while the participant with code MISX noted the situation similarly, "*Just a lot of paperwork*".

5. DISCUSSION

The combination of information and communication technology (ICT) developments and global organizational changes has resulted in an international division of labor. As economies integrate into the new knowledge economy, industrialization increasingly depends on developing more complex ICT. There is now a greater need for a highly skilled and technologically educated workforce (Sanyal and Martin, 2007). Socio-economic and cultural transformations have put pressure on the system focus of education to shift from quantitative expansion to quality-oriented expansion (Sahney et al., 2004). Globalization, in turn, has created the need to facilitate the comparability of educational standards. These developments require adopting a typical competence structure and comparable systems for quality assurance (Hernes and Martin, 2008). Accreditation, defined as verifying compliance with specific standards by an independent institution, plays an important role here.

This study aims to examine the perceptions and perspectives of faculty members working in MIS departments regarding accreditation. The analysis of the interviews conducted with 11 faculty members working in the MIS department reveals important findings regarding the accreditation process. According to the study results, most faculty members consider accreditation and standardization together. Adelman and Silver (1990) defined accreditation as a centuries-old and voluntary process that aims to guarantee standards in schools and higher education. When examined as a process, it is expressed by many researchers as the verification of compliance with specific standards by an independent institution (Eaton, 2012; Van Berkel and Wijnen, 2010; Bruening et al., 2002; Aktan and Gencel, 2010; Erkuş, 2009; Chowdhury et al., 2013). While faculty members evaluate the primary purpose of accreditation as ensuring standards, they also draw attention to the process management associated with it. Alaskar et al. (2019) defined the accreditation process as actions implemented by academic staff and senior administrators that lead to specific results or outcomes. Institutional accreditation processes include standards set by accrediting bodies that encourage many activities to demonstrate compliance with measurable criteria within an institution and its programs (Alaskar et al., 2019). While the faculty members participating in the research provide the standards that accreditation will bring, they believe that process management should also be carried out. Research is being conducted on modeling, implementing, and improving business processes in higher education (Thomas et al., 2017; Ahmad et al., 2007; Seeman et al., 2006; Balzer, 2020). Menzli et al. (2019) investigated the relationship between business process modeling and quality by

focusing on accreditation processes in higher education. Bittick (2003) stated that the accreditation process provides an opportunity to look deeply into the entire process and solve problems that have not been noticed for a long time. Faculty members generally interpret the purpose of accreditation as described in the literature.

Considering the recent interest in accreditation, especially in Türkiye, whether accreditation is regarded as a temporary fashion constitutes an important research question. Accreditation is a deep-rooted process that has been used for many years to provide quality assurance. Academicians have stated that the first period in the historical development of accreditation began with the establishment of the University of the State of New York in 1784 as a board of King's College (now Columbia University) in New York City (Harcleroad 1980 and Wilkins 1959 as cited in Alstete, 2007). Eaton (2012) stated that the roots of accreditation date back to the end of the 19th century. It has become an important tool for the quality assurance of higher education institutions, especially in the United States. Although accreditation has deep roots in some countries, higher education accreditation studies in Türkiye began in 1994 with the application of some engineering departments for ABET accreditation. For accreditation to serve its purposes, a long-term quality assurance mechanism is needed. Most faculty members who participated in the research consider it a necessity or need for quality, although it is sometimes perceived as a fashion in Türkiye. It can be said that the faculty members have a general awareness of the contribution of accreditation to the quality of education and continuous improvement.

When evaluating the success of a good accreditation system, faculty members prioritize the recognition of the accreditation agency and the experience of the accreditation agency. Another critical issue is that stakeholders support and contribute to the process. The primary purpose of this research is to examine the perceptions and perspectives of stakeholders whose contributions to the process are of great importance in obtaining an idea about the applicability and sustainability of the accreditation process. The findings revealed that faculty members perceive that compliance with standards will be achieved and quality education will be achieved with accreditation. Along with the idea that the recognition that accreditation provides will bring cooperation, there is also the idea that accreditation increases the prestige of departments and will positively attract students. The perspective on the benefits of accreditation is parallel to many studies in the literature (Roller et al., 2003; Bitter, 2014; Wilson-Hail et al., 2019; Semerci et al., 2021; Alpaydın and Topal, 2022; Bougherira and Elasmar, 2023; Mady et al. 2023; Ito et al., 2024). According to the research results of Wilson-Hail et al. (2019), faculty members believe that the national accreditation process is important for status and prestige. Bitter (2014) concluded that faculty members believe that it increases the reputation of their programs in the eyes of their stakeholders. Bougherira and Elasmar (2023) found that faculty members have perceptions that it positively affects teaching and learning. The research findings also revealed that faculty members have some perceptions about the difficulties of the accreditation process.

Faculty members perceive intense effort is required to meet accreditation criteria and manage the process. They stated that the documentation and evidence creation processes, in particular, increase workload. This finding parallels the literature on the applicability and implementation of the accreditation process. According to Alstete (2004), the fact that most faculty members do not expect to be involved in tasks that will facilitate the accreditation process causes a contradiction between faculty expectations and accreditation needs and processes. Campbell and Rozsnyai (2002) stated that although they believe in the staff monitoring and review process, the combination of regional, specialized, and state reviews sometimes hinders teaching and research. Again, in the study of Wilson-Hail et al. (2019), faculty members think that workload is a disadvantage unless they are recognized for their work. In addition, the bureaucratic process of accreditation and the inadequacy of the necessary resources prevent this process from being effectively conducted. Shaiban (2024) emphasized that the accreditation process requires faculty members' effort and may cause stress. According to the research findings, the biggest obstacles to accreditation were expressed by faculty members as insufficient budget and resources. The process of obtaining initial accreditation and then periodically renewing it is generally considered time- and resource-intensive (Hegji, 2014). In the Campbell and Rozsnyai (2002) study, it is thought that the accreditation process is expensive but valuable because it provides an opportunity for most schools to review their work and ensure that they meet the accreditation criteria. Over seventy percent of the Mandy et al. (2023) survey respondents agreed that resources invested in AACSB accreditation will pay off in the long term. When asked whether the cost was worth it, no faculty member responded that it was not. Faculty members believe that evaluations should be performed differently depending on the accrediting agency or institution to be accredited. Considering that the benefits of accreditation are related to the competence of the accrediting agency, it can be considered an important constraint.

Another question the research seeks to answer is how the accreditation process is viewed regarding MIS departments and how readiness status is evaluated in line with the ABET criteria. Faculty members expressed their opinions that MIS departments should be evaluated differently from social sciences departments, considering their interdisciplinary structure. ABET establishes commissions to accredit programs in different disciplines and evaluates the process in line

with field-specific criteria. The fact that accreditation is relatively new in Türkiye, the few MIS departments accredited by international accreditation agencies, and the fact that there are no MIS departments with ABET accreditation also emerge as factors affecting the awareness of ABET. While only 2 of the faculty members who participated in the research stated that they had information about the ABET accreditation agency, 3 faculty members indicated that they had no information at all. 6 faculty members said they had heard about the organization but did not have any information. Faculty members believe that there are deficiencies mostly in institutional support and facilities criteria. ABET's expectation from the institutional support criteria is that "Institutional support and leadership should be sufficient to ensure the quality and continuity of the program." The facilities criterion is that "Classrooms, offices, laboratories, and related equipment should be sufficient to support the achievement of student outcomes and provide an atmosphere conducive to learning" (ABET, 2024). When evaluated together with the criticisms received from the Turkish Higher Education System, it can be concluded that the facilities are inadequate due to capacity planning problems. In their research in Türkiye, Semerci et al. (2021) concluded that institutional deficiencies are one of the obstacles to accreditation.

6. CONCLUSION

The research findings show that faculty members have a positive perspective on the accreditation process, which aligns with concepts such as quality education, standardization, continuous improvement, and process management. However, difficulties such as documentation, lack of awareness, resistance, failure to perform accreditation by its purpose, difficulty following the processes, and the workload imposed on faculty members have been identified. Faculty members have expressed their views on the issue of evaluating MIS departments differently from social sciences, considering their interdisciplinary structure. However, the scarcity of MIS departments accredited by international accreditation agencies and the absence of a MIS department evaluated with field-specific criteria affect faculty members' awareness of this issue. When this finding in the research results and the lack of awareness of ABET accreditation are evaluated together, this research reveals the importance of ABET accreditation, which assesses MIS programs using program-specific criteria. In this respect, this research can be considered as a guide.

The inadequacy of financial resources required to meet the standards is a significant obstacle to the sustainability of the process. The higher education system can support infrastructure development, laboratory equipment, and technical needs related to university education. It can control the increase in quotas by taking into consideration capacities. Long-term quality policies should be established at the higher education level to prevent accreditation from being seen as temporary and to create sustainable quality assurance systems. By adopting a quality-based education model, accreditation processes can be ensured as a permanent quality assurance tool in higher education. Special funds can be established for universities to provide the necessary resources for accreditation. These funds will provide valuable support, especially for universities that have difficulty meeting the workload and cost of the process. Financial support can be provided to universities to reduce the workload of the accreditation process and to encourage faculty members. This can facilitate process management by meeting the need for additional resources, especially during the documentation and evidence-creation stages. Policymakers can make the process more accessible by offering state incentives to cover accreditation costs. Supporting investments in tools that digitize documentation and evidence-creation processes can reduce faculty members' workload and increase process efficiency. Such tools can simplify and streamline accreditation preparations.

Accreditation agencies can organize regular introductory meetings, training, and guidance studies to increase their recognition in universities and departments. They can also create awareness and trust among faculty members and other stakeholders, mainly by providing informative materials about accreditation's contribution to the quality of education. Organizations' guidance services in the documentation and evidence-creation processes are essential to ease the difficulties of the accreditation process and reduce faculty members' anxiety about the process.

Future research can evaluate the efficiency of accreditation processes by examining their long-term effects on education quality, student satisfaction, and graduate success. It can analyze the effects of accreditation on the increase in education quality and universities' overall performance. Such studies can provide concrete data on the necessity of accreditation. Conducting qualitative and quantitative research to deeply understand faculty members' perceptions of accreditation will contribute to implementing the accreditation process more actively and effectively. Studies need to be conducted within and with accreditation agencies to examine the factors affecting faculty members' motivation during the accreditation process and to determine the factors that reduce the workload in the process. These studies can make the process more efficient and participatory for faculty members. Studies can be conducted to analyze the effectiveness of the contributions of students, sector representatives, and faculty members to the accreditation process. Such studies will provide a deeper understanding of the effects of stakeholder support on the success of accreditation. Comparative

analyses can be conducted between different universities and departments to assess the effects of accreditation on the recognition and prestige of universities at national and international levels. These studies can reveal the effects of accreditation processes on attracting students, graduate employment, and sectoral collaboration with more concrete data.

Peer Review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- E.C., N.S., E.C.T.; Data Acquisition- E.C., E.C.T.; Data Analysis/Interpretation- E.C., N.S., E.C.T.; Drafting Manuscript- E.C.T.; Critical Revision of Manuscript- E.C., N.S.; Final Approval and Accountability- E.C., N.S., E.C.T.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

ORCID IDs of the authors

Esra Cengiz Tırpan 0000-0001-7675-5635
Erman Coşkun 0000-0001-8712-3246
Nihal Sütütemiz 0000-0002-8964-7198

REFERENCES

AACSB (2023). *2020 Guiding principles and standards for business accreditation.* Retrieved July 10, 2024 from https://www.aacsb.edu/-/media/documents/accreditation/2020-aacsb-business-accreditation-standards-june-2023.pdf

ABET (2023). Criteria for accrediting computing programs, 2024 – 2025. Retrieved August 18, 2024 from https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-computing-programs-2024-2025/

ABET (2024). *About ABET*. Retrieved August 24, 2024 from https://www.abet.org/about-abet/

Adelman, C., & Silver, H. (1990). *Accreditation: The American Experience*, Council for National Academic Awards: London.

Ahmad, H., Francis, A. & Zairi, M. (2007), Business process reengineering: critical success factors in higher education", *Business Process Management Journal, 13*(3), 451-469. https://doi.org/10.1108/14637150710752344

AKPINAR, H., (2011). Türkiye'de Enformasyon Sistemleri Öğretiminde 20. Yıl. Uluslararası 9. Bilgi, Ekonomi ve Yönetim Kongresi Bildirileri (pp.565-665).

Aktan, C. C., & Gencel, U. (2010). Yüksek öğretimde akreditasyon. *Organizasyon Ve Yönetim Bilimleri Dergisi*, 2(2), 137-146.

Al Shobaki, M. J., & Naser, S. S. A. (2017). The role of the practice of excellence strategies in education to achieve sustainable competitive advantage to institutions of higher education-faculty of engineering and information technology at Al-Azhar University in Gaza a model. *International Journal of Digital Publication Technology, 1*(2), 135-157.

Alaskar, A., D'Errico, E., Alipoon, L., & Dehom, S. (2019). Institutional accreditation in Saudi Arabian higher education: perceptions and involvement. *Quality in Higher Education*, 25(3), 245–260. https://doi.org/10.1080/13538322.2019.1667630

Alpaydın, Y., & Topal, M. (2022). Eğitim fakültelerindeki akreditasyon deneyimleri üzerine nitel bir araştırma. *İnsan ve Toplum, 12*(2), 232-265.

Alstete, J. W. (2004). *Accreditation matters: Achieving Academic Recognition and Renewal. ASHE-ERIC Higher Education Report.* (1st edition). John Wiley & Sons.

Alstete, J.W. (2007). A brief history of college accreditation. *In: College accreditation managing internal revitalization and public respect* (pp.11-21). Palgrave Macmillan, New York. https://doi.org/10.1057/9780230601932_2

Angulo-Ruiz, F., Pergelova, A. & Cheben, J. (2016). The relevance of marketing activities for higher education institutions. In Wu, T. & Naidoo, V. (Eds) International marketing of higher Education (pp.13-45). Palgrave Macmillan, New York. https://doi.org/10.1057/978-1-137-54291-5_2

AQAS (2024). *Criteria & indicators for programme accreditation (decision of the accreditation commission of 01.12.2015, version of 26.09.2019*) Retrieved July 18, 2024 from https://www.aqas.eu/downloads/Criteria%20for%20Programme%20Accreditation.pdf

Aytaçlı, B. (2012). Durum çalışmasına ayrıntılı bir bakış. *Adnan Menderes Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Dergisi, 3*(1), 1-9.

Balzer, W.K. (2020). *Lean higher education: Icreasing the value and performance of university processes*, (2nd ed.). Productivity Press. https://doi.org/10.4324/9781351216944

Başkale, H. (2016). Nitel araştırmalarda geçerlik, güvenirlik ve örneklem büyüklüğünün belirlenmesi. *Dokuz Eylül Üniversitesi Hemşirelik Fakültesi Elektronik Dergisi, 9*(1), 23-28.

Bengshir Kaya, T. (2002). Türkiye'de yönetim bilişim sistemleri disiplininin gelişimi üzerine düşünceler. *Amme İdaresi Dergisi, 35*(1), 77-103.

Bitter, M. E. (2014). The benefits of AACSB accounting accreditation: Perceptions of administrators of accounting accredited programs. *Academy of Educational Leadership Journal, 18*(1), 85.

Bittick, J. C. (2003). Evolving Through Accreditation. Corrections Today, 65(2), 8. Retrieved August 18, 2024 from https://research.ebsco.com/linkprocessor/plink?id=66b9441d-87a7-3b24-a01d-df6ae344f76b

Blanco Ramírez, G. (2015). International accreditation as global position taking: An empirical exploration of US accreditation in Mexico. *Higher Education, 69*, 361-374.

Bougherira, M. R., & Elasmar, M. H. (2023). Impact of academic accreditation on teaching and learning: faculty members' perceptions. *Journal of Further and Higher Education, 47*(2), 167–181. https://doi.org/10.1080/0309877X.2022.2102412

Bruening, K. S., Mitchell, B. E., & Pfeiffer, M. M. (2002). Accreditation standards for dietetics education. *Journal of the Academy of Nutrition and Dietetics, 102*(4), 566-577.

Caeiro, S., Sandoval Hamón, L. A., Martins, R., & Bayas Aldaz, C. E. (2020). Sustainability Assessment and Benchmarking in Higher Education Institutions—A Critical Reflection. Sustainability, 12(2),1-30. https://doi.org/10.3390/su12020543

Campbell, C., & Rozsnyai, C. (2002). *Quality assurance and the development of course programmes.* Papers on Higher Education Regional University Network on Governance and Management of Higher Education in South East Europe Bucharest, UNESCO.

Ceccucci & White (2008). IS faculty perceptions of ABET. *Accreditation. Information Systems Education Journal, 6* (54), 1-8.

Cho, J. & Trent, A. (2006). Validity in qualitative research revisited. *Qualitative Research, 6*(3), 319–340.

Chowdhury, H., Alam, F., Biswas, S. K., Islam, M. T., & Islam, A. S. (2012). Quality Assurance and Accreditation of Engineering Education in Bangladesh. *Procedia Engineering, 56*, 864-869. https://doi.org/10.1016/j.proeng.2013.03.208

Coşkun E., Özşahin M., Bozkuş Kahyaoğlu S. & Alma Çallı B. (2020, 23-25 September). *Accreditation process in turkish business schools: recommendations based on awareness, perceptions, and problematic areas.* 19. Uluslararası İşletmecilik Kongresi, Kayseri, Türkiye.

Coşkun, E., & Alma, A. G. B. (2018). Copıng Wıth Qualıty Issues in Management Informatıon Systems Departments In Turkey Through Accredıtatıon. *The Online Journal of Quality in Higher Education, 5*(2), 29-37.

Council for Higher Education Accreditation (2024). *About accreditation*. Retrieved July 28, 2024 from https://www.chea.org/about-accreditation#what.

Creswell, J. W. 2003. *Research design: Qualitative, quantitative, and mixed methods approaches.* (2nd ed.) Thousand Oaks, CA: Sage

Cropley, A. J. (2002). *Qualitative research methods: An introduction for students of psychology and education.* Zinatne

Eaton, J. S. (2012). What future for accreditation: The challenge and opportunity of the accreditation–federal government relationship. *Inquiry, evidence, and excellence: The promise and practice of quality assurance* (77-88). Teacher Education Accreditation Council. https://files.eric.ed.gov/fulltext/ED541693.pdf#page=89

Eaton, J. S. (2015). An overview of us accreditation. revised november 2015. Council for Higher Education Accreditation, 1-12. https://files.eric.ed.gov/fulltext/ED569225.pdf

Erkuş, L. (2009). *Eğitim fakültelerinin akreditasyon sürecine hazır olma durumuna ilişkin öğretim elemanlarının görüşlerinin değerlendirilmesi.* (unpublished master's thesis) Kırıkkale Üniversitesi.

Erlandson, D. A., Harris, E. L., Skipper, B. L., & Allen, S. D. (1993). *Doing naturalistic inquiry: A guide to methods*. Sage Publications.

Frasquet, M., Calderón, H., & Cervera, A. (2012). University–industry collaboration from a relationship marketing perspective: An empirical analysis in a Spanish University. Higher Education, 64, 85-98. https://doi.org/10.1007/s10734-011-9482-3

Gilbert, G. (2010). Making faculty count in higher education assessment. *Academe, 96*(5), 25-27.

Grix J. (2010). *Foundations of research,* (2nd ed.). Palgrave Macmillan.

Guba, E. G., & Lincoln, Y. S. (1982). Epistemological and methodological bases of naturalistic inquiry. *Educational Technology Research and Development, 30*(4), 233-252. https://doi.org/10.1007/BF02765185

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1), 59-82. https://doi.org/10.1177/1525822X05279903

Harvey, L. (2004). The power of accreditation: views of academics1. *Journal of Higher Education Policy and Management, 26*(2), 207–223. https://doi.org/10.1080/1360080042000218267

Harvey, L., & Knight, P. T. (1996). *Transforming Higher Education*. Open University Press, Taylor & Francis, 1900 Frost Road, Suite 101, Bristol, PA 19007-1598.

Harvey, L., & Williams, J. (2010). Fifteen years of quality in higher education. *Quality in Higher Education, 16*(1), 3–36. https://doi.org/10.1080/13538321003679457

Hegji, A. (2014). *An overview of accreditation of higher education in the United States*. Retrieved May 18, 2024 from https://www.everycrsreport.com/files/20141212_R43826_b164b03be4ae59db66b18659613e7befa573edf2.pdf

Hernes G. & Martin, M. (Eds.). (2008). *Accreditation and the global higher education market.* International Institute for Educational Planning. https://coilink.org/20.500.12592/6djhbg6

Herr, K. & Anderson, G. L. (2015). *The action research dissertation: a guide for students and faculty* (2nd Ed.). SAGE Publications.

Hijazi, R. (2016). Statistics education in gcc business schools. Global Business and Economics Review, 18(1), 1-14. https://doi.org/10.1504/gber.2016.073256

Hilton, T. S (2003). MIS Program Accreditation: Comparing AACSB and ABET. *In the Proceedings of Information Systems Education Conference (ISECON)*. Retrieved August 24, 2024 https://iscap.us/proceedings/isecon/2003/2211/ISECON.2003.Hilton.pdf

Hilton, T. S., Johnson, D. A., & Kasper, G.M. (2004), ABET Accreditation of MIS Programs in AACSB Schools, *In the Proceedings of Information Systems Education Conference (ISECON)*, 21, 1-16.

Hoare, A., & Goad, P. (2022). The quality continuum: perceptions of institutional accreditation. Quality Assurance in Education, 30(1), 102-117. https://doi.org/10.1108/QAE-08-2021-0135

Ito, H., Takeuchi, S., Yokoyama, K., Makita, Y., & Ishii, M. (2024). Impact of AACSB accreditation on education quality: perceptions of faculty in an accredited school in Japan. *International Journal of Educational Management, 3*8(3), 769-785. https://doi.org/10.1108/IJEM-05-2023-0258

İstanbul Gelişim Üniversitesi (2024). *Bölüm tanıtımı*. Retrieved August 18, 2024 from https://iisbf.gelisim.edu.tr/tr/akademik-bolum-yonetim-bilisim-sistemleri-icerik-bolum-tanitimi

King, W. R. (1978). Strategic planning for management information systems. MIS Quarterly, 2(1), 27-37.

Kroenke, David M. (2012), Experiencing MIS. (3th Ed.). Pearson.

Kumar, P., Shukla, B., & Passey, D. (2020). Impact of accreditation on quality and excellence of higher education institutions. *Investigación Operacional, 41*(2), 151-167.

Laudon, Kenneth C., & Jane P. Laudon. 2014. *Management information systems: Managing the digital firm*. (13th Ed.). Pearson.

LeCompte, M. D., & Goetz, J. P. (1982). Problems of reliability and validity in ethnographic research. *Review of educational research, 52*(1), 31-60.

Leidig, P., Salmela, H., Anderson, G., Babb, J., de Villiers, C., Gardner L., Nunamaker, J.F., Scholtz, B., Shankararaman, V., Sooriamurthi, R. & Thouin, M., (2021). The joint ACM/AIS IS2020 task force, IS2020 a competency model for undergraduate programs in information systems. Association for Computing Machinery, New York, NY, USA. Retrieved August 18, 2024 from https://www.acm.org/binaries/content/assets/education/curricularecommendations/is2020.pdf

Lewis, S. E. (2016). *Perceptions of university faculty regarding accreditation in a college of education* (Unpublished doctoral thesis), University of South Florida.

Lozano, R., Lukman, R., Lozano, F. J., Huisingh, D., & Lambrechts, W. (2013). Declarations for sustainability in higher education: becoming better leaders, through addressing the university system. *Journal Of Cleaner Production, 48*, 10-19.

Lundquist, R. (1997). Quality systems and ISO 9000 in higher education. *Assessment & Evaluation in Higher Education,* 22(2), 159-172.

MacKinnon, R. J., Elder, K. L., & Dyer, J. N. (2012). The Current Status of ABET Accreditation of Information Systems Programs. *Issues in Information Systems, 13*(1), 151-159.

Mady, M. T., Al-Hussainan, A., & Mady, T. T. (2023). AACSB-Accreditation benefits as predictor of faculty attitude toward the accreditation process: an exploratory study. *Journal of Management and Business Education, 6*(1), 24-42. https://doi.org/10.35564/jmbe.2023.0002

Menzli, L. J., Ayouni, S., & Elsadig, M. (2019). Quality assessment of business process models: Case of accreditation process in higher education. *International Journal of Civil Engineering and Technology, 10*(5), 874-884.

Merriam, S. B. & Tisdell, E. J. (2016).*Qualitative research: a guide to design and implementation* (4th Ed.). Jossey-Bass Publishers.

Merriam, S. B. (1998). *Qualitative research and case study applications in education. revised and expanded from case study research in education.* (2nd Revised & Expanded). Jossey-Bass Publishers

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.

Mpinganjira, M. (2011). Retaining Africa's talent: The role of Africa's higher education. *International Journal of Emerging Markets, 6*(2), 168-179.

Mursidi, A., Raharjo, T. J., Sugiyo, S., & Yulianto, A. (2020). Factual model of internal quality assurance system of private higher education institutions in indonesia. *Journal of Education, Teaching and Learning,* 5(1), 46-52

Oudshoorn, M., Raj, R. K., Thomas, S., & Parrish, A. (2018, June). *The value of abet accreditation to computing programs.* In 2018 ASEE Annual Conference & Exposition. University of Pittsburgh. Utah, USA.

Özyeğin Üniversitesi (2024). *Akreditasyon ve kalite güvencesi*. Retrieved August 18, 2024 from https://www.ozyegin.edu.tr/tr/isletme-fakultesi/akreditasyon-ve-kalite-guvencesi

Patton, M. Q. (1987). *How to use qualitative methods in evaluation.* (2nd Ed.) Sage Publications

Patton, M. Q. (2015). *Qualitative research and evaluation methods* (4th Ed.). Sage Publications.

Prendergast, J., Saleh, M., Lynch, K., & Murphy, J. (2001). *A revolutionary style at third level education towards TQM. Journal of Materials Processing Technology, 118*(1-3), 362-367.

Rhodes, T. L. (2012). Show me the learning: Value, accreditation, and the quality of the degree. *Planning for Higher Education, 40*(3), 36-42.

Roller, R. H., Andrews, B. K., & Bovee, S. L. (2003). Specialized accreditation of business schools: A comparison of alternative costs, benefits, and motivations. *Journal of Education for Business, 78*(4), 197-204.

Shaiban, A. S. (2024). Perceived stress among staff in Saudi Arabian dental colleges before and after an accreditation process: A cross-sectional study. *World Journal of Clinical Cases, 12*(4), 758. doi: 10.12998/wjcc.v12.i4.758

Sahney, S., Banwet, D. K., & Karunes, S. (2004). Conceptualizing total quality management in higher education. *The TQM Magazine, 16*(2), 145-159.

Sakarya Üniversitesi (2024). AACSB akreditasyonu resmi sonuç. Retrieved August 18, 2024 from https://aacsb.sakarya.edu.tr/tr/duyuru/goster/116660/aacsb-akreditasyonu-resmi-sonuc

Sanyal, B. C., & Martin, M. (2007). *Quality assurance and the role of accreditation: An overview.* Retrieved July 24, 2024 from http://hdl.handle.net/2099/8095

Seeman, E. D., & O'Hara, M. (2006). Customer relationship management in higher education: using information systems to improve the student-school relationship. *Campus-wide Information Systems, 23*(1), 24-34.

Semercı, Ç., Semerci, N., Ünal, F., Yılmaz, E., Kaygın, H., Ulus, İ. Ç., ... & Yılmaz, Ö. (2021). Akademisyenlerin akreditasyon algılarının incelenmesi. *Asian Journal of Instruction (E-AJI)*, 9(1), 1-14.

Sosyal Beşeri Temel Bilimler Akreditasyon ve Rating Derneği (2024). *Hakkımızda*. Retrieved August 18, 2024 from https://www.star.org.tr/kurumsal/hakkimizda

Tantekin Ersolmaz Ş (2018). MÜDEK: Yükseköğretimde İlk Ulusal Akreditasyon Kuruluşu, *İTÜ Vakfı Dergisi*, 79, 76-79.

Thomas, A., Antony, J., Haven-Tang, C., Francis, M., & Fisher, R. (2017). Implementing Lean Six Sigma into curriculum design and delivery–a case study in higher education. *International Journal of Productivity and Performance Management*, 66(5), 577-597.

van Berkel, H., & Wijnen, W. (2010). Accreditation in the Netherlands: does accountability improve educational quality?. *Research in Comparative and International Education, 5*(1), 88-97.

Van Vught, F. A., & Westerheijden, D. F. (1994). Towards a general model of quality assessment in higher education. *Higher Education, 28*(3), 355-371.

Welzant, H., Schindler, L., Puls-Elvidge, S., & Crawford, L. (2011). Definitions of quality in higher education: A synthesis of the literature. *Higher learning research communications*, 5(3), 1-11.

Wilson-Hail, C. K., Hurst, B., Chang, C. W., & Cooper, W. (2019). Accreditation in education: One institution's examination of faculty perceptions. *Critical Questions in Education*, 10(1), 17-28.

Yıldırım, A., Şimşek, H. (2018). *Sosyal bilimlerde nitel araştırma yöntemleri (10. Baskı)*. Ankara: Seçkin Yayıncılık

Yin, R. (2003). *Case study research: design and methods.* (3rd Ed.). Sage Publications.

Yükseköğretim Kalite Kurulu (2024). *Tescil süresi devam edenler.* Retrieved August 24, 2024 from https://yokak.gov.tr/akreditasyon-kuruluslari/tescil-suresi-devam-edenler

Zayachuk, Y., & Yamelynets, M. (2021). Accreditation procedures as a factor of quality assurance of higher education in Germany. *Pedagogika. Studia i Rozprawy, 30*(30), 179-189

Zohrabi, M. (2013). Mixed method research: Instruments, validity, reliability and reporting findings. *Theory and Practice in Language Studies, 3*(2), 254-262.

How cite this article

Cengiz Tırpan, E., Coşkun, E., & Sütütemiz, N. (2024). A Qualitative Study on Management Information Systems Faculty Members' Perceptions of Accreditation in Türkiye. *Acta Infologica, 8*(2), 222-244. https://doi.org/10.26650/acin.1540603

İSTANBUL UNIVERSITY PRESS

Acta Infologica
acin 2024, 8 (2), 245–259
DOI: 10.26650/acin.1543650

Research Article

House Value Estimation using Different Regression Machine Learning Techniques

Tarek Ghamrawi[1], Müesser Nat[2]

[1]Cyprus International University, School of Applied Sciences, Haspolat, Lefkosa, Mersin 10 Turkey

Corresponding author : Müesser NAT
E-mail : mnat@ciu.edu.tr

ABSTRACT

This study investigates the effectiveness of various regression algorithms in estimating house values using a dataset sourced from Zillow.com, encompassing 15,000 residential properties from Denver, Colorado. Comparisons of different models such as linear regression, Ridge regression, Lasso regression, Elastic Net, Decision Tree, Random Forest, Gradient Boosting, and XGBoost. The models were evaluated using R-squared (R^2) and Mean Absolute Error (MAE) as performance metrics. The results demonstrated that the Random Forest Regressor and XGB Regressor outperformed other models, achieving the highest R^2 scores and the lowest MAE values. These findings underscore the potential of these models for accurate house price estimation, which can be instrumental for the real estate market. Accurate valuations can help prevent overpricing, which causes properties to remain unsold for extended periods, and under-pricing, leading to financial losses. Implementing these regression models can enhance pricing strategies, ensuring efficient buying and selling processes and contributing to the overall financial health of the real estate market. Future research will explore the use of a broader range of regression models with fewer features to assess their performance and robustness in house price prediction.

Keywords: "House Price Estimation", "Machine Learning", "ElasticNet", "Lasso Regression", "Decision Tree Regressor", "Random Forest Regressor", "Linear Regression", "Ridge Regressor", "Gradient Boosting Regressor", "XGB Regressor"

Submitted : 05.09.2024
Revision Requested : 09.12.2024
Last Revision Received : 16.12.2024
Accepted : 17.12.2024

1. INTRODUCTION

Determining how much a house is worth, also known as home valuation, involves figuring out its market value in a fair and objective way. It's more than just checking the price tag; it uses different methods to get a good estimate. It's super important for pretty much every house deal. Sellers need accurate valuations to set reasonable prices, buyers want to make sure they're not paying too much, and agents need the right info to sell properties well, banks also need to know a house's value to decide how much of a risk it is to lend money for it. The value affects how big the loan is and what the interest rate will be. Insurance, companies, as well, use valuations to figure out how much coverage a homeowner needs. Getting it right means homeowners are protected financially if something bad happens to their house (Binu, 2020).

Accuracy matters because if a house is priced too high, it might sit on the market for ages, and if it's too low, the seller could lose out financially. Accurate valuations keep the real estate market healthy and stop people from getting ripped off, it is the basis for a well profound decision-making process (Thomas, 2023).

Machine learning (ML) offers exciting possibilities to overcome traditional house valuation methods limitations and enhance house value estimation accuracy by using the predictive performances of Linear Regression, Ridge Regression, Lasso regression, Decision Tree regressor, Random Forest regressor, ElasticNet, Gradient Boost Regression, and XG Boost Regression as regressor analytics

2. LITERATURE REVIEW

Multiple studies have used regressor methods on different data sets and a literature review was conducted on the following:

Wang (2018) proposed the utilization of Random Forests, a machine-learning approach, for developing a house price estimation model. The primary objective was to compare its performance against a benchmark linear regression model. The findings revealed that the Random Forest model excelled in capturing hidden nonlinear relationships between house prices and their features, resulting in more accurate estimations than the linear regression model. In this study, 27649 data points were used for the house assessment price of 2015 in Arlington country, Virginia USA

Similarly, Li (2023) focused on predicting house prices in King County, Washington, employing four machine learning models: linear regression, Random Forest (RF), Artificial Neural Network (ANN), and XGBoost. The aim was to identify key features influencing house prices and provide insights for future investments. Their results indicated that XGBoost achieved the highest accuracy among the models tested. Additionally, the study identified grade, square footage of living space, and latitude as the most influential features affecting house prices using 21,611 pieces of data and 21 features.

In a comparative study (Rana, 2020), various regression techniques were evaluated for their effectiveness in house price prediction. Support vector regression, XGBoost, Decision tree regression, and Random Forest regression. The results demonstrated that advanced regression techniques provided more accurate predictions compared to traditional linear models. The study underscored the importance of selecting relevant features to enhance model performance using 13,320 data points and 9 features.

Furthermore, Maida (2022) applied the Xtreme Gradient Boosting Model (XGBoost) to predict house prices in Karachi, Pakistan, using data from an open real estate portal. The XGBoost model achieved a remarkable 98% accuracy, highlighting its efficiency and flexibility in house price prediction. This study emphasized the model's superior performance in comparison to other models utilized for similar purposes using 38,961 data points and 14 features.

In another study by Truong et al. (2020) explored the performance of advanced machine learning models, including Random Forest, XGBoost, LightGBM, and ensemble techniques like Hybrid Regression and Stacked Generalization. Using a dataset of 231,962 housing records and 19 features from Beijing, they reported that Stacked Generalization achieved the best test set performance, leveraging multiple model outputs for robust predictions. The study underscores the effectiveness of combining machine learning models to enhance accuracy in house price estimation even though in this study the R2 was addressed as high or decent accuracy instead of providing real values but a root mean square logarithmic error was provided (RMSLE).

Additionally, Hernes et al. (2024) analysed the primary residential real estate market in Wroclaw, Poland, using a dataset of 15,000 records collected via a web scraping approach. The study implemented multiple machine learning models, including Gradient Boosting Regression, Random Forest Regression, LASSO, Multiple Linear Regression, and Simple Linear Regression, to predict housing prices based on attributes like area, number of rooms, year of construction, and more. Gradient Boosting Regression achieved the highest performance with an R^2 of 0.989, followed

closely by Random Forest Regression with an R^2 of 0.986. The study highlighted the utility of machine learning models for achieving high prediction accuracy and their potential application for real estate investors and institutions through a developed web-based prediction tool.

On the other hand, in this study Ali Soltani et al. (2022) analysed 428,000 residential transactions over a 32-year period (1984–2016) in Metropolitan Adelaide, Australia, using 38 explanatory variables and additional spatiotemporal lag features. The performance of four machine learning models—Linear Regression, Decision Tree, Random Forest, and Gradient-Boosted Tree—was evaluated. The Gradient-Boosted Tree model delivered the best results, achieving an R^2 of 0.896 and an RMSE of 0.086. This research underscored the significance of incorporating spatiotemporal dependency into machine learning models to enhance predictive accuracy in house price estimation.

3. AIM

The foundation of artificial intelligence is data. Models cannot be trained without sufficient data, which means pricey and sophisticated technology is left idle. The information found in data is what the models use to identify trends, glean insights, make predictions, and grow into more sophisticated models.

Although gathering reliable, and high-quality data can be an expensive and time-consuming procedure because there is a science to it. Certain types of data are extremely controlled, requiring lengthy lead times to obtain access and authorization Additionally the size of the data might be so small, that even when secured, training models might not end up finding it useful (Li, 2023). This research aims to demonstrate that, with a relatively small dataset (approximately 11,078 data points) and a limited set of carefully chosen features (7) from zillow.com for the state of Denver, machine learning models can achieve highly accurate house price estimations. This is in contrast to existing research, which often utilizes significantly larger datasets (average of 25,385.25 data points) and a broader range of features (average of 13).

In contrast to the methodologies adopted in the aforementioned studies, the current research focuses on employing a broader range of regression models with a significantly reduced feature set. This strategy aims to investigate whether a smaller set of features, which usually leads to underfitting (IBM, 2024), can still produce high-performing scores, thereby simplifying the model-building process and potentially enhancing interpretability without sacrificing accuracy.

4. METHODOLOGY

A. Data preperations

In this project, the dataset is gathered straight from Zillow.com, a trusted hub for real estate information. It was a dataset packed with details about residential properties, totalling 15,000 entries from the city of Denver in Colorado. To dig into what makes house prices tick,

One big task when preparing the data was handling missing data. missing values were not replaced with the mean, so a trick called "dropna()." Basically, it's a way to drop the blanks. It has been done to keep things legit and make sure the data stays reliable and after this dropping only 11,078 data points were left with a minimum value of $147,767 and a maximum value of $1,600,000.

B. Feature Engineering

In this study the following numerical features were selected which are bedrooms, bathrooms, rooms, square footage, lot size, year built, and last sale amount for this study. These features directly influence house prices by representing key aspects of a property's size, condition, and historical transaction values. Non-numerical features like id and address, longitude, altitude, and prior sale Date, were excluded as they do not directly impact house prices,

C. Technology and tools

For this study, JupyterNotebook was used, an open-source web application, as the primary technology platform. JupyterNotebook provides an interactive environment for conducting data analysis, exploratory research, and model development. Its flexibility and ease of use allowed the study to efficiently explore the dataset, perform feature engineering, and train predictive models. The dataset was first divided into training, validation, and testing sets to ensure the robustness of the model and to prevent overfitting. The data was split as follows:

- Training Set: 70% of the data
- Test Set: 30% of the data

This splitting was done for all learning models. Additionally, to enhance the predictive performance of the regression models, polynomial features were applied engineering and normalization. The features were expanded to polynomial degrees 1, 2, and 3, capturing non-linear relationships in the data. After polynomial feature expansion, Standard Scaler

was applied to normalize the features, ensuring that they have a mean of zero and a standard deviation of one, which helps improve the convergence and performance of the regression models.

The models were assessed using the R^2 Test score to determine the proportion of variance in the dependent variable predictable from the independent variables in order to maintain transparency. This methodology is represented in the workflow figure 1 below (Géron, 2019)

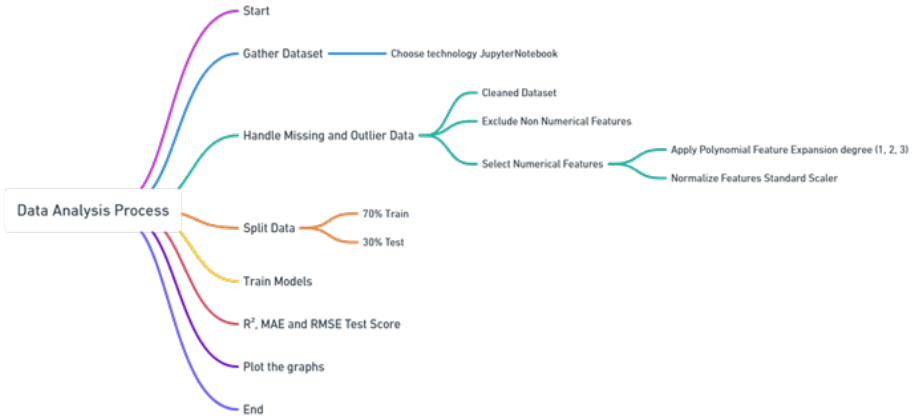


Figure 1. *Methodology Workflow (Géron, 2019)*

D. Model Selection and Training

Several regression techniques were used in this work to estimate house values; these strategies were selected based on their individual merits and contributions to the overall performance of the model. The chosen techniques consist of the following: Gradient Boosting Regressor, XGBoost Regressor, Decision Tree Regressor, Random Forest Regressor, Ridge Regression, Lasso Regression, Elastic Net Regression, and Linear Regression. The following factors led to the selection of these models:

Diverse Capabilities: The integration of linear and non-linear models guarantees the collection of straightforward and intricate interactions within the data. While more sophisticated models like Decision Trees and ensemble techniques successfully manage non-linear relationships, linear models like Linear Regression offer a baseline and findings that are easy to understand models (Montgomery, Peck, & Vining, 2021).

Regularization Techniques: The inclusion of Ridge, Lasso, and Elastic Net regressions stems from their capacity to provide regularization, which lessens overfitting and enhances the model's generalizability. These methods also help with feature selection, which improves the interpretability and performance of the model (Zou, 2005).

Ensemble Methods: Models like Random Forest and Gradient Boosting Regressor leverage the power of ensemble learning to improve prediction accuracy and robustness. These methods reduce variance and bias by combining multiple weak learners to form a strong predictive model (Breiman, 2001).

Advanced Optimization: XGBoost is incorporated for its advanced optimization techniques and efficiency in handling large datasets. Its flexibility in hyperparameter tuning allows for fine-tuning the model to achieve superior performance (Chen, 2016).

By using a diverse set of regression methods, the study ensures a comprehensive evaluation of the dataset. This approach allows for comparing the effectiveness of different algorithms in predicting house values, ultimately identifying the best-performing model.

1) Linear Regression

This study uses a linear regression model as a baseline to forecast house prices.

linear regression helps predict how a dependent variable changes as the independent variable changes. The equation for linear regression looks like this:

$$y = \alpha + \beta 1x1 + \beta 2x2 + \ldots + \beta nxn + \varepsilon.$$

Here, the betas represent the estimated parameters for the independent variables, y is the dependent variable. In this case, y represents the predicted house price, and xi represents the different features chosen.

The reason behind choosing linear regression is that it provides a straightforward approach to understanding the relationship between features and target variables. Also, it is useful for comparing the performance of more complex models (Montgomery, Peck, & Vining, 2021).

2) Ridge regression

Aso known as Tikhonov regularization, is a technique used to analyze multiple regression data that suffer from multicollinearity. When it occurs, least squares estimates are unbiased, but their variances are large, which may lead to overfitting. Ridge Regression adds a penalty to the regression coefficients, effectively shrinking them and reducing the model complexity.

$$\beta = (X^T X + \lambda I)^{(-1)}\{X^T\}\_y$$

where:

- X is the matrix of input features,
- y is the vector of target values,
- λ is the regularization parameter,
- I is the identity matrix.

The term λ added to *XTX*

The reason this method has been chosen it is because it ensures that the matrix is invertible, thus stabilizing the coefficient estimates to better generalizability on unseen data (Montgomery, Peck, & Vining, 2021).

3) Lasso regression

Least Absolute Shrinkage and Selection Operator is a type of linear regression that includes L1 regularization. This technique is particularly useful for feature selection and regularization, making it effective for models with many predictors, especially when some of those predictors are irrelevant or redundant.

The main idea behind Lasso Regression is to minimize the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant. The reason this method has been chosen is because it encourages sparsity in the model by shrinking some coefficients to zero, effectively selecting a simpler and more interpretable model.

$$min\beta((\| y - \{Xb \|\}\_2^2 + \lambda \| \beta \| \_1)$$

where:

- y is the target variable,
- X is the feature matrix,
- β is the coefficients,
- λ is the regularization parameter.

By adjusting the λ parameter, Lasso can control the strength of the regularization. A larger λ value increases regularization, resulting in more coefficients being shrunk to zero. This can lead to models that are simpler and potentially less prone to overfitting, especially in high-dimensional datasets.

Lasso Regression is powerful for predictive modelling because it not only prevents overfitting but also performs automatic feature selection, enhancing the interpretability and performance of the model (Hastie, Tibshirani, & Wainwright, 2015).

4) Elastic Net Regression

A regularized regression method that linearly combines the penalties of Lasso (L1) and Ridge (L2) regression techniques. It aims to address the limitations of both methods by balancing the trade-off between feature selection (Lasso) and coefficient shrinkage (Ridge). This makes Elastic Net particularly effective when dealing with datasets with highly correlated predictors or when the number of predictors exceeds the number of observations. The Elastic Net objective function is given by:

$$min\beta(\| y - X\beta \| \_2^2 + \lambda\_1 \| \beta \| \_1 + \lambda\_2 \| \beta \| \_2^2)$$

where:

- y is the target variable,
- X is the feature matrix,
- β are the coefficients,
- λ 1 is the L1 regularization parameter (Lasso)

- λ 2 is the L2 regularization parameter (Ridge).

By incorporating both L1 and L2 penalties, Elastic Net encourages a sparse model with fewer features (like Lasso) while maintaining regularized coefficients to prevent overfitting (like Ridge). This dual approach helps in improving model performance and interpretability, especially in high-dimensional datasets the reason it was chosen is because it is effective when dealing with datasets with highly correlated predictors or when the number of predictors exceeds the number of observations. (Hastie et al., 2015).

5) Decision tree regressor

Decision tree regression is a non-parametric technique that predicts continuous values by building a tree-like model. Each internal node represents a feature, and branches represent decisions based on feature values. The terminal nodes (leaves) hold predicted target values. This model is constructed through recursive partitioning:

Feature Selection: The algorithm selects the most informative feature (e.g., using information gain) to split the data at each node, aiming to maximize the difference in the target variable between resulting child nodes.

- Splitting: The chosen feature is used to create a split rule, separating data points.
- Recursion: This splitting continues recursively until a stopping criterion (e.g., minimum data points) is met

The reason it was chosen is because it provides a non-parametric approach that is easy to visualize and interpret. (James, Witten, Hastie, & Tibshirani, 2021).

6) Random Forests

Random forests as shown in the figure 2 leverage decision trees by building an ensemble of them. Each tree is trained on a random subset of data (bootstrapping) and utilizes a random subset of features at each split. This injects randomness to improve model generalization:

- Diversity: Random subsets create diverse trees capturing different data aspects.
- Aggregation: Predictions from all trees are aggregated (e.g., averaged) for the final prediction, reducing variance and improving robustness, the reason it was chosen is because it reduces variance and improves robustness by aggregating predictions from multiple trees. (Hastie, Tibshirani, & Friedman, 2021).

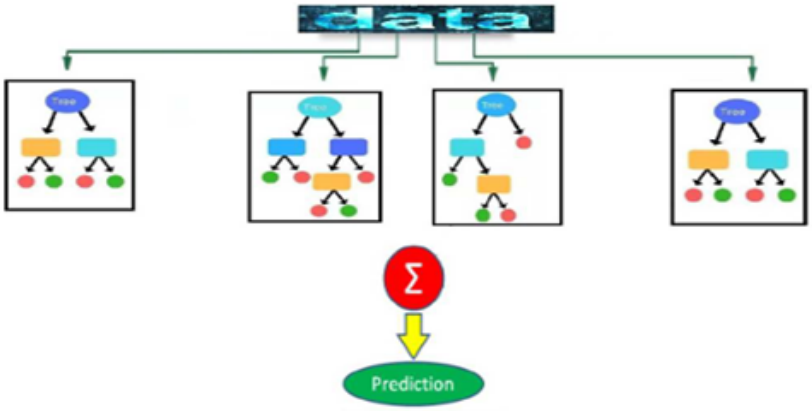


Figure 2. *Random Forest Model (Hastie, Tibshirani, & Friedman, 2021)*

7) Gradient Boost Regressor

It is an ensemble learning technique that builds a model in a stage-wise fashion from multiple weak learners, usually decision trees. It is built by iteratively fitting new models to the residual errors of the combined model built so far. It is powerful for predictive modelling because it combines the strengths of multiple models, reducing variance and bias. The main idea is to minimize the loss function by adding weak learners using a gradient descent-like procedure:

$$F\_m(x) = F\_(m - 1)(x) + h\_m(x)$$

where:

- *Fm(x)*: is the combined model after m iterations,
- *hm(x)*: is the weak learner added at the *m*-th iteration

(Friedman, 2001).

8) Extreme Gradient Boosting

XGBoost (Extreme Gradient Boosting) is a powerful machine learning algorithm belonging to the family of ensemble methods, specifically gradient boosting. It excels in both regression and classification tasks it builds an ensemble of decision trees, sequentially combining them to create a more robust and accurate model than any single

tree. Each tree learns from the errors (residuals) of the previous tree, focusing on improving the overall prediction accuracy. Formula:

$$f\_m(x) = f\_(m-1)(x) + \gamma * h\_m(x)$$

- *Fm(x)*: Prediction of the model at the m-th iteration (ensemble prediction) for a given input x.
- *Fm(m-1)(x)*: Prediction of the model at the previous (m-1) th iteration for input x.
- $\gamma(gamma)$: Learning rate (hyperparameter)
- *hm(x)*: Prediction of the m-th decision tree in the ensemble for input x.

XGBoost builds upon the core idea of gradient boosting by incorporating several advancements as represented in figure 3 below (Chen & Guestrin, 2016): Regularization techniques like L1 and L2 regularization that are employed to prevent overfitting, a common challenge where the model becomes too specific to the training data and performs poorly on unseen data. XGBoost leverages parallel processing and feature subsampling to handle large datasets efficiently, making it suitable for big data applications.

- Flexibility: A wide range of hyperparameters are available for tuning, enabling customization to optimize performance for specific tasks and data characteristics.
- Scalability: In essence, XGBoost combines the strengths of ensemble learning with regularization and efficient algorithms, making it a powerful tool for various machine-learning problems (Chen & Guestrin, 2016).

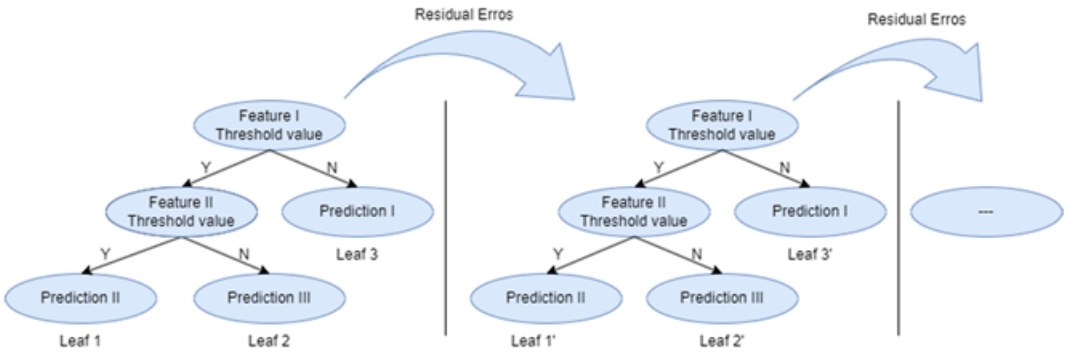


Figure 3. *XGB Model (Chen & Guestrin, 2016).*

5. RESULTS

In this study for an accurate house price prediction model, a comparison performance of various regression algorithms was conducted. a diverse range of models were applied, including:

Linear models: Elastic Net, Lasso, and Ridge regression are well-established techniques that provide interpretable results.

Decision tree and ensemble methods: Decision Tree Regressor and Random Forest Regressor offer flexibility in capturing complex relationships within the data.

Gradient boosting methods: Gradient Boosting Regressor and XGB Regressor are powerful ensemble methods known for their ability to handle non-linear relationships and complex datasets.

This study evaluated each model's performance on a separate test set using two key metrics:

R-squared (R^2): This metric measures the proportion of variance in the target variable (house prices) explained by the model's predictions. A higher R^2 signifies a better fit between the predicted and actual values.

Mean Absolute Error (MAE): This metric calculates the average absolute difference between the predicted and actual house prices. A lower MAE indicates a closer match between the predictions and the real values by analysing these metrics, the aim is to identify the model that delivers the most accurate and reliable house price predictions, ultimately assisting in informed decision-making.

Table 1. Test R^2 Scores

	Result Table			
	Model Name	***Test R²***	***Test MAE***	***Test RMSE***
1	ElasticNet	0.61	108,700.36	315735.74
2	Lasso	0.75	131,959.56	253678.56
3	DecisionTreeRegressor	0.74	114,648.88	256545.37
4	RandomForestRegressor	0.82	85,124.75	211055.52
5	LinearRegression	0.75	131,959.93	253678.55
6	Ridge	0.83	88,409.49	206209.01
7	GradientBoostingRegressor	0.82	83,888.93	218895.59
8	XGBRegressor	0.83	85,236.33	209536.04

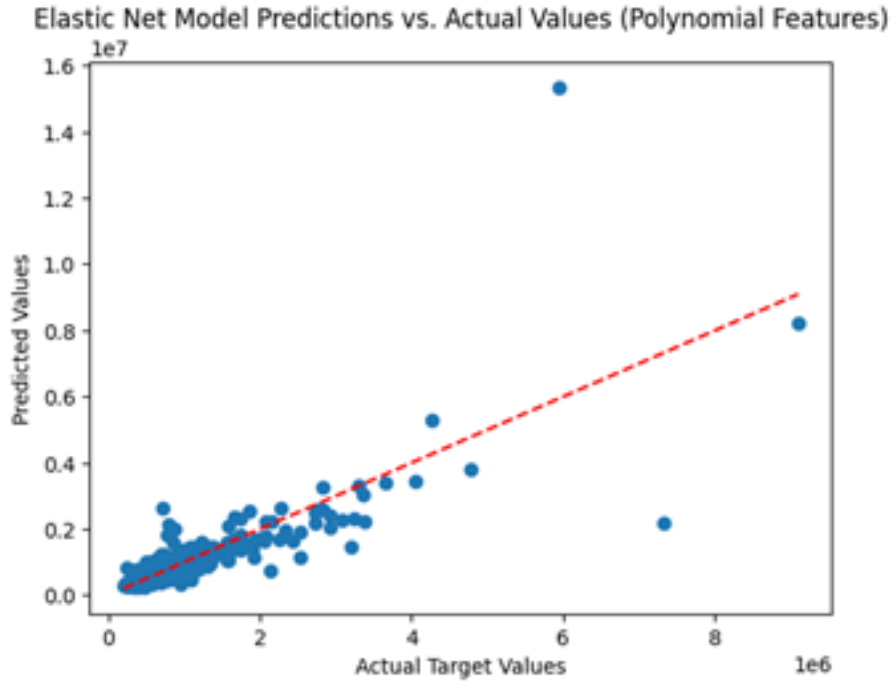


Figure 4. *Elastic Net*

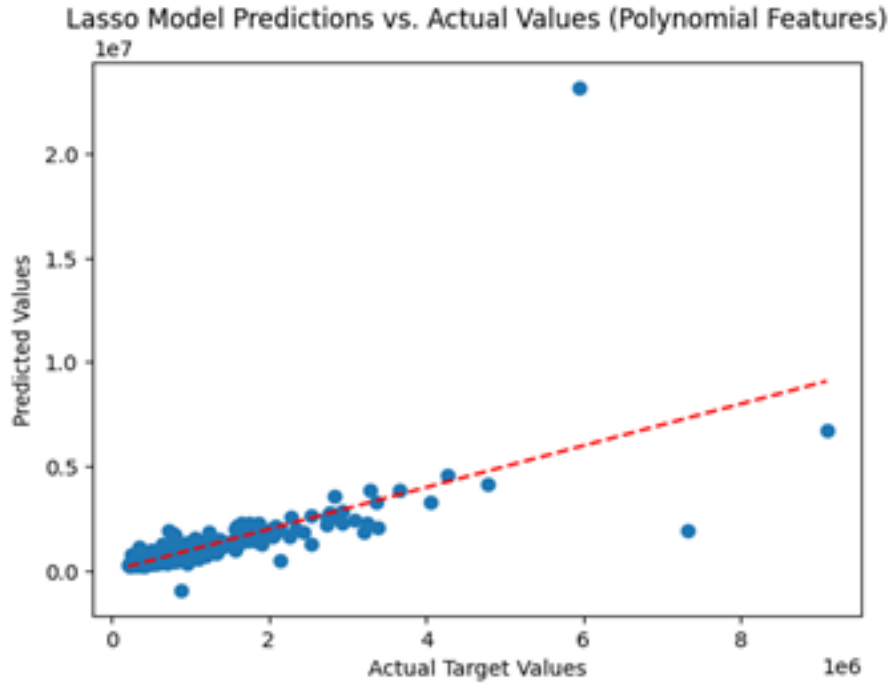


Figure 5. *Lasso Model*

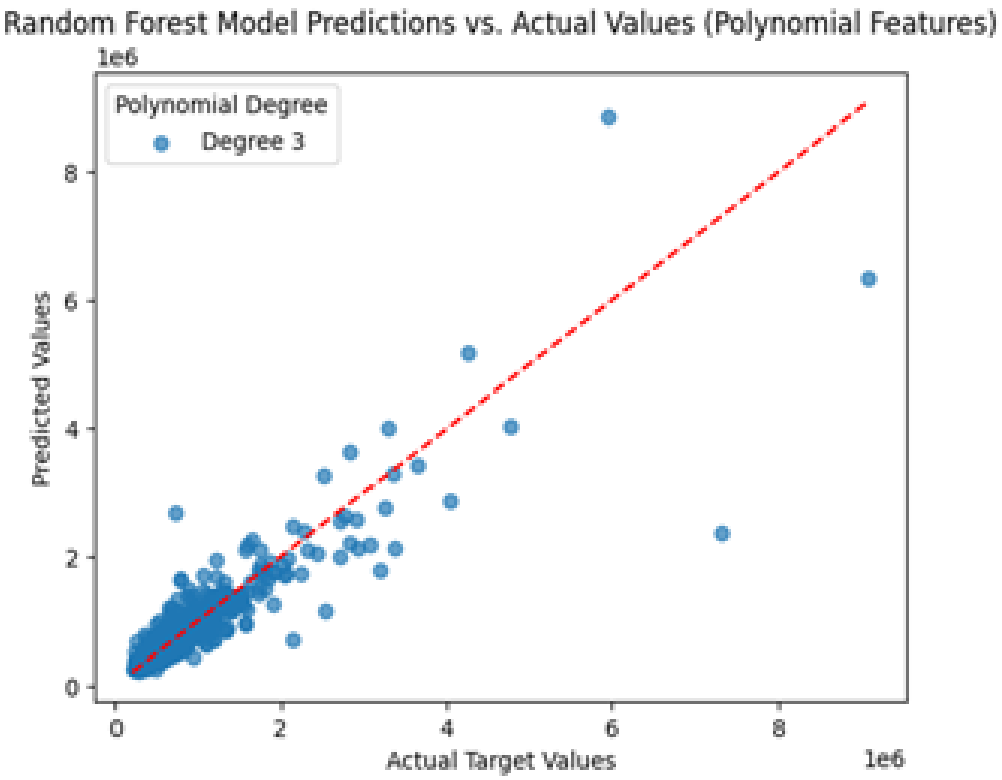


Figure 6. *Random Forest*

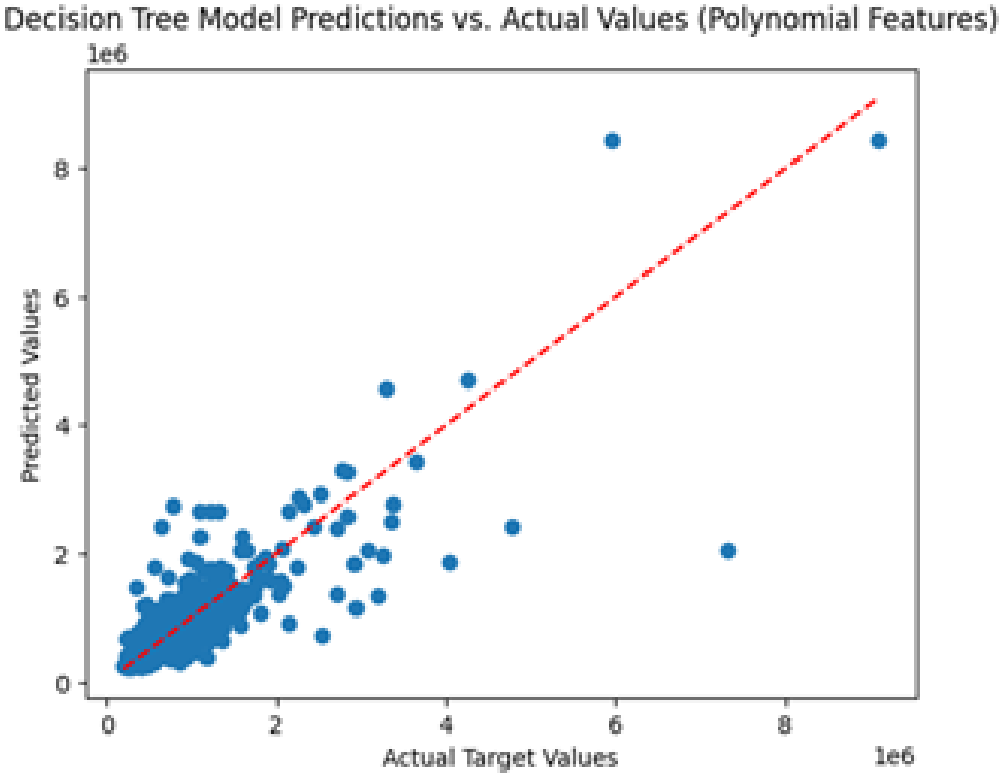


Figure 7. *Decision Tree*

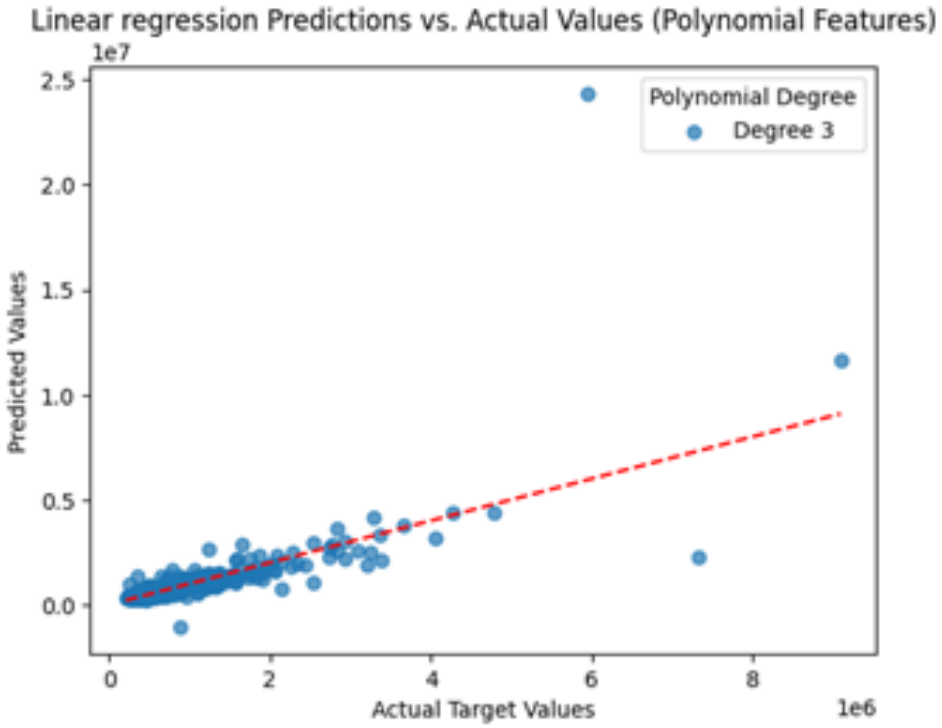


Figure 8. *Linear Regression*

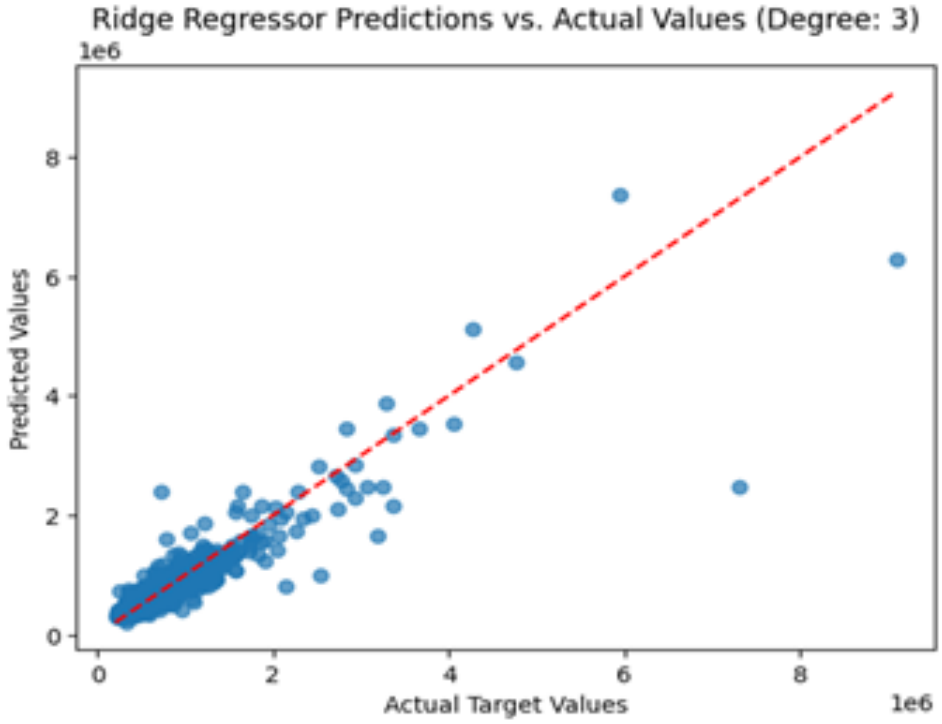


Figure 9. *Ridge Regressor*

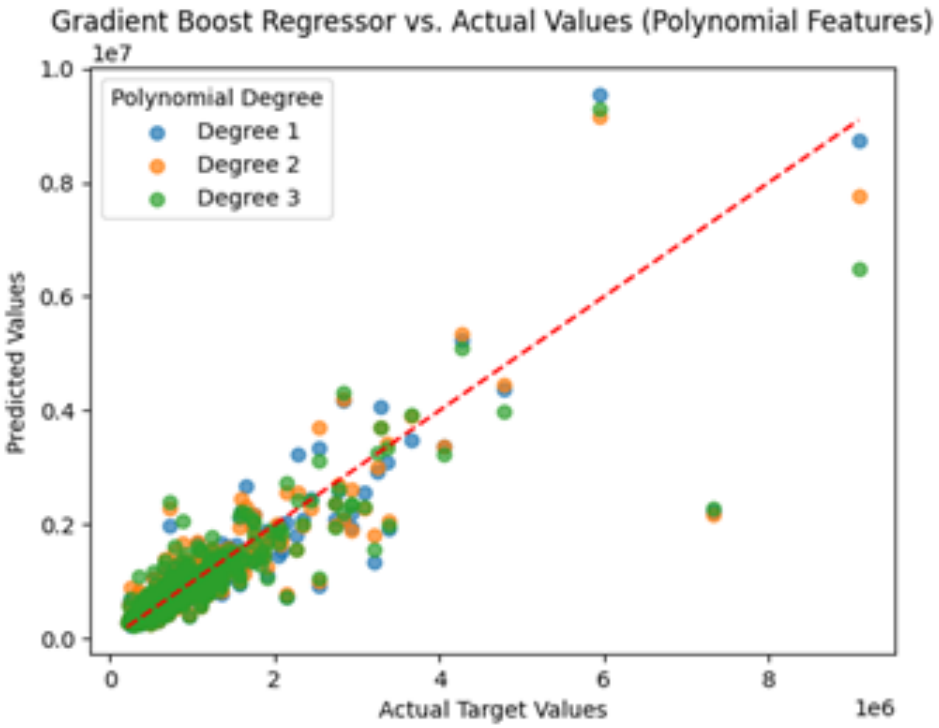


Figure 10. *Gradient Boost Regressor*

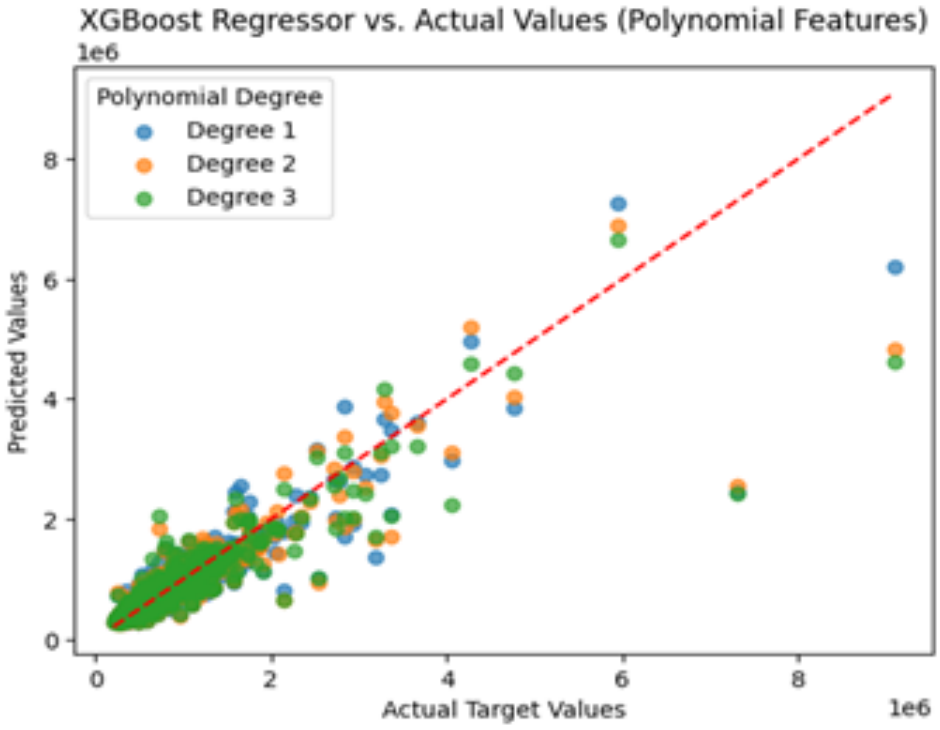


Figure 11. *XGBoost*

6. FINDINGS

In the quest to identify the most suitable model for house price prediction, an empirical evaluation was conducted comparing the performance of various regression algorithm as represented in the findings of this study represented in the result table 1 Test R^2 Scores. The candidate models included:

Linear models: ElasticNet, Lasso, and Ridge regression are established techniques known for their interpretability.

Decision tree and ensemble methods: Decision Tree Regressor and Random Forest Regressor offer flexibility in capturing complex relationships within the data.

Gradient boosting methods: Gradient Boosting Regressor and XGB Regressor are powerful ensemble methods known for their ability to handle non-linear relationships and complex datasets.

Separate test set for model evaluation, ensuring the models were not biased towards the training data. Two key metrics were used to assess performance:

R-squared (R^2): This metric measures the proportion of variance in the target variable (house prices) explained by the model's predictions. A higher R^2 signifies a better fit between the predicted and actual values.

Mean Absolute Error (MAE): This metric calculates the average absolute difference between the predicted and actual house prices. A lower MAE indicates a closer match between the predictions and the real values.

Random Forest Regressor and XGB Regressor achieved the highest R^2 scores (0.82 and 0.83, respectively), indicating a strong correlation between their predictions and the actual house prices.

Random Forest Regressor and XGB Regressor also achieved relatively low MAE values (85,124.75 and 83,888.93 respectively), suggesting a close match between the predicted and actual house prices in terms of absolute difference.

While Ridge regression achieved a comparable R^2 score to the Random Forest Regressor and XGB Regressor, its MAE value was slightly higher.

One notable aspect of this study is the relatively small number of features used compared to other literature. Despite the limited feature set, the models still produced highly accurate predictions. This demonstrates the efficiency and robustness of the Random Forest Regressor and XGB Regressor models, which can achieve strong predictive performance even with fewer input variables. This efficiency is particularly advantageous for real-world applications where gathering extensive data can be time-consuming and costly.

Figures 4 through 11 display the actual vs. predicted values for various models helping visually represent the accuracy of each model, they include, ElasticNet, Lasso Regression, Random Forest Regressor, Decision Tree Regressor, Linear Regression, Ridge Regressor, Gradient Boosting Regressor, and XGB Regressor. These figures were generated through machine learning techniques and visualized using Matplotlib in Jupyter Notebook.

A. Comparison and Analysis

Five studies were taken into consideration for comparison presented in the table 2 Comparison Table which show the Test R2 Score the Test MAE and the Test RMSE of each model used in each study the following studies taken into consideration are:

- Wang, 2018
- Chen Li, 2023
- Madhuri 2019
- Hernes et al., 2024
- Soltani et al., 2022

1) ***Strengths:***

a) XGBRegressor: In this study, the R^2 (0.83) is competitive, though slightly lower than the highest R^2 in the Chenxi Li study (0.888).

The XGBRegressor in this study has one of the highest R-squared values across all studies, indicating strong performance even though less data and features were used.

b) Ridge Regression:

The R^2 in this study (0.83) outperforms the Ridge Regression in the comparative study by CH. Raga Madhuri (0.732164).

The Ridge model in this study has a high R-squared value, showing its robustness.

c) Random Forest Regressor:

The R^2 in this study (0.82) is higher than in the Changchun Wang and Hui Wu study (0.701310346391) and close to the Chenxi Li study (0.878).

This indicates that Random Forest model is highly effective in this study as well.

d) Gradient Boosting Regressor:

The R^2 in this study (0.82) is strong and comparable to the Gradient Boosting Regression in CH. Raga Madhuri's study (0.9177022).

The Gradient Boosting model performs well, indicating its capability to handle non-linear relationships.

e) Gradient Boosting, Random Forest

Hernes et al. achieved 98% precision in predicting housing prices using Gradient Boosting, Random Forest. This precision reflects the use of ensemble methods that are well-suited to capturing non-linear relationships and ensuring robust performance.

f) Gradient Boosting

Soltani et al. incorporated a spatiotemporal lag variable to improve prediction accuracy, achieving an R^2 of 0.896 with Gradient Boosting. RMSE and MAE metrics for Soltani et al. showed improvements when the spatiotemporal lag variable was included, with MAE reduced to 0.058 and RMSE to 0.086 on the training set.

2) ***Weaknesses:***

a) ElasticNet:

The R^2 in this study (0.61) is lower compared to the Elastic Net Regression in CH. Raga Madhuri's study (0.665228).

This suggests that ElasticNet might not be the best fit for this dataset.

b) Linear Regression:

The R^2 in this study (0.75) is higher than the values in the other studies (0.539887986037 in Changchun Wang and Hui Wu, 0.706 in Chenxi Li, and 0.732072 in CH. Raga Madhuri).

However, it still doesn't match the best-performing models like XGBRegressor or Gradient Boosting but because of the feature selection and the standardization used it was able to get a better result than other studies.

c) Lasso Regression:

The R^2 in this study (0.75) is higher than in the CH. Raga Madhuri study (0.732072) but still lower compared to your top models like XGBRegressor and Ridge.

d) Decision Tree Regressor:

The R^2 in this study (0.73) is comparable but not superior to the other models. This suggests it may be less effective with less data and less feature selection compared to ensemble methods like Random Forest and Gradient Boosting.

e) Simple Linear Regression

Hernes et al. achieved 60.7% accuracy using their Simple Linear Regression technique while comparing to this study which had 75% and additionally, the study of Hernes et al. relied on web-scraped data for real-time updates, which provided dynamic insights but posed risks of instability due to changes in website structures. This study's reliance on pre-existing datasets offers stability and consistency. While Hernes et al. focused on a single market (Wroclaw primary real estate), this study's broader range of datasets and methods underscores its adaptability to diverse scenarios without losing predictive accuracy.

f) Decision Tree

Soltani et al.'s results concerning the decision tree were much less than this study with R2 of 0.579 compared to ours which was 0.73 even though it had relied on a 32-year dataset of 428,000 records that provided a historical depth that can also lead to legacy biases. This study focuses on fewer features and more recent data, achieving predictions that are both accurate and timely. Soltani et al.'s preprocessing efforts included handling extensive spatiotemporal dependencies, which increased computational demands. This study's streamlined approach reduces computational overhead while maintaining high accuracy, making it more practical for real-time applications.

This study's strengths lie in the performance of the XGBRegressor, Ridge Regression, and Gradient Boosting Regressor, which show high R-squared values, indicating strong predictive power. Random Forest Regressor also performs well but slightly below the top models in other studies. Weaknesses include ElasticNet and Linear Regression, which have lower R-squared values, indicating they might not capture the complexity of the data as effectively as other models.

The models in this study generally perform well, with certain models like XGBRegressor and Ridge standing out. Focusing on improving models like ElasticNet and exploring ensemble methods further could enhance the study's predictive accuracy.

The accuracy strength of this study lies in the polynomial features when they were created and normalized as part of feature engineering. The data was expanded to include polynomial degrees of 1, 2, and 3, capturing non-linear patterns. Following this expansion, the features were normalized using Standard Scaler to achieve a mean of zero and a standard deviation of one, enhancing the performance and convergence of the regression models.

3) Comparison Table

Table 2. Comparison Table

Study	*Algorithm*	*R^2 Score*	*MAE*	*RMSE*	*MSE*
Madhuri, 2019	Multiple Linear Regression	0.732072	-	48.88446	391,875,744
	Ridge Regression	0.732164	29.73141	51699	391,740,496
	LASSO Regression	0.732072	34.32263	46466	391,875,537
	Elastic Net Regression	0.665228	85.00798	76781	489,642,930
	AdaBoost Regression	0.7801099	79.94242	32161481	-
	Gradient Boosting	0.9177022	88.27804	109,713	179,336
Chen Li, 2023	Linear Regression	0.706	-	210,649.771	44,373,326,009.81
	Random Forest	0.878	-	136,170.257	18,542,338,996.45
	Neural Network	0.846	-	143,075.123	22,825,808,222.92
	XGBoost	0.888	-	130,281.363	16,973,233,719.78
Wang, 2018	Random Forests	0.701310	-	352.065	-
	Linear Regression	0.539887	-	381.280	-
Soltani et al., 2022	Decision Tree (ST Lag)	0.579	0.106	0.135	-
	Decision Tree (+ST Lag)	0.797	0.064	0.094	-
	Gradient-Boosted Tree (ST)	0.674	0.084	0.101	-
	Gradient-Boosted (+ST Lag)	0.896	0.058	0.086	-
	Random Forest (ST Lag)	0.662	0.086	0.109	-
	Random Forest (+ST Lag)	0.875	0.059	0.087	-
Hernes et al., 2024	Simple Linear Regression	0.607	-	-	101,789.11
	Gradient Boosting	0.989	-	-	5,173.12
	Multiple Linear Regression	0.912	-	-	41,404.69
	LASSO	0.912	-	-	41,405.14
	Random Forest	0.986	-	-	5,493.95
This Study	ElasticNet	0.61	108,700.36	315735.74	
	Lasso	0.75	131,959.56	253678.56	
	DecisionTreeRegressor	0.74	114,648.88	256545.37	
	RandomForestRegressor	0.82	85124.75	211055.52	
	Linear Regression	0.75	131,959.93	253678.55	
	Ridge	0.83	88,409.49	206209.01	
	GradientBoostingRegressor	0.82	83888.93	218895.59	
	XGBRegressor	0.83	85,236.33	209536.04	

7. CONCLUSION

This study evaluated the effectiveness of various regression algorithms for estimating house values, comparing models such as linear regression, decision trees, ensemble methods, and gradient-boosting techniques. The models were assessed on a separate test set using R-squared (R^2) and Mean Absolute Error (MAE) as performance metrics.

The results showed that the Random Forest Regressor and XGB Regressor models achieved the highest R^2 scores and relatively low MAE values, indicating a strong correlation between their predictions and actual house values with minimal absolute difference. While models like Ridge regression also performed well, Random Forest Regressor and XGB Regressor emerged as the top performers.

The choice between Random Forest Regressor and XGB Regressor should be guided by the specific needs of the application:

Interpretability: If model transparency and the ability to explain predictions to clients are important, Random Forest Regressor is preferable due to its interpretable decision tree structure.

Absolute Error Sensitivity: If minimizing the absolute difference between predicted and actual house values is crucial, the XGB Regressor is the better choice due to its slightly lower MAE.

The application of these regression models is particularly valuable for evaluating house prices in small estate markets such as North Cyprus. Accurate and reliable house value estimations can significantly enhance pricing strategies, preventing overpricing that can lead to properties remaining unsold for extended periods and avoiding under-pricing that results in financial losses. By providing precise value estimates, these models can aid real estate professionals in setting competitive listing prices, ensuring properties are bought and sold efficiently, and maintaining the financial health of the real estate market.

Real estate professionals benefit from enhanced pricing strategies and more informed client consultations, which contribute to greater market efficiency. Investors gain clearer insights for better decision-making through reliable value assessments. Meanwhile, accurate property valuations empower buyers, sellers, and homeowners to make well-informed choices regarding transactions and refinancing...

8. FUTURE WORK

These models have demonstrated the ability to anticipate house estimation values with great accuracy, thus future research in real estate markets will incorporate these models, and North Cyprus is one such market where such

research will be carried out. Future research could further enhance these models by including additional features like neighbourhood amenities, school quality, and property condition to capture a more comprehensive range of factors influencing house values and refine-tuning the hyperparameters for each algorithm to improve model performance. Applying the models to different regions to understand price variations across locations continuing to refine these models will contribute to the development of robust and precise house value estimation, ultimately leading to better decision-making in the housing market.

Peer Review: Externally peer-reviewed.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

ORCID IDs of the authors

Tarek Ghamrawi 0009-0008-9107-7375
Müesser Nat 0000-0002-1539-3586

REFERENCES

Ahtesham, M., Bawany, N., & Fatima, K. (2020). *House price prediction using machine learning algorithm - The case of Karachi City. Pakistan*, 1-5. doi:10.1109/ACIT50332.2020.9300074

Ali, S., Mohammad, H., Fatemeh, A., Christopher, J. P. (2022). Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms. *Cities*, 131, 103941. https://doi.org/10.1016/j.cities.2022.103941.

Binu, J. (2020). *A data analytics model for extended real estate comparative market analysis.* Pace University, New York.

Breiman, L. (2001). *Random forests.* Machine Learning, 45, 5-32. doi:10.1023/A:1010950718922

Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system.* Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794. doi:10.1145/2939672.2939785

Friedman, J. H. (2001). *Greedy function approximation: A gradient boosting machine.* IMS 1999 Reitz Lecture. Modified March 15, 2000, April 19, 2001.

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: The lasso and generalizations.* CRC Press.

IBM. What is underfitting? Retrieved from https://www.ibm.com/cloud/learn/underfitting

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R (2nd ed.)*. Springer.

Li, C. (2023). *House price prediction using machine learning*. Proceedings of the 4th International Conference on Signal Processing and Machine Learning. School of International Education, GuangDong University of Technology, Guangzhou, China. doi:10.54254/2755-2721/53/20241426

Li, D. (2020, December 25). *Overcoming data scarcity and privacy challenges with synthetic data.* InfoQ. Retrieved from https://www.infoq.com/articles/overcoming-privacy-challenges-synthetic-data/#idp_register/

Madhuri, C. R., Anuradha, G., & Pujitha, M. V. (2019). *House price prediction using regression techniques: A comparative study.* 2019 International Conference on Smart Structures and Systems (ICSSS), Chennai, India, pp. 1-5. doi:10.1109/ICSSS.2019.8882834

Marcin, H., Piotr, T., & Mateusz, S. (2024). Prediction of residential real estate price on primary market using machine learning. *Procedia Computer Science*, 246, 3142-3147. https://doi.org/10.1016/j.procs.2024.09.358.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis (6th ed.)*. Wiley.

Quang, T., Minh, N., Hy, D., & Bo, M. (2020). Housing Price Prediction via Improved Machine Learning Techniques. *Procedia Computer Science,* 174, 433-442. https://doi.org/10.1016/j.procs.2020.06.111.

Rana, V. S., Mondal, J., Sharma, A., & Kashyap, I. (2020). *House price prediction using optimal regression techniques*. 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, pp. 203-208. doi:10.1109/ICACCCN51052.2020.9362864

Thomas, D. (2023). *The importance of data in property valuation and the key role of comparative method.* doi:10.13140/RG.2.2.35313.86881

Wang, C., & Wu, H. (2018). A new machine learning approach to house price estimation. *New Trends in Mathematical Sciences*, 6(4).

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301-320. doi:10.1111/j.1467-9868.2005.00503.

How cite this article

Ghamrawi, T., & Nat, M. (2024). House Value Estimation using Different Regression Machine Learning Techniques. *Acta Infologica, 8*(2), 245-259. https://doi.org/10.26650/acin.1543650

İSTANBUL UNIVERSITY PRESS

Acta Infologica
acin 2024, 8 (2), 260–273
DOI: 10.26650/acin.1567382

Research Article

A New Approach to Grouping Learners Based on Behavioral Engagement in CSCL Environments

Souhila Zerdoudi[1], Houda Tadjer[1], Yacine Lafifi[1], Zohra Mehenaoui[1]

[1]Computer Science Department, LabSTIC Laboratory, University 8 Mai 1945, Guelma, Algeria

Corresponding author : Yacine Lafifi
E-mail : lafifi.yacine@univ-guelma.dz

ABSTRACT

In Computer-Supported Collaborative Learning (CSCL) environments, forming a group is essential for the success of the learning process. Furthermore, several studies on forming groups in CSCL environments have been conducted recently to form ones that promote learners' engagement and collaboration among their members. Forming group-based approaches requires data on learners' actions (or traces) during the learning process. In this study, behavioral traces of learners are used to form groups. In other words, we used a clustering algorithm based on learners' behavioral engagement to form homogeneous groups of learners. The learners must have different levels of engagement within each group to enhance their engagement and cognitive levels. The basis of the proposed grouping algorithm is a set of indicators of learners' engagement. Furthermore, the proposed approach is based on an artificial intelligence algorithm, the k-means clustering method, which is used to find the maximum possibilities for the best clusters. Then, another algorithm is applied to obtain groups of learners with different levels of behavioral engagement. The validation of the proposed approach on a dataset containing behavioral traces from 100 learners was encouraging and promoting.

Keywords: CSCL, Group formation, K-means, Learning analytics, Learner engagement

Submitted : 15.10.2024
Revision Requested : 13.12.2024
Last Revision Received : 17.12.2024
Accepted : 23.12.2024

1. INTRODUCTION

Numerous studies have been conducted on strategies and tactics for grouping learners. The criteria for group formation were cognitive profile, learning style, and soft skills. The aim is to obtain homogeneous or heterogeneous groups according to the objective of the grouping process. All actions performed by learners are saved and input to grouping algorithms. Thus, what standards and components should be applied to determine the best method for assigning students to groups in CSCL (Computer-Supported Collaborative Learning) environments?

Our challenge is to combine learning groups with practical and academic skills. We must determine each learner's behavioral engagement level to build a group of learners. First, behavioral engagement refers to learners' participation in extracurricular activities, attendance, and disciplinary actions are all critical indicators of behavioral engagement (O'Donnell & Reschly, 2020). Fredricks et al. (2004) claimed that behavioral engagement involves student behavior and participation in educational and school-related activities. According to (Reschly et al., 2017), this term encompasses good classroom conduct and engagement in extracurricular activities.

In general, learners' engagement is characterized by a qualitative value, such as being very engaged or quietly engaging. Therefore, measuring this skill is a challenge in online learning environments that support collaboration (like CSCL environments). Traces left by learners during their learning process can be used as data to calculate the value of the learner's engagement degree. However, many research questions can be addressed: How can the degree of engagement of learners be calculated, and what indicators are used to calculate this value? Then, after determining this value, how do we group learners into groups with learners having different engagement values? The aim is to obtain groups of learners who can benefit from each other on one side and all the obtained groups that are homogeneous on the other side. In other words, the objective is to form homogenous groups with heterogeneous learners within the groups.

This study simulates a learner grouping system based on behavioral indicators to place students into diverse learning groups.

This paper comprises five sections and is organized as follows. Section 2 discusses related works on learners' engagement and grouping. The third section presents the proposed approach for grouping learners based on behavioral engagement. The simulation study will be presented and discussed in section four. Finally, the last section of this paper presents the conclusions of this study and future research directions.

2. RELATED WORKS

Computer Supported Collaborative Learning (CSCL) aims to maximize student learning achievement by promoting student interaction and collaboration through the use of technology in collaborative learning environments (Stahl et al., 2006; Scheuer et al., 2010). To optimize the collaboration in CSCL contexts, the authors in (Kirschner et al., 2002) and (Isotani et al., 2013) outline certain factors that need to be considered to gain a deeper understanding of (a) How to assign students to groups and (b) How to enhance student participation and communication in group work.

Some researchers have studied the tools and techniques used to improve learners' engagement, such as the "Students' Engagement Scale" tool. Sun and Rueda (2012) developed the student engagement scale to assess student engagement in online learning environments. Furthermore, many techniques are used to improve learners' engagement. Learning analytics are among them. Karaoglan et al. (2022) adopted personalized meta-cognitive feedback support based on learning analytics in online learning for learners' recommendations and guidance to improve student engagement.

Some researchers have developed a Social Learning Analytics "SLA toolkit, " which combines social network analysis with lexical analysis to produce information on student forum participation. Hence, the toolbox promotes the behavioral engagement of learners (Chen et al., 2018; Ouyang et al., 2021). Additionally, visualization of the topic network enhances all students' perspective expressions, indicating that demonstrating students' interest in topics can increase cognitive engagement in terms of students' levels of knowledge sharing, construction, and creation (Ouyang et al., 2021).

Learner engagement can be categorized into several types: cognitive, emotional, social, behavioral, etc. The literature reports several ways to measure students' cognitive engagement in classroom and online learning environments. The Online Student Engagement (OSE) Questionnaire is one such instrument. The questions in the Feedback activity were primarily open-ended and aligned with the OSE questionnaire and the "Sloan" instrument for measuring student satisfaction in online courses (Rajabalee & Santally, 2021). Other researchers have used different concepts and technologies for curriculum design, teaching methods, assessments, and the range of academic support that should be included in open-access-enabling courses (Atherton et al., 2017).

Collaboration among learners can improve their engagement level. Furthermore, learning in groups can improve the outcomes of learners and their profiles. In the education field, numerous approaches have been proposed to address

the issue of group formation. Bekele (2005) stated that the formation of learning groups is centered around three main points:

- Group size based on learning objectives
- Dividing learners into groups
- Heterogeneity within groups.

While group homogeneity is a requirement that guarantees the group's productivity (Anzieu & Martin, 1971), it is a criterion that ensures the diversity of the members' viewpoints, ideas, and personalities (Bekele, 2005).

The formation group process can be performed manually or automatically (Matazi et al., 2014). Forming a group manually involves either self-selection or instructor selection (Resta & Laferrière, 2007; Srba & Bielikova, 2014; Ounnas et al., 2007). In the self-selection approach, members have the right to choose the most appropriate group. The second approach is managed by the instructor, who decides which members will comprise the group (Abnar et al., 2012; Srba & Bielikova, 2014). This type of selection guarantees better results in balanced grouping; however, it is a reasonably complex process when many members are grouped manually (Srba & Bielikova, 2014; Mujkanovic et al., 2012). To form groups, CSCL environments can automatically create groups with or without human intervention (Abnar et al., 2012). Random selection is a way to create groups automatically (Srba & Bielikova, 2014). Other approaches form groups on the basis of the context (Maqtary et al., 2019).

The collaborative learning research community has presented numerous algorithmic approaches to address this difficulty. Probabilistic algorithms, clustering, semantic web, anthologies, and other techniques are examples of such methods (Bouyzem et al., 2021; Combaudon, 2018). Due to its ability to handle several variables and its speed in generating optimal solutions, researchers have recently employed genetic algorithms to execute cluster compositions in CSCL systems (Da Rocha, 2019). However, the number of such studies has been limited. Darwin's theory of evolution provides the foundation for the meta-heuristic theory known as the genetic algorithm (Zheng et al., 2018).

Cole and co-authors (Cole et al., 2021) proposed a machine-learning technique that uses a positive label to predict numerous labels for a given input. The proposed method is predicated on handling labels that are absent in the training set or missing labels. Other researchers have examined the application of multi-label classification models to identify diabetes complications (Zhou et al., 2021).

A technique for automatically creating learning groups for MSCL (Mobile Computer Supported Collaborative Learning) systems was presented by Amara et al. (2016). The k-means algorithm was used for the following three grouping criteria:

1. Leaner's attributes: abilities, age, gender, and religion.
2. Interactions between learners and teachers, including the learner's behavior.
3. Contextual data: learner location and learning time.

Many researchers employ k-means techniques to organize cooperative groups. Each time, a k-means variant is created with a distinct set of parameters based on the situation's specific needs, and the proposed approach is then evaluated. Thus, the CSCL community needs a guide to inform it on how a k-means algorithm can be applied to clustering students and which configurations are appropriate to compose an efficient cluster and improve learning outcomes.

In the Table 1, we present a comparative analysis of some studies on the formation of groups to promote collaborative learning.

Table 1. Overview of forming group techniques

Reference	Techniques	Contributions
(Christodoulopoulos & Papanikolao, 2007)	Fuzzy C-means for random selection	Homogeneous and heterogeneous groups. Manual interference by the instructor. Equality is provided by the group size.
(Abnar et al., 2012)	Genetic algorithm Greedy algorithm	The homogeneous and heterogeneous groupings of mixed groups are formed with equal dimensions. An iterative process that satisfies the teacher-tutor on the grouping.
(Jozan & Taghiyareh, 2013)	Genetic algorithm	Inter-homogenous and intra-heterogeneous groups. Three algorithms are compared: random, genetic, and the proposed method.
(Amara et al., 2016)	k-means	Three types of grouping criteria : Learners' characteristics, Learners' learning behaviors and contextual information.
(Cole et al., 2021)	Solve the forming group problem using semi-supervised learning algorithms based on a unique positive label for each element.	Develop a method for performing multi-label learning in environments where only a few labels are available.
(Zhou et al., 2021)	Multi-label classification models based on medical data processing algorithms are used to address this problem.	Develop a method for diagnosing diabetic complications based on medical data.

3. A NEW APPROACH TO GROUP FORMATION BASED ON BEHAVIORAL ENGAGEMENT

As mentioned earlier in this paper, there are many forming group techniques. However, few studies have used learners' traces as inputs. In this study, the proposed process relies primarily on the learner's behavior while studying in a distance learning environment.

The proposed approach should meet the following objectives:

1. We construct a model for each learner in the learning environment.
2. A set of indicators is computed using the traces available in the constructed model.
3. Evaluate the engagement level of each learner based on the previously calculated indicators.
4. Group learners according to their behavior, i.e., their level of behavioral commitment in intra-heterogeneous groups.

Our approach comprises three steps (Figure 1):

1. Collecting learners' digital traces.
2. Measuring the learner's engagement level.
3. Automatic grouping of learners.

Figure 1. Process of proposed approach.

3.1. COLLECTION OF STUDENTS' DIGITAL TRACES

In this module, all learners' traces are collected during their interactions with their peers to measure their behavior engagement. To collect these traces, the learner performs any action from the moment of their first access to the platform until their disconnection. In this study, we separated the collected traces into four categories (Table 2):

- ***Participation traces:*** This category records all actions performed by learners regarding their interaction with other learners when using communication tools (Messaging tool, Forum, etc.) offered by the system. For example, messages sent by the learner, messages answered by the learner, posts (topics, answers), and comments posted by the learner in the system's Forum.
- ***Presence traces*** concern the availability of learners, the number of connections to the system, and so forth.
- ***Effort traces:*** This category contains traces left by learners while completing their learning activities, such as consulting or downloading pedagogical resources.
- ***Meeting deadlines traces:*** consistently meeting deadlines that show strong engagement with academic or professional obligations, indicating effective time management skills and a sense of accountability.

Table 2 determines the learner's behavioral engagement indicators.

Table 2. Indicators related to learners' traces

Traces category	Action
Participation	Sending emails Consulting emails Responding to electronic messages Creating a new subject in the forum tool Answering some submitted questions
Presence	System Access Connections to the system
Effort	Consulting pedagogical resources Downloading pedagogical resources Research tasks on the proposed system
Meeting deadlines	Submission of assignments/homework within given deadlines

3.2. MEASURING LEARNERS' ENGAGEMENT LEVEL

In this step, we propose a set of indicators to measure each learner's level of engagement based on traces collected according to their behavior. We recall that not all actions have the same weight regarding learners' engagement, so weighing actions according to their importance is essential to obtain a more precise and meaningful measure. For example, viewing an email demonstrates superficial engagement, whereas responding to an email indicates deeper engagement.

In our work, to evaluate learners' engagement levels based on their traces, we create "a points system" where each type of trace is assigned a weight based on its importance. We propose the following weighting:

- Less important trace=1 point.
- Moderately important trace=2 points.
- Important trace=3 points.

For example, viewing an email's content indicates less important engagement than answering an email. Therefore, if we give a weight equal to 1 point for '**Consult emails**, ' we should give a weight equal to 2 points for "**Response to an email**."

To calculate the level of engagement according to each trace category, we used the following formula:

$$\mathbf{En\_cat}_k\mathbf{(Li)} = \Sigma_{j=1}^{n} P_j * AC_{jk} \quad (1)$$

$\mathbf{En\_cat}_k$(Li): The engagement value of learner Li for category k.
$\mathbf{AC}_{jk}$: Action/trace j for the engagement category k.
$\mathbf{P}_j$: The weight allotted to each action or trace j corresponds to its level of importance.
n: number of actions/traces.

Thus, the behavioral engagement levels En_cat_1, En_cat_2, En_cat_3, and En_cat_4 represent the categories of traces related to *Participation, Presence, Effort, and Meeting deadlines*, respectively. These levels were normalized to a data scale between 0 and 1. This normalization preserves the original value distribution and transforms all values in the interval [0, 1]. For this purpose, min-max normalization is applied (Indira et al., 2019), and the new normalized value **$\mathbf{En\_cat}_k$(Li)**' is given by the following formula:

$$\boldsymbol{En\_cat_k(Li)'} = (En\_cat_k(Li) - min(En\_cat_k)) \;/\; (max(En\_cat_k) - min(En\_cat_k)) \quad (2)$$

3.3. AUTOMATIC GROUPING OF LEARNERS

In this work, our objective is to form heterogeneous groups of learners based on their behavioral engagement levels in an online learning system. We propose two steps to achieve this: First, we use the k-means technique to form k groups containing the most similar learners (Algorithm 1). The k-means and clustering algorithms, in particular, all have a common goal: grouping similar elements into clusters. These elements can be anything as long as they are encoded in a data matrix. K-means has numerous fields of application; it is notably used in clustering data mining during data exploration to detect similar individuals.

A point (a learner X) is assigned to a cluster according to its distance from a different K centroids. In addition, this point (learner X) is assigned to a cluster if it is closer to its centroid (minimum distance). Finally, the distance between two points in the k-means case is calculated using formula (3).

Generally, depending on the need, other techniques can be used once these populations are detected. Our approach used another method or algorithm to group learners into heterogeneous groups. Second, we use one learner from each group to form new groups with varied engagement levels (see Algorithm 2). From this algorithm, we form heterogeneous groups of learners. These groups are composed according to the learners' activities in the online learning system. In heterogeneous groups, learners are grouped according to differences in levels of engagement. Heterogeneous groups based on abilities positively affect students, regardless of their abilities. However, high-potential students may be dissatisfied because helping low-potential students requires additional effort.

The following figure (c.f. Figure 2) illustrates the learners' dynamic grouping process.

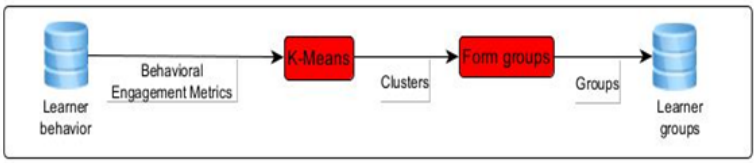


Figure 2. Proposed grouping process.

Step 1: K means clustering

In this step, we use the K-means algorithm to form groups of learners based on their level of engagement (algorithm 1). K-means clustering is a widely used method for partitioning a set of data points into K clusters, which are measured by the number of connected nearby objects (neighbors) to form a cluster. We use this method to group learners in K clusters based on similarity or distance. The distance between two learners was calculated using the following formula:

$$d(x1, x2) = \sqrt{\Sigma_{j=1}^{n}(x_{1j} - x_{2j})^2} \quad (3)$$

In this study, the number k of clusters (groups) was equal to the number of categories. In other words, because we have 4 categories, the number of clusters is 4.

Algorithm 1. *k-means algorithm*

Inputs :

- K: number of groups to form.
- Data matrix: L learners (lines of the data matrix), where each learner Li is represented by K engagement levels (columns of the data matrix) (En_cat_{1i}, En_cat_{2i}, En_cat_{3i}, … En_cat_{ki})

Outputs: k initial groups or clusters: G1, G2, …, Gk

Begin

Randomly select K points (one row of the data matrix). These points are the cluster centers (called centroids).

Repeat

Each point (element of the data matrix) is assigned to the group closest to its center.

The center of each cluster is recalculated, and the centroid.

Until convergence (or stabilization of the total population inertia)

End

Step 2: Form heterogeneous groups

Based on the groups obtained in the previous step, we propose to form new groups with different levels of engagement using the following algorithm:

Algorithm 2. *New group formation algorithm*

Inputs : K groups: G1, G2, G3, …, Gk

N: Number of learners (***Lij: Learner number i from the group j)***

Outputs: M *New groups (NG1, NG2, ……, NGM)*

Begin

If N mod K=0, then M=N div K

else M=N div K + 1

endif

For i:=1 **to** K,

For j:=1 **to** M,

Each learner Lij is assigned from group Gi to group NGj.

If there are unaffected learners in group Gi, then

Assign these learners to the incomplete groups.

End For

End For

End

3.4. ARCHITECTURE OF THE PROPOSED SYSTEM

To validate the proposed approach, we proposed a collaborative learning system composed of groups with different levels of behavioral engagement. Figure 3 shows the architecture of the proposed system, which is composed of the following components:

- Course management module: This module considers all actions related to creating course content (concepts and learning objects).
- The trace management sub-system: It collects the learner's traces and calculates the indicators that help group learning. This sub-system is composed of two modules:
  1. Traces collection module.
  2. Indicators calculating module.
- Grouping module: This module applies the proposed algorithm (described above) for grouping learners into groups with heterogeneous engagement profiles.

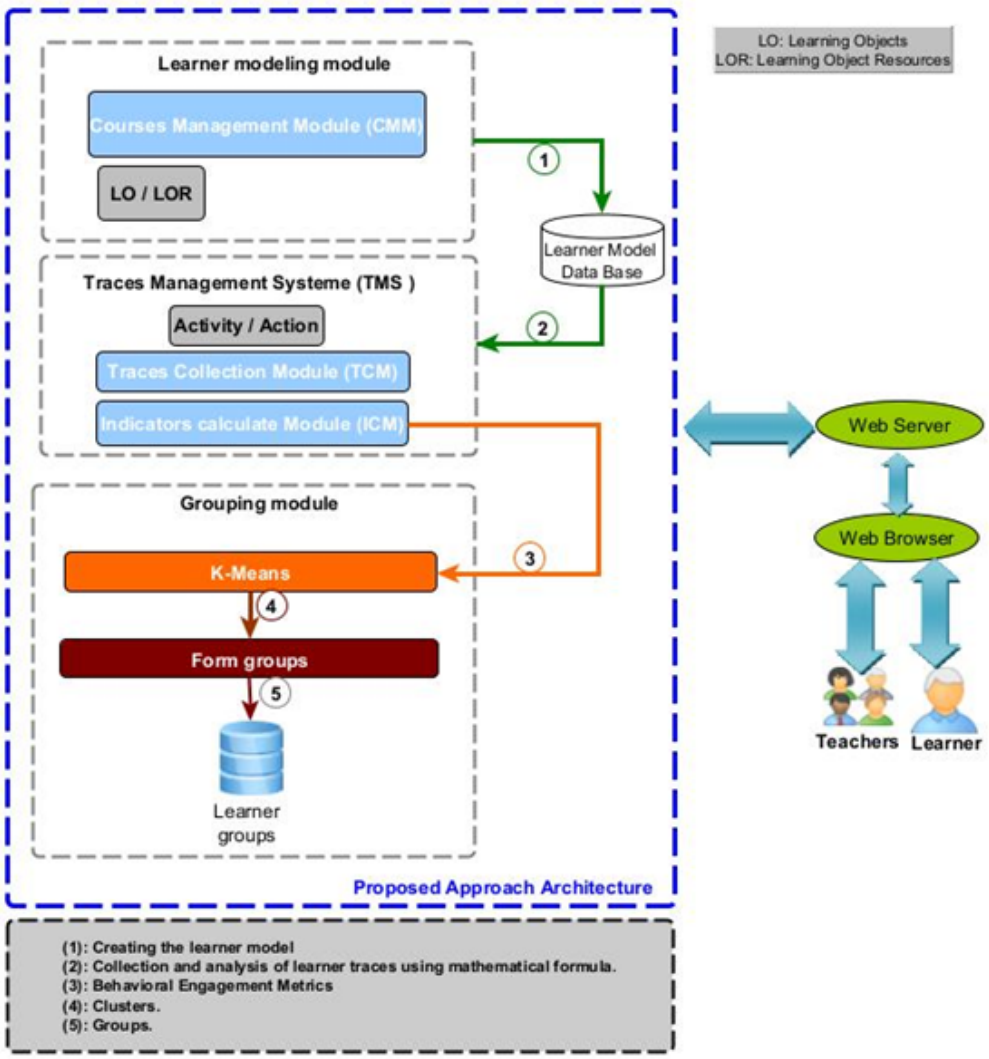


Figure 3. Proposed Approach Architecture.

4. VALIDATION OF THE PROPOSED APPROACH: RESULTS AND DISCUSSION

To test the proposed approach, we generated a dataset of 100 learners. Therefore, we have 100 learners, indicated by L1 until L100, whose details are given in the Table 3.

Step 1

After applying the K-means algorithm to form four (4) clusters, their centers were initially chosen as learners L9, L18, L40, and L22.

Table 3. Values of the tested datasets

Learner	Participation	Presence	Effort	Meeting deadline
L1	1.00	0.30	1.00	0.02
L 2	0.06	1.00	1.00	0.97
L3	0.56	0.20	0.50	0.55
L4	0.72	0.90	1.00	0.23
L5	0.89	0.90	0.50	0.58
L6	1.00	0.60	0.50	0.15
L7	0.78	0.10	0.00	0.77
L8	0.50	0.30	1.00	0.79
L9	0.28	0.60	0.00	0.75
L10	0.00	0.00	0.75	0.21
L11	0.94	0.70	0.00	0.65
L12	0.11	0.80	0.50	0.03
L13	0.67	1.00	0.50	0.89
L14	0.78	0.90	0.00	0.26
L15	1.00	0.90	1.00	0.93
L16	0.33	0.90	0.50	0.76
L17	0.83	0.70	0.50	0.59
L18	0.28	1.00	0.25	0.23
L19	0.89	0.50	0.25	0.00
L20	0.72	0.90	0.75	0.49
L21	0.33	0.60	1.00	0.52
L22	1.00	0.60	0.75	0.48
L23	0.22	0.80	0.75	0.32
L24	0.50	0.80	1.00	0.62
L25	0.50	0.80	0.75	0.58
L26	0.83	0.70	1.00	0.55
L27	1.00	0.80	0.25	1.00
L28	0.83	0.40	0.50	0.08
L29	0.89	0.60	0.50	0.55
L30	0.78	0.10	0.75	0.15
L31	0.44	0.50	0.50	0.88
L32	0.11	0.40	1.00	0.35
L33	0.06	0.30	0.00	0.48
L34	1.00	0.80	1.00	0.80
L35	0.83	0.40	0.75	0.28
L36	0.78	0.00	0.75	0.31
L37	0.33	0.90	0.50	0.98
L38	0.33	0.90	0.00	0.62
L39	0.50	0.00	0.25	0.18
L40	0.33	0.00	0.50	0.51
L41	0.12	0.55	0.4	0.33
L42	0.80	0.90	0.2	0.60
L43	0.20	0.80	0.75	0.40
L44	0.60	0.40	1.00	0.90
L45	0.10	0.20	0.80	0.66
L46	0.40	0.80	0.15	0.11
L47	0.95	0.12	0.67	0.32
L48	0.45	0.20	0.14	0.62
L49	0.13	0.82	0.82	1.00
L50	0.56	0.22	0.34	0.18
L51	0.44	0.66	0.22	0.36
L52	0.87	0.88	0.68	0.90
L53	0.99	0.17	0.55	0.30
L54	1.00	0.55	0.88	0.77
L55	0.98	1.00	0.16	0.23
L56	0.64	0.66	0.48	0.66
L57	0.12	0.11	0.41	0.33
L58	0.57	0.16	0.14	0.25
L59	0.46	0.14	0.28	0.64
L60	0.36	0.75	0.18	0.19
L61	0.15	0.28	0.84	0.36
L62	0.32	0.19	0.68	0.12
L63	0.58	0.54	0.56	0.57

Table 3. Continued

L64	0.14	0.87	0.69	0.26
L65	0.55	0.57	0.49	0.83
L66	0.68	0.19	0.15	0.87
L67	0.68	0.23	0.58	0.69
L68	0.94	0.88	0.46	0.78
L69	0.52	0.26	.0.61	0.94
L70	0.11	0.23	0.16	0.45
L71	0.28	0.16	0.58	0.17
L72	0.48	0.64	0.98	0.36
L73	0.16	0.45	0.18	0.58
L74	0.45	0.88	0.26	0.17
L75	0.55	0.48	0.12	0.78
L76	0.99	0.59	0.45	0.16
L77	1.00	1.00	1.00	0.68
L78	0.22	0.14	0.16	0.89
L79	0.78	0.32	0.65	0.23
L80	0.50	0.30	0.45	0.59
L81	0.14	0.26	0.98	0.78
L82	0.47	0.66	0.14	0.16
L83	0.78	0.69	0.57	0.78
L84	1.00	0.60	0.90	0.67
L85	0.44	0.22	0.66	0.96
L86	0.66	0.35	0.68	0.45
L87	1.00	0.99	0.88	0.69
L88	0.69	0.25	0.87	0.64
L89	0.32	0.66	0.17	0.98
L90	0.15	0.17	0.88	0.35
L91	0.68	0.69	0.17	0.16
L92	0.45	0.80	0.90	0.74
L93	0.17	0.19	0.43	0.16
L94	0.16	0.18	0.55	0.45
L95	0.22	0.80	0.67	0.78
L96	0.85	0.84	0.99	1.00
L97	1.00	0.98	0.96	0.85
L98	0.65	0.79	1.00	0.98
L99	0.98	0.80	1.00	0.99
L100	0.94	0.58	0.80	1.00

Let μ1, μ2, μ3, and μ4 be the centers of gravity of clusters 1, 2, 3, and 4, respectively (See Table 4).

1st iteration:

Table 4. Gravitational centroids of the first iteration.

Centers of gravity	The « center » learner	Participation	Presence	Effort	Meeting deadline
μ1	L9	0.28	0.60	0.00	0.75
μ2	L18	0.28	1.00	0.25	0.23
μ3	L40	0.33	0.00	0.50	0.51
μ4	L22	1.00	0.60	0.75	0.48

To calculate the distance, we applied the formula 3. After calculating all the distances between the learners, we find the following clusters were identified:

Group 1: L7, L9, L11, L16, L31, L33, L37, L38, L49, L65, L66, L73, L75, L78, L89

Group 2: L2, L12, L14, L18, L23, L41, L42, L43, L46, L48, L51, L55, L60, L64, L67, L74 L82, L91

Group 3: L3, L8, L10, L30, L32, L36, L39, L40, L45, L50, L57, L58, L59, L61, L62, L69
L70, L71, L80, L81, L85, L90, L93, L94

Group 4: L1, L4, L5, L6, L13, L15, L17, L19, L20, L21, L22, L24, L25, L26, L27, L28
L29, L34, L35, L44, L47, L52, L53, L54, L56, L63, L68, L72, L76, L77, L79, L83
L84, L86, L87, L88, L92, L95, L96, L97, L98, L99, L100

Updating the centers of gravity: The centers of the formed clusters are represented by averages; thus, we can find new centers of gravity below (See Table 5).

2nd iteration:

Table 5. Gravitational centroids of the second iteration.

Centers of gravity	Participation	Presence	Effort	Meeting deadline
μ1	0.41	0.55	0.24	0.79
μ2	0.42	0.75	0.36	0.34
μ3	0.35	0.18	0.60	0.44
μ4	0.82	0.68	0.76	0.61

The new clusters resulting from the last iteration are as follows:

Group 1: L2, L7, L9, L11, L27, L31, L33, L37, L38, L48, L49, L65, L66, L73, L75, L78 L89, L95

Group 2: L12, L14, L16, L18, L19, L23, L28, L41, L42, L43, L46, L51, L55, L60, L64, L70 L74, L82, L91

Group 3: L3, L8, L10, L30, L32, L36, L39, L40, L45, L47, L50, L57, L58, L59, L61, L62, L67 L69, L71, L79, L80, L81, L85, L86, L90, L93, L94

Group 4: L1, L4, L5, L6, L13, L15, L17, L20, L21, L22, L24, L25, L26, L29, L34, L35, L44, L52, L53, L54, L56, L63, L68, L72, L76, L77, L83, L84, L87, L88, L92, L96, L97. L98, L99, L100

Step 2

After obtaining four groups from the previous step, algorithm 2 was used to form new groups. These groups are shown in the Table 6.

Table 6. Forming heterogeneous groups

	1st Learner	2nd Learner	3rd Learner	4th Learner
Group 1	L2	L12	L3	L1
Group 2	L7	L14	L8	L4
Group 3	L9	L16	L10	L5
Group 4	L11	L18	L30	L6
Group 5	L27	L19	L32	L13
Group 6	L31	L23	L36	L15
Group 7	L33	L28	L39	L17
Group 8	L37	L41	L40	L20
Group 9	L38	L42	L45	L21
Group 10	L48	L43	L47	L22
Group 11	L49	L46	L50	L24
Group 12	L65	L51	L57	L25
Group 13	L66	L55	L58	L26
Group 14	L73	L60	L59	L29
Group 15	L75	L64	L61	L34
Group 16	L78	L70	L62	L35
Group 17	L89	L74	L67	L44
Group 18	L95	L82	L69	L52
Group 19	L93	L91	L71	L53
Group 20	L94	L92	L79	L54
Group 21	L77	L96	L80	L56
Group 22	L83	L97	L81	L63
Group 23	L84	L98	L85	L68
Group 24	L87	L99	L86	L72
Group 25	L88	L100	L90	L76

4.1. DEGREE OF HETEROGENEITY

As mentioned in this paper, each learner is represented by his engagement level, which is described as a quadruplet (Participation, effort, presence, meeting deadline). We propose assigning a color to each category according to its value to obtain groups of learners with heterogeneous behavioral engagement levels. The colors of each category are as follows:

If $En\_cat_i \in [0, 0.2[$, then En_cat_i(color)="Red"
If $En\_cat_i \in [0.2, 0.4[$, then En_cat_i(color)="Orange"
If $En\_cat_i \in [0.4, 0.6[$, then En_cat_i(color)="Blue"
If $En\_cat_i \in [0.6, 0.8[$, then En_cat_i(color)="Purpil"
If $En\_cat_i \in [0.8, 1]$, then En_cat_i(color)="Green"

For each learner, we have four colors (CC1, CC2, CC3, CC4), and CCi is the color of the category i.
Therefore, in a group composed of four learners L1, L2, L3, and L4, we obtain the following:
Li(CCi1, CCi2, CCi3, CCi4), where CCij is the color of category j of learner i.
This study aimed to form groups with members with heterogeneous levels of each of the four categories (in other words, having different colors). To this end, we propose the following algorithm to calculate the degree of heterogeneity of each group.

Algorithm calcul_degree_group_het (K: integer);
Input: En_cat (Color) of all learners:(CCi1, CCi2, CCi3, CCi4)
Output: Degree_group_het$_k$ //the heterogeneity degree of group K
Begin
number_color$_i$ (CCi1, CCi2, CCi3, CCi4)=j // j ϵ **[1, 4]**
If j=1 then het_cat_i=0
Else het_cat_i=((j-1) +1/(5-j))/ 4
Degree_group_het$_k$=$\Sigma_{i=1}^{4}$(het_cat$_{ik}$)/4
End.

Example:
We assume we have a group G1 composed of learners represented by their engagement levels.
L1(Red, Green, Green, Green)
L2(Red, Green, Blue, Red)
L3(Blue, Orange, Blue, Blue)
L4(Green, Blue, Green, Red)
Therefore, the *Degree_group_het* of this group is 0.54

- number_color$_1$(Red, Red, Blue, Green)=3 and het_cat$_1$=2, 5/4=0.62
- number_color$_2$(Green, Green, Orange, Blue)=3 and het_cat$_2$=2, 5/4=0.62
- number_color$_3$(Green, Blue, Blue, Green)=2 and het_cat$_3$=1.33/4=0.33
- number_color$_4$(Green, Red, Blue, Red)=3 and het_cat$_4$=2, 5/4=0.62

Degree_group_het=0.62+0.62+0.33+0.62=2.19/4=0.54

4.2. DEGREE OF HETEROGENEITY IN THE PROPOSED APPROACH FOR FORMING GROUPS BASED ON LEARNERS' BEHAVIORAL ENGAGEMENT

We obtained a set of groups by applying the two steps of the proposed method. To determine whether the formed groups contain learners with heterogeneous engagement levels, we calculated the degree of heterogeneity for these groups. As a result, Table 7 summarizes the degree of heterogeneity among the formed groups using our proposed approach.

Table 7. Degree of heterogeneity in intra-group.

Groups	D_H_Sub-Profile 1	D_H_Sub-Profile 2	D_H_Sub-Profile 3	D_H_Sub-Profile 4	A_D_H
Group1	(Red, Red, Blue, Green)=0.62	(Green, Green, Orange, Blue)=0.62	(Green, Blue, Blue, Green)=0.33	(Green, Red, Blue, Red)=0.62	0.54
Group2	(Purpil, Purpil, Blue, purpil)=0.33	(Red, Green, Orange, Green)=0.62	(Red, Red, Green, Green)=0.62	(Purpil, Orange, Purpil, Orange)=0.62	0.54
Group3	(Orange, Orange, Red, Green)=0.62	(Purpil, Green, Red, Green)=0.62	(Red, Blue, Purpil, Blue)=0.62	(Purpil, Purpil, Orange, Blue)=0.62	0.62
Group4	(Green, Orange, Green, Green)=0.33	(Purpil, Green, Red, Purpil)=0.62	(Red, Orange, Purpil, Blue)=1	(Purpil, Orange, Red, Red)=0.62	0.64
Group5	(Green, Green, Red, Purpil)=0.62	(Green, Blue, Blue, Green,)=0.62	(Orange, Orange, Green, Blue)=0.62	(Green, Red, Orange, Green)=0.62	0.62
Group6	(Blue, Orange, Purpil, Green)=1	(Blue, Green, Red, Green)=0.62	(Blue, Purpil, Purpil, Green)=0.62	(Green, Orange, Orange, Green)=0.62	0.71
Group7	(Red, Green, Blue, Green)=0.62	(Orange, Blue, Red, Purpil)=1	(Red, Blue, Orange, Blue)=0.62	(Blue, Red, Red, Blue)=0.62	0.71
Group8	(Orange, Red, Orange, Purpil)=0.62	(Green, Blue, Red, Green)=0.62	(Blue, Blue, Blue, Purpil)=0.33	(Green, Orange, Blue, Blue)=0.62	0.54
Group9	(Orange, Green, Red, Orange)=0.62	(Green, Green, Orange, Purpil)=0.62	(Red, Orange, Green, Green)=0.62	(Purpil, Purpil, Purpil, Blue)=0.33	0.54
Group10	(Blue, Orange, Green, Green)=0.62	(Orange, Green, Blue, Purpil)=1	(Red, Blue, Purpil, Purpil)=0.62	(Purpil, Blue, Blue, Blue)=0.33	0.64
Group11	(Red, Blue, Blue, Blue)=0.33	(Green, Green, Orange, Green)=0.33	(Green, Red, Orange, Green)=0.62	(Green, Red, Red, Purpil)=0.62	0.47
Group12	(Blue Blue, Red, Blue)=0.33	(Blue, Purpil, Red, Green)=1	(Blue, Orange, Blue, purpil)=0.62	(Green, Orange, Orange, Blue)=0.62	0.64
Group13	(Purpil, Green, Red, Green)=0.62	(Red, Green, Red, Purpil)=0.62	(Red, Orange, Red, Green)=0.62	(Green, Orange, Orange, Blue)=0.62	0.62
Group14	(Red , Orange, Blue, Green)=1	(Blue, Purpil, Red, Purpil)=0.62	(Red, Red, Orange, Blue)=0.62	(Blue, Red, Purpil, Blue)=0.62	0.71
Group15	(Blue Red, Red, Green)=0.62	(Blue, Green, Orange, Green)=0.62	(Red, Purpil, Green, Green)=0.62	(Purpil, Orange, Orange, Green)=0.62	0.62
Group16	(Orange, Red, Orange, Green)=0.62	(Red, Orange, Red, Blue)=0.62	(Red, Red, Purpil, Purpil)=0.62	(Green, Blue, Red, Orange)=1	0.71
Group17	(OrangeBlue, Purpil, Purpil)=0.62	(Blue, Green, Orange, Blue)=0.62	(Red, Orange, Blue, Green)=1	(Green, Red, Purpil, Green)=0.62	0.71
Group18	(Orange, Blue, Blue, Green)=0.62	(Green, Purpil, Orange, Green)=0.62	(Purpil, Red, Purpil, Purpil)=0.33	(Purpil, Red, Green, Green)=0.62	0.54
Group19	(Red, Purpil, Orange, Green)=1	(Red, Purpil, Red, Red)=0.33	(Blue, Red, Blue, Blue)=0.33	(Red, Red, Red, Orange)=0.33	0.49
Group20	(Red, Blue, Purpil, Green)=1	(Red, Green, Orange, Blue)=1	(Blue, Green, Purpil, Green)=0.62	(Blue, Purpil, Orange, Purpil)=0.62	0.81
Group21	(Green, Green, Blue, Green)=0.33	(Green, Green, Orange, Green)=0.33	(Green, Green, Blue, Red)=0.62	(Purpil, Green, Blue, Orange)=1	0.57
Group22	(Blue, Green, Red, Blue)=0.62	(Purpil, Green, Orange, Blue)=1	(Red, Green, Green, Blue)=0.62	(Red, Green, Purpil, Blue)=1	0.81
Group23	(Green, Purpil, Blue, Green)=0.62	(Purpil, Purpil, Orange, Green)=0.62	(Green, Green, Purpil, Blue)=0.62	(Purpil, Green, Green, Purpil)=0.62	0.62
Group24	(Green, Green, Purpil, Blue)=0.62	(Green, Green, Orange, Purpil)=0.62	(Green, Green, Purpil, Green)=0.33	(Purpil, Green, Blue, Orange)=1	0.64
Group25	(Purpil, Green, Red, Green)=0.62	(Orange, Blue, Red, Blue)=0.62	(Green, Green, Green, Blue)=0.33	(Purpil, Green, Orange, Red)=1	0.64
Average_Degree_Heterogeneity_Groups					**0.60**

The greater the degree of heterogeneity, the more distant the learners are (Heterogeneous). Therefore, this degree of heterogeneity provides a better result with low heterogeneity in small groups.

4.3. DISCUSSION OF THE OBTAINED RESULTS

An analysis of the results in Table 7 reveals that 18 groups (72%) have learners with at least three different engagement categories. In other words, learners of these groups have complementary behavioral engagement values, which is the main objective of this study. In addition, the five groups were composed of learners having at least two different behavioral engagement values. Finally, only two groups (8%) had learners close to each other regarding behavioral engagement values. In conclusion, these results are encouraging and promotive because more than 70% of the obtained groups contained at least three of four learners with different engagement levels. In other words, learners within the formed groups can share.

During this research, we encountered some obstacles. First, we had to figure out how to measure each learner's behavior during his interactions within the system and with his teammates. The issue lies in measuring qualitative values using learners' actions. These actions were used to calculate learners' engagement levels. Second, another issue is classifying learners' digital traces to measure different engagement levels. Experts in psychology must validate the classification of each action into categories. In addition, associating the weights of each action requires further study using more developed techniques like machine learning techniques, to obtain the adequate weight of each learner's digital trace.

Third, in this study, learners were regrouped into groups according to their engagement levels. The aim of collaboration within the group is not indicated. Therefore, learners can communicate with each other without evaluating their collaboration results, which constitutes a limitation of this research. To address this limitation, we propose creating projects to be carried out during the collaboration process by the members of each formed group. So, we can, in the end, evaluate the regrouping process through the analysis of the results obtained by each group during the realization of the assigned projects.

Finally, we believe that adopting learners' engagement levels as a criterion for forming heterogeneous groups of learners will improve their cognitive levels. Further, researchers in the CSCL field can adapt our approach to form groups of learners to perform collaborative tasks, complete common projects, or resolve common exercises or problems. Furthermore, artificial intelligence techniques, like deep learning techniques, can be used to form groups adaptively in collaborative learning contexts.

5. CONCLUSION AND FUTURE WORK

Recently, after the COVID-19 pandemic, we have seen growing interest in adapting online learning platforms. However, after the actual use by students, most users, especially teachers, noticed the students' poor interaction or commitment to the activities requested by them. This situation has enabled researchers to search for methods and techniques to improve students' engagement following distance classes. This involvement has several types: emotional, cognitive, social, and behavioral. In this study, behavioral engagement was used because of its importance in learning.

To improve this type of engagement, we group learners into groups with different levels of behavioral engagement. We must first propose indicators to measure commitment to achieve this goal. These indicators were calculated from traces left by learners during their learning and collaboration processes with other students in the e-learning system.

To answer the questions posed in the introduction section, we can say that in this research, we have proposed a new method of grouping learners based on their levels of behavioral engagement. This level is calculated using four indicators: participation, presence, effort, and meeting deadlines. The latter are calculated according to the learners' traces. In addition, the groups must have complementary levels to enable learners to benefit from each other. To this end, we use the k-means algorithm in the first step of the proposed algorithm. Then, in the second step, a new algorithm was used to obtain groups with heterogeneous or complementary engagement profiles.

We conducted a series of tests on a randomly created dataset to validate the obtained results. The results obtained are considered encouraging and promising. By simulating the grouping method, our system effectively grouped people into different profiles according to complementary engagement levels.

In future work, we plan to test the proposed system in an online learning environment to make learning more user-friendly and effective. We also propose to use other methods for grouping learners based on other clustering algorithms and compare them with the proposed algorithm.

Peer Review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- S.Z., H.T., Y.L.; Data Acquisition- S.Z., H.T., Y.K.; Data Analysis/Interpretation- S.Z., H.T., Y.L., Z.M.; Drafting Manuscript- S.Z., H.T., Y.L.; Critical Revision of Manuscript- H.T., Y.L., Z.M.; Final Approval and Accountability- S.Z., Y.L., H.T., Z.M.; Technical or Material Support - S.Z., Y.L., H.T., Z.M.; Supervision – H.T., Y.L., Z.M.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

ORCID IDs of the authors

Souhila Zerdoudi	0000-0002-2982-1633
Houda Tadjer	0000-0001-7624-1343
Yacine Lafifi	0000-0001-8232-4196
Zohra Mehenaoui	0000-0002-6732-7839

REFERENCES

Abnar, S., Orooji, F., & Taghiyareh, F. (2012, January). *An evolutionary algorithm for forming mixed groups of learners in web based collaborative learning environments.* In the IEEE international conference on technology enhanced education (ICTEE) (pp. 1-6). IEEE.

Amara, S., Macedo, J., Bendella, F., & Santos, A. (2016). Group formation in mobile computer-supported collaborative learning contexts: A systematic literature review. *Journal of Educational Technology & Society, 19*(2), 258-273.

Anzieu, D., & Martin, J. Y. (1971). *La dynamique des groupes restreints.* Presses univ. de France.

Atherton, M., Shah, M., Vazquez, J., Griffiths, Z., Jackson, B., & Burgess, C. (2017). Using learning analytics to assess student engagement and academic outcomes in open access enabling programmes. Open Learning: *The Journal of Open, Distance and e-Learning,* 32(2), 119-136.

Bekele, R. (2005). *Computer-assisted learner group formation based on personality traits* (Doctoral dissertation, Staats-und Universitätsbibliothek Hamburg Carl von Ossietzky).

Bouyzem, M., AL Meriouh, Y., & Moustakim, O. (2021). Le e-learning à l'université Abdelmalek Essaâdi: Une analyse descriptive du point de vue des enseignants. *Revue Economie, Gestion et Société, 1*(32).

Chen, B., Chang, Y. H., Ouyang, F., & Zhou, W. (2018). Fostering student engagement in online discussion through social learning analytics. *The Internet and Higher Education*, 37, 21-30.

Christodoulopoulos, C. E., & Papanikolaou, K. A. (2007, October). *A group formation tool in an e-learning context*. In 19th IEEE international conference on tools with artificial intelligence (ICTAI 2007) (Vol. 2, pp. 117-123). IEEE

Cole, E., Mac Aodha, O., Lorieul, T., Perona, P., Morris, D., & Jojic, N. (2021). *Multi-label learning from single positive labels*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 933-942).

Combaudon, S. (2018). *MySQL 5.7: administración y optimización*. Ediciones Eni.

Da Rocha, H. (2019). *Learn Chart. js: Create interactive visualizations for the web with chart.* js 2. Packt Publishing, Ltd.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59-109.

Indira, B., Valarmathi, K., & Devaraj, D. (2019). An approach to enhance packet classification performance of software-defined network using deep learning. *Soft Computing*, 23, 8609-8619.

Isotani, S., Mizoguchi, R., Isotani, S., Capeli, O. M., Isotani, N., De Albuquerque, A. R., ... & Jaques, P. (2013). A Semantic Web-based authoring tool to facilitate the planning of collaborative learning scenarios compliant with learning theories. *Computers & Education*, 63, 267-284.b

Jozan, M. M. B., & Taghiyareh, F. (2013). An evolutionary algorithm for homogeneous grouping to enhance web-based collaborative learning. *International Journal of Computer Science Research and Application,* 3(1), 74-85.

Karaoglan Yilmaz, F. G., & Yilmaz, R. (2022). Learning analytics intervention improves students' engagement in online learning. *Technology, Knowledge and Learning*, 27(2), 449-460.

Kirschner, P. A., Jochems, W., Dillenbourg, P., & Kanselaar, G. (2002). *Three worlds of CSCL: Can we support CSCL.* Heerlen: Open University of the Netherlands.

Maqtary, N., Mohsen, A., & Bechkoum, K. (2019). Group formation techniques in computer-supported collaborative learning: A systematic literature review. *Technology, Knowledge and Learning, 24*, 169-190.

Matazi, I., Messoussi, R., & Bennane, A. (2014, May). *The design of an intelligent multi-agent system for supporting collaborative learning.* In the 9th International conference on intelligent systems: Theories and applications (SITA-14) (pp. 1-8). IEEE.

Mujkanovic, A., Lowe, D., Willey, K., & Guetl, C. (2012, June). *Unsupervised learning algorithm for adaptive group formation: Collaborative learning support in remotely accessible laboratories.* In International Conference on Information Society (i-Society 2012) (pp. 50-57). IEEE.

O'Donnell, K., & Reschly, A. L. (2020). Assessment of student engagement. In A. L. Reschly, A. J. Pohl, & S. L. Christenson (Eds.), Student engagement: Effective academic, behavioral, cognitive, and affective interventions at school (pp. 55–76). Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-37285-9_3.

Ounnas, A., Davis, H. C., & Millard, D. E. (2007). *Semantic modeling for group formation.* Workshop on Personalisation in E-Learning Environments at Individual and Group Level (PING) at the 11th International Conference on User Modeling UM2007, Corfu, Greece.

Ouyang, F., Chen, S., & Li, X. (2021). Effect of three network visualizations on students' social-cognitive engagement in online discussions. *British Journal of Educational Technology*, 52(6), 2242-2262.

Rajabalee, Y. B., & Santally, M. I. (2021). Learner satisfaction, engagement and performances in an online module: Implications for institutional e-learning policy. *Education and Information Technologies*, 26(3), 2623-2656.

Reschly, A. L., Pohl, A., Christenson, S. L., & Appleton, J. J. (2017). Engaging adolescents in secondary schools. *School Mental Health Services for Adolescents*, 45-77.

Resta, P., & Laferrière, T. (2007). Technology in support of collaborative learning. *Educational Psychology Review*, 19, 65-83.

Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. M. (2010). Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning, 5*, 43-102.

Srba, I., & Bielikova, M. (2014). Dynamic group formation as an approach to collaborative learning support. *IEEE Transactions on Learning Technologies, 8*(2), 173-186.

Stahl, G., Koschmann, T., & Suthers, D. (2006). *Aprendizaje Colaborativo apoyado por computador: Una perspectiva histórica.* G. Stahl, T. Koschmann, & D. Suthers, Aprendizaje Colaborativo apoyado por computador: Una perspectiva histórica.(pág. 426). Cambridge: RK Sawyer.

Zheng, Y., Li, C., Liu, S., & Lu, W. (2018). An improved genetic approach for composing optimal collaborative learning groups. *Knowledge-Based Systems*, 139, 214-225.

Zhou, L., Zheng, X., Yang, D., Wang, Y., Bai, X., & Ye, X. (2021). Application of multi-label classification models for the diagnosis of diabetic complications. *BMC Med Inform Decis., 21*(1), 182.

How cite this article

Zerdoudi, S., Tadjer, H., Lafifi, Y., & Mehenaoui, Z. (2024). A New Approach to Grouping Learners Based on Behavioral Engagement in CSCL Environments. *Acta Infologica, 8*(2), 260-273. https://doi.org/10.26650/acin.1567382

İSTANBUL UNIVERSITY PRESS

Acta Infologica
acin 2024, 8 (2), 274–286
DOI: 10.26650/acin.1395045

Review Article

International Standards for Digital Twins

Esra Kumaş[1]

[1]Ostim Teknik Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Yönetim Bilişim Sistemleri Bölümü, Ankara, Türkiye

Corresponding author : Esra Kumaş
E-mail : esra.kumas@ostimteknik.edu.tr

ABSTRACT

Interest in digital twins has grown significantly due to the maturation of enabling technologies, including cloud computing, 5G, data storage, computing capabilities, the Internet of Things (IoT), Artificial Intelligence (AI), and machine learning. As these technologies have reached a sufficient maturity level and costs have decreased, the concept of digital twins has gained prominence. Given the growing interest in digital twins, which are widely acknowledged as a fundamental component in the digitalization of production underlying smart manufacturing, the absence of standards concerning digital twin terminology, architecture, and models during application development has resulted in divergent user interpretations. This lack of standardization leads to significant confusion regarding the concept of digital twins. To alleviate this confusion, establishing guidelines and developing unified terminology and implementation procedures are crucial steps to promote the widespread adoption of digital twins. This study aims to contribute to the development of digital twins by analyzing studies that present different technologies, procedures, and standards for implementing digital twins, with a particular focus on the ISO 23247 Digital Twin Framework for Manufacturing.

Keywords: Digital twin, Standards of digital twin, ISO 23247

Submitted : 24.11.2023
Accepted : 11.07.2024
Published Online : 19.09.2024

1. INTRODUCTION

The foremost pioneer of digital twins was NASA, which innovatively used simulators from 200,000 miles away in 1970 to diagnose and repair damage on the Apollo 13 spacecraft. While the concept of digital twins was initially proposed in 2002, the technologies required to make this concept widely accessible recently reached a turning point (Accenture, 2021). Grieves and Vickers initially defined a digital twin as "a digital information structure of a physical system as an entity in itself." In this context, the term "twin" implies that digital information is connected to a physical system throughout its lifecycle. The application of this concept to production enables manufacturers to create purposeful digital representations of production systems and processes using collected data (Shao & Helu, 2020).

Digital twins facilitate the modeling, analysis, and optimization of problems that are challenging in the physical world. By modeling the performance of humans, physical entities, and processes in a virtual environment, digital twins help us understand how humans behave under various conditions. By leveraging machine learning, digital twins can simulate complex scenarios in numerous new ways, capturing potentialities that might otherwise go unnoticed (Accenture, 2021). With the ability to work with real-time data from the physical world, digital twins can represent the physical world with unparalleled precision and accuracy and conduct mathematical modeling. This capability allows decision makers to conduct limitless "what-if" analyses by altering numerous variables to model potential outcomes.

Digital twins not only enable the real-time visual modeling of products and processes but also aid decision makers in understanding how productivity can be enhanced, risks mitigated, issues resolved, and potential future states of a product or system (Shao & Helu, 2020). Due to tthesepromised benefits, the uuseof digital twins is increasingly expanding across various sectors, from manufacturing to energy, health care to defense, logistics to supply chains. However, the concept of digital twins leads to confusion and difficulty for practitioners because of its foundation in the convergence of new technologies and different disciplines, its nascent application, and the lack of sufficient practical examples. Overcoming these practical challenges requires the standardization of terms, concepts, and reference models to serve as guiding principles (Shao et al., 2023). This study aims to contribute to the development of digital twins and eliminate the conceptual confusion that arises in practice by examining works that define standards applicable to implementing digital twins, primarily the ISO 23247 Digital Twin Framework for Manufacturing. At the same time, a very limited amount of research exists on the standards of digital twins. Therefore, this study can serve as a guide for future research in this area.

2. LITERATURE REVIEW

The number of studies focusing on standards for implementing digital twins, particularly ISO 23247, remains limited. Although 96 articles concerning digital twins are accessible in national reference journals indexed in LUlabim, none specifically explore digital twin standards. Within the WoS and Scopus databases, 16 studies were documented

Notably, the initial publication dates vary, with one in 2020, two in 2021, four in 2022, six in 2023, and three in 2023.

Shao and Help, affiliated with the American National Institute of Standards and Technology (NIST), elucidated the foundational framework of the ISO 23247 standard in their 2020 study, "Framework for a digital twin in manufacturing: Scope and requirements." This research explored and categorized the diverse applications of digital twins in manufacturing, underscoring the necessity of standards to ensure interoperability among digital twins designed for distinct purposes and employing different technologies. This study emphasizes the potential of the ISO 23247 standard to bridge this gap.

Huiyue and Run (2021) conducted a case analysis leveraging edge computing advantages to access and analyze real-time data in digital twins. They outlined a general framework aligned with the ISO 23247 standard for the realization phase of this application.

Jacoby et al. (2021) developed integrated software for digital twins to enhance their business interoperability. This software adheres to the ISO 23247 reference architecture and can be seamlessly integrated with other open production standards.

Another 2022 literature study by Huile, Tang, and Xun surveyed digital twin platforms used in academia and industry, emphasizing the significance of delineating requirements and complying with the ISO 23247 standard during their development.

Eirinakis et al. (2022) proposed a digital twin focused on predicting and mitigating interruptions in production using the ISO 23247 reference architecture.

Lidell et al. (2022) discussed current and future challenges associated with digital twins, highlighting the imperative need for standards and the pivotal role of the ISO 23247 standard in addressing these challenges.

Kim et al. (2022) presented a digital twin application architecture tailored for additive manufacturing that was structured based on the ISO 23247 framework.

Huan et al. (2022) aimed to present a literature review of digital twin platforms in manufacturing. This study first proposes a generalized definition of a digital twin platform, and then, based on this definition, a literature review on the digital twin platform is conducted using the Web of Science database. The importance of ISO 23247 is also discussed to give an overview of the requirements for building a digital twin platform.

Kibira et al. (2023) analyzed existing standards, technologies, and methodologies to create a digital twin for a robot work cell and developed it in accordance with the ISO 23247 framework.

Ferko et al. (2023) scrutinized how existing digital twin architectures align with the ISO 23247 architecture, highlighting the incipient adoption of this standard by multinational companies and challenges in precisely measuring compliance due to differences between existing architectures and the recently proposed standard.

In their study, Ferrero et al. (2023) defined the functional requirements for adapting lean manufacturing to the digital twin model based on the ISO 23247 standard.

In their study, Spaney et al. (2023) created a digital twin of the milling process in manufacturing based on the ISO 23247 standard. This study presents a digital twin architecture framework that optimizes the manufacturing process.

Shao et al. (2023) reviewed the ISO 23247 set of standards to inform the manufacturing community in general and for applications in emerging industry sectors, such as bio-manufacturing, and new manufacturing technologies, such as 3D printers.

Cabral et al. (2023) devised a digital twin for implementation in a CNC machine tool by designing protocols for data acquisition, storage, visualization, simulation, and cloud-based transfer aligned with the ISO 23247 reference.

Triphati et al. (2024) interviewed experts on the digital twin ecosystem and conducted a systematic literature review. Based on these interviews and a literature review, the study identified various stakeholders and their roles in adding value to the digital twin ecosystem. The study also revealed the technical and nontechnical challenges faced by ecosystem-driven digital twins and highlighted the importance of standardization as a solution.

In their work, Caiza and Sanz (2024a) developed a digital twin to be implemented in the Industry 4.0 lab. The requirements for the design and implementation of the digital twin architecture are based on ISO 23247. The architecture includes 3D design and visualization, a communication entity through the OPC UA protocol for the collection of state changes of production elements, digital modeling and updating according to the collected data, and the use of AR and VR, all built according to the ISO 23247 standard. The results demonstrate that the proposed architecture provides interoperability between different platforms and control subsystems. The results demonstrate that the ISO 23247 standard makes an important contribution to interoperability, which is one of the most serious problems of digital twins.

In another study (2024b), Caiza and Sanz developed a digital twin architecture for flexible manufacturing systems based on the ISO 23247 standard. The proposed system is based on the integration of digital twin technologies together with augmented reality and motion tracking and aims to increase the interaction and flexibility between physical and virtual environments in real time. As a result of this study, it was stated that ISO 23247 facilitates the integration of different technologies, but its use with the ISO 16792 standard in the creation of the digital twin architecture allows for more detailed development of the 3D model.

Digital twins are among the most important technologies of Industry 4.0 and smart manufacturing, and their use in both academia and industry has recently become significantly widespread. It has a wide range of applications and research areas, from manufacturing to smart cities, to healthcare and accounting systems. However, the platforms used for digital twin architectures are also changing significantly. Although it is a new and developing field, the wide range of application areas and platforms and the fact that it is not yet fully mature have created some difficulties in the implementation of digital twins. Among the above-mentioned studies, it has been reported that digital twin architectures created with reference to the ISO 23247 standard facilitate the integration of different technologies. These studies emphasize the importance of the standards to be applied to digital twins and the potential of the ISO 23247 standard to overcome the difficulties encountered in the implementation of digital twins. However, studies and applications in academia and business are still not at a sufficient level.

3. DIGITAL TWIN STANDARDS

A standard can be delineated as a universally agreed-upon set of regulations and principles for interoperability (Türkiye Bilimler Akademisi, 2023). These standards serve to ensure the seamless operation of technology and foster trust to facilitate efficient market functionality.

Moreover, they establish a shared framework for assessing and appraising performance, thus allowing for the compatibility of components manufactured by diverse entities (Shao et al., 2023). Standardization denotes the process of instituting and implementing specific regulations, involving the cooperation of all stakeholders, aimed at yielding

economic advantages in a particular sphere of activity. Essentially, standardization dictates the attributes to be pursued in the production of goods and services (Ministry of Industry and Technology, 2023).

Various organizations, such as the International Standards Organization (ISO), International Electrotechnical Commission (IEC), International Telecommunication Union (ITU), Institute of Electrical and Electronics Engineers (IEEE), American National Institute of Standards and Technology (NIST), and American National Standards Institute (ANSI), conduct standardization studies on digital twins or related technologies (ANSI, 2020; Wang et al., 2022).

Standards developed for digital twins encompass a spectrum of focal areas. Wang et al. (2022) categorized these standards based on physical assets, virtual assets, data, connectivity, and services, as illustrated in Figure 1.

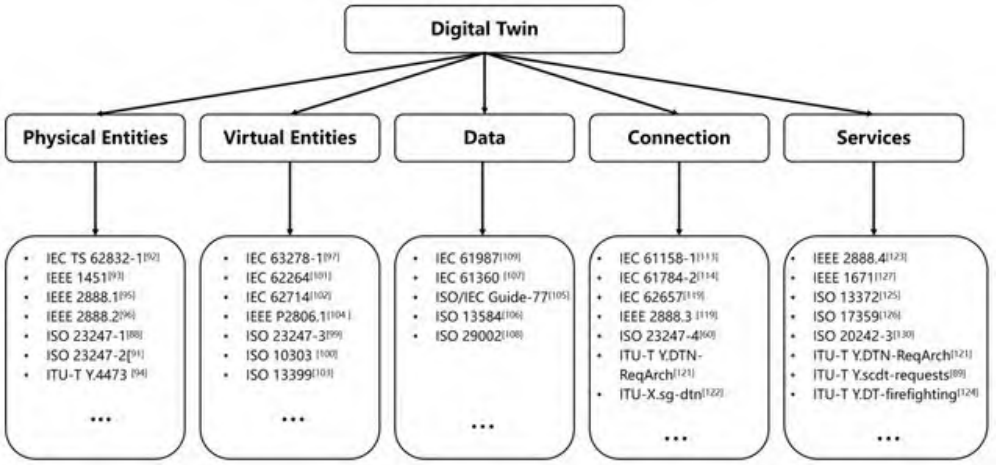


Figure 1. *Framework of digital twin standards (Wang vd., 2022)*

ISO functions as an international standardization body composed of representatives from national standard organizations. The ISO 23247 series defines a framework for the establishment of digital twins. Founded in 1906, the IEC is the foremost global organization for developing and disseminating international standards on electrical, electronics, and related technologies. Collaborative efforts focusing on digital twins are being undertaken by both the ISO and the IEC. ISO/IEC JTC 1/SC 41 strives to standardize aspects of the Internet of Things and Digital Twinning and offers guidance to other entities developing applications in this domain (International Electrotechnical Commission, 2023). Within ISO/IEC JTC 1/SC 41, the working group WG6 is presently engaged in five projects, two of which are dedicated to standard development: ISO/IEC 30172: This document outlines digital twin use cases across diverse sectors like autonomous mobility, energy, smart cities, buildings, manufacturing, business management, and healthcare. ISO/IEC 30173: Focused on digital twin concepts and terminology, this document delves into applications, ecosystems, life cycle processes, and the classification of digital twins (IEC, 2022).

The International Telecommunication Union (ITU) functions as the specialized agency of the United Nations for telecommunications and information and communication technologies (ICTs). Within the ITU, the Telecommunications Standardization Sector (ITU-T) is a permanent body responsible for examining and issuing recommendations for standardizing global telecommunications (ITU, 2022). The ITU-T recommendation "Y.3090: Digital Twin Network" delineates the requirements and architecture of digital twin networks, encompassing functional requirements, service criteria, architectural aspects, and security concerns. Unlike ISO 23247 and ISO/IEC initiatives, ITU-T Y.3090 focuses primarily on network-oriented standards. The Institute of Electrical and Electronics Engineers Standards Association (IEEE-SA) is an integral part of IEEE, developing global standards spanning various industries. Within the IEEE-SA, the C/SM/DT_WG Digital Twin Working Group developed the P 3144 standard. This standard defines a digital twin maturity model catering to industry needs, encompassing digital twin capability areas, related sub-areas, and assessment methodologies comprising assessment content, processes, and maturity levels (IEEE-SA, 2022). The American National Standards Institute (ANSI) developed the IPC-2551 standard for digital twins. This standard allows any manufacturer, design organization, or solution provider to initiate the implementation of interoperability to create smart value chains and assess their current level of IPC Digital Twin readiness.

This standard provides the information and guidance necessary to understand the IPC Digital Twin, Digital Twin Product, Digital Twin Manufacturing, and Digital Twin Lifecycle. The standard also provides information and guidance on how organizations benefit from the IPC Digital Twin, how to assess readiness, and how an organization of any size can prepare to apply the IPC Digital Twin approach to itself and/or its products (ANSI, 2020).

The US National Institute of Standards and Technology (NIST) released draft report NISTIR 8356 titled "Digital

Twin Technology and Emerging Standards Considerations" in 2021. This report comprehensively details digital twins, elucidating their motivations, utility, common operational procedures, use cases, and illustrative examples. With a specific focus on technical facets pertinent to the cybersecurity of digital twins, this study scrutinizes both existing and emerging cybersecurity challenges stemming from the deployment of digital twin architectures. Moreover, it analyzed trust-related concerns, explored the ramifications of lacking standards on the functionality and quality of digital twins, and correlated these evaluations with existing NIST cybersecurity directives (Voas et al., 2021).

Within the purview of NIST, evaluations were conducted regarding ISO 23247 studies, generating scenarios aligned with the framework established by ISO 23247 (Shao, 2021). A summary of the above-mentioned and other standards for digital twins and their constituent technologies is presented in Table 1.

Table 1. Standards for digital twins and related technologies

Standard Name	*Full Name*	*Year*	*Purpose*	*Application Areas*	*Focus*
ISO 10303	Automation systems and integration — Product data representation and exchange	1994	Exchange of product data between computers.	CAD systems, industrial data management.	Data Representation & Exchange
IEEE 1451	Standard for a Smart Transducer Interface for Sensors and Actuators	1997	Network-based integration of smart sensors.	IoT, industrial automation, smart sensors.	Sensors & Actuators
ISO 13584	Industrial automation systems and integration — Parts library	1998	Representation and exchange of part libraries.	Digital manufacturing, data exchange.	Parts Library & Exchange
IEC 62264	Enterprise-control system integration	2003	Model for integration between manufacturing and business management.	Manufacturing management, Industry 4.0, business processes.	Enterprise Integration
ISO 29002	Industrial automation systems and integration — Exchange of characteristic data	2004	Use of metadata in product data management.	Product data management, industrial automation.	Metadata Exchange
ISO 13372	Condition monitoring and diagnostics of machines — Vocabulary	2004	Terminology and methodology for condition monitoring and diagnostics.	Maintenance management, industrial facilities.	Condition Monitoring
IEEE 1671	Standard for Test System Interface Architecture	2006	Information modeling standard for test systems.	Automated test systems, electronic testing.	Test Systems & Interfaces
ISO 13399	Cutting tool data representation and exchange	2006	Digital format representation of cutting tool data.	Cutting tools, digital manufacturing processes.	Data Representation
IEC TR 62541	Industrial communication protocol — OPC Unified Architecture	2006	Unified architecture for communication in automation systems.	Industrial automation, communication systems.	Industrial Communication
IEC 61987	Industrial-process measurement and control — Data structures and elements in process equipment catalogues	2007	Data structures for devices used in industrial processes.	Industrial automation, process control.	Data Structures

Table 1. Continued

ISO/IEC Guide-77	Guide for description of reference models and general requirements for classification of products	2008	Provides a guide for product classification and data management.	Product lifecycle, supply chain management.	Product Classification
IEC 61360	Standard data element types with associated classification scheme for electric components	2010	Methodology for product classification and data elements.	Product lifecycle management, data classification.	Data Classification
IEC 61784	Industrial communication networks – Profiles	2010	Standards for industrial communication profiles.	Industrial automation, factory communication systems.	Communication Profiles
ISO 17359	Condition monitoring and diagnostics of machines — General guidelines	2011	Procedures for monitoring the condition of industrial machinery.	Industrial maintenance, equipment management.	Condition Monitoring
ISO/IEC 27001:2013	Information security management systems — Requirements	2013	Requirements for information security management systems.	Information security, risk management.	Information Security
IEC 62714	Engineering data exchange format for use in industrial automation systems engineering	2014	Exchange and integration of engineering data.	Industrial engineering, data exchange.	Data Exchange
ITU-T Y.DTN-ReqArch	Requirements and architecture for delay-tolerant networking (DTN)	2016	Architecture and requirements for delay-tolerant networks.	Space, defense, large data transmission.	Delay-Tolerant Networking
IEC 62657	Industrial communication networks – Wireless communication networks	2017	Spectrum management for industrial wireless communication systems.	Wireless communication, industrial automation.	Wireless Communication
ISO/IEC 38505-1:2017	Information governance — Framework for information and records management	2017	Governance standards for information management.	Information management, governance.	Information Governance
ISO 20242	Industrial automation systems and integration — Distributed application protocol	2018	Interface structures for distributed systems.	Automation systems, distributed networks.	Distributed Systems

Table 1. Continued

ITU-T Y.DT-firefighting	Digital twin standard for firefighting applications	2018	Requirements for digital firefighting systems.	Fire safety, digital monitoring.	Digital Twin for Safety
IEC 62443	Cybersecurity for industrial automation systems	2018	Cybersecurity framework for industrial systems.	Cybersecurity, industrial automation, digital twins.	Cybersecurity for Industrial Systems
ISO/IEC 20889:2018	Privacy enhancing technologies — Data deletion and destruction techniques	2018	Standards for data deletion and destruction techniques.	Data security, privacy management.	Data Privacy & Security
IEC 61158	Industrial communication networks – Fieldbus specifications	2019	Communication protocols for industrial networks.	Industrial networks, automation.	Industrial Communication
IEC TS 62832	Industrial-process measurement, control and automation – Digital factory framework	2020	Provides a reference model for digital factory systems.	Industrial automation, digital factories.	Digital Factory
IEEE P.2806	Standard for Digital Reality — Reference Architecture	2020	Provides a reference architecture for digital reality technologies.	Digital reality, augmented reality, virtual reality.	Digital Reality
ITU-X.sg.dtn	Delay-Tolerant Networking (DTN) standardization efforts	2020	Standardization efforts for delay-tolerant networks.	Space and defense, large data transmission.	Delay-Tolerant Networking
IEEE P2048	Standard for Digital Reality — Framework for interaction	2020	Interaction framework for digital reality applications.	Digital reality, virtual and augmented reality.	Digital Reality
IEEE 2888	Standard for Sensory and Exchanged Information in the Internet of Things	2021	Standards for human sensory and experience sharing over the internet.	Multimedia systems, virtual reality, IoT.	IoT & Sensory Systems
ISO 23247	Automation systems and integration — Digital twin framework for manufacturing	2021	Creation of digital twins for smart manufacturing.	Digital twins, Industry 4.0, smart manufacturing.	Digital Twin
ITU-T Y.4473	Requirements and capabilities for IoT data sharing framework	2021	Technical requirements for data sharing in IoT.	IoT, data sharing, network management.	Data Sharing & IoT

Table 1. Continued

ISO/IEC TS 27110:2021	Internet of Things (IoT) — Reference architecture	2021	General framework for IoT architectures.	IoT architecture, digital systems.	IoT Reference Architecture
ISO/IEC 21823-3:2021	Internet of Things (IoT) — Interoperability for IoT systems	2021	Interoperability for the Internet of Things.	IoT, integration, cyber-physical systems.	IoT Interoperability
ISO/IEC 30147:2021	Internet of Things (IoT) — Bridging IoT protocols	2021	Bridging protocols and architectures for IoT.	IoT, communication protocols.	IoT Communication
IEC 63278	Industrial-process measurement, control, and automation — Digital factory framework for system configuration	2022	Guidelines for configuring and managing digital factory systems.	Digital factories, industrial automation.	Digital Factory
ISO/IEC DIS 23894	Cybersecurity — Data security and privacy framework	2022	Framework for data security and privacy.	Data management, privacy and security.	Cybersecurity Framework
IPC-2551	Generic Standard for Smart Manufacturing Systems	2022	Provides guidelines for smart manufacturing systems integration.	Manufacturing, smart systems, automation.	Smart Manufacturing
ITU-T Y.3090	Framework for AI-Based Network Management and Control	2022	Provides a framework for managing and controlling networks using AI.	Network management, artificial intelligence, telecommunications.	AI in Network Management
ISO/IEC DIS 23894	Cybersecurity — Data security and privacy framework	2022	Framework for data security and privacy.	Data management, privacy and security.	Cybersecurity Framework
ISO/IEC DIS 27400	Cybersecurity and IoT — Security and privacy management guidelines	2023	IoT security risk management standard.	IoT, cybersecurity, risk management.	Cybersecurity & IoT
ISO/IEC AWI 30172	Internet of Things (IoT) — Bridging mechanisms in IoT devices	In Progress	Standards for bridging mechanisms in IoT devices.	IoT, network management.	IoT Mechanisms
ISO/IEC AWI 30173	Internet of Things (IoT) — Bridging protocols in IoT devices	In Progress	Development of standards for bridging protocols in IoT devices.	IoT, network protocols.	IoT Protocols

Table 1. Continued

ISO/IEC AWI 5339	Internet of Things (IoT) — Open data sharing framework	In Progress	Standards for open data sharing in IoT.	IoT, data sharing, open networks.	IoT & Data Sharing
ISO/IEC AWI 5339	Internet of Things (IoT) — Open data sharing framework	In Progress	Standards for open data sharing in IoT.	IoT, data sharing, open networks.	IoT & Data Sharing
ISO/IEC AWI 5392	Internet of Things (IoT) — IoT device lifecycle management	In Progress	Standards for managing IoT device lifecycle.	IoT, device management, lifecycle systems.	IoT Device Management
ISO/IEC AWI TR 5469	Information security — Privacy impact assessment	In Progress	Guidelines for conducting privacy impact assessments.	Data security, privacy management, risk assessment.	Privacy & Risk Assessment
ISO/IEC FDIS 22989	Artificial intelligence — Artificial Intelligence concepts and terminology	In Progress	Standardizing terminology and concepts for AI.	AI, data science, digital systems.	AI & Terminology
ISO/IEC FDIS 38507	Information governance — Governance implications of the use of AI	In Progress	Governance of AI in information systems.	AI governance, information management.	AI Governance
P 3144	Standard for Digital Twins for Industrial Systems	In Progress	Guidelines for implementing digital twins in industrial systems.	Industrial systems, digital twins, manufacturing.	Digital Twin

3.1. ISO 23247

To facilitate the creation of ISO digital twins, the concept of "observable production elements" was identified. These elements include personnel, equipment, materials, production processes, facilities, the environment, products and supporting documentation. The primary objective of the ISO 23247 series is to establish a framework that offers comprehensive guidelines, reference architectures, methods, and approaches for developing digital twins by monitoring these "observable production elements" within production contexts (ISO, 2021).

Digital twins play a crucial role in identifying anomalies in production processes, and they are aligned with functional goals such as real-time control, predictive maintenance, in-process adaptation, big data analytics, and machine learning. They achieve this by continuously updating pertinent operational and environmental data, thus enabling the monitoring of these 'observable production elements'. Transparency in processes and execution provided by digital twins contributes to optimizing manufacturing operations (ISO, 2021).

The applicability of the ISO 23247 framework to supporting production types hinges on the standards and technologies available for modeling "observable production elements." Different data standards might be employed across diverse production domains.

The scopes of the four parts within this series are as follows:

ISO 23247-1: Overview and principles

ISO 23247-2: Reference architecture,

ISO 23247-3: Digital representation of production elements,

ISO 23247-4: Information exchange

Figure 2 illustrates the interrelations among the four parts of this series.

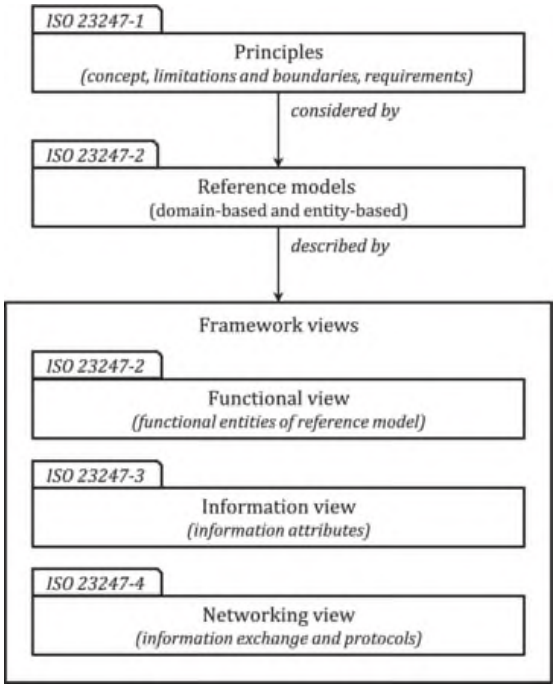


Figure 2. *ISO 23247 series structure (ISO, 2021)*

ISO 23247 Part 1: Overview and General Principles

This section outlines the general principles and requirements governing the development of digital twins in production environments. It establishes terminologies for each segment of the standard and delineates synchronization and communication protocols between digital twins and observable production elements, ensuring the optimization and real-time status of data sourced from these elements.

ISO 23247 Part 2: Reference Architecture

Part 2 covers the reference architecture for digital twins in manufacturing, considering perspectives from both domain and entity perspectives. The architecture comprises four domains: the observable production domain, data collection and device control domain, core domain, and user domain. Each domain delineates a logical array of tasks and functions executed by functional entities. Figure 3 illustrates the functional entity view of the reference architecture model.

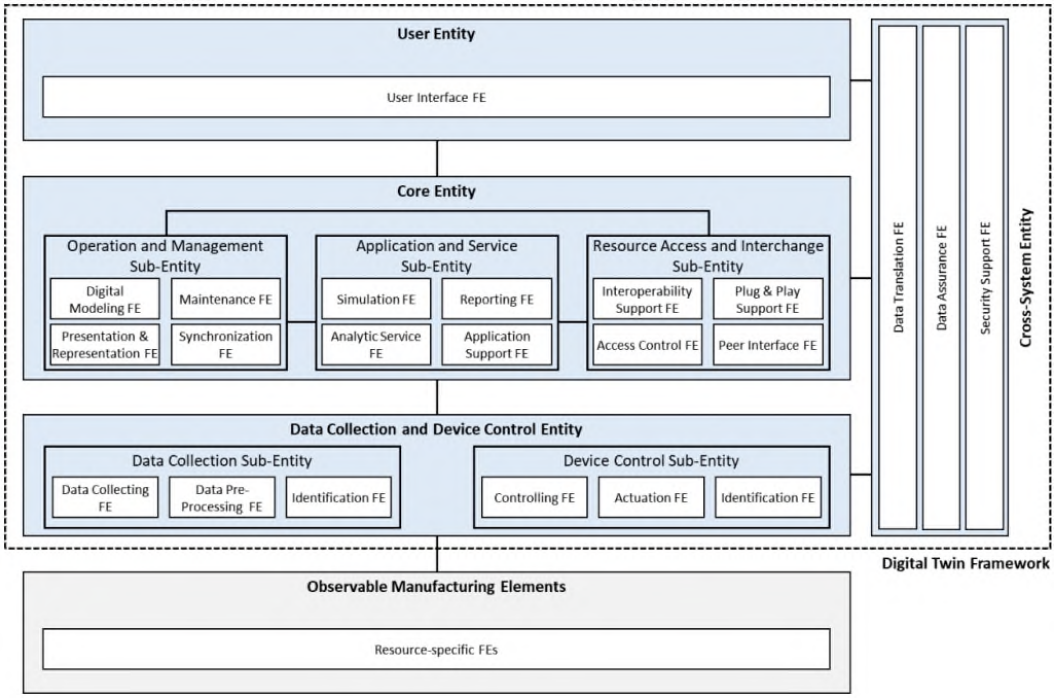


Figure 3. Digital twin manufacturing framework *(Shao, 2021)*

ISO 23247 Part 3: Digital Representation

Part 3 details the digital representation aspect and elucidates the fundamental information attributes—both static and dynamic—pertaining to observable production elements.

ISO 23247 Part 4: Information Exchange

This section outlines the technical requirements for information exchange among entities in the reference architecture. It defines the user network, service network, access network, and proximity network. Furthermore, it provides illustrative use cases illustrating the framework and presents a selection of standards and technologies applicable to information exchange.

4. DISCUSSION and CONCLUSION

Manufacturing is undergoing a digital transformation dubbed Industry 4.0. As a key technology to enable this digital transformation, digital twins enable manufacturers to digitally represent their assets, diagnose problems in advance, collect and manage relevant data, predict and optimize the response of their assets under different conditions. Recently, there has been a growing interest from both academia and industry in the potential benefits of digital twins. As an emerging technology, the digital twin creates a virtual representation of physical objects and develops predictive strategies. Digital twins are virtual representations of resources that organize and manage knowledge and are tightly integrated with AI, machine learning and IoT to further optimize and automate production. All these technologies and models that make up the digital twin are not new. They have been around for some time and have reached a level of maturity to prove their effectiveness. So the power of the digital twin, but also its complexity and difficulty, is not that it is a new technology, but that it brings together many new but mature disruptive technologies. Its power comes when the digital twin brings all these technologies and applications together, and this integration needs standards and common concepts. The Internet of Things, which makes it possible to collect a wide variety of types of data from a variety of objects, advances in powerful but low-cost processing and storage, artificial intelligence applications to help model and optimize the acquired data, and advances in virtual and augmented reality that enable cost-effective visual viewing of digital representations have been important building blocks for the expected benefits of digital twin applications. While the standards established for all these applications are important building blocks for the standards to be established for the digital twin, they are also important challenges for integration.

The potential applications and benefits of digital twins have been demonstrated in both academic and industry applications. The next step is to develop standards and harmonize existing standards to make these applications widespread. While individual companies are starting to use digital twins, there are significant challenges for manufacturers, especially small and medium-sized enterprises, to implement digital twin applications correctly and effectively. The lack of relevant standards for digital twins is a barrier to wider adoption. Because digital twins involve highly complex data collections and functional subsystems, many manufacturers struggle to know where to start when implementing digital twins. While standards are often controversial and seen as premature at first when a new technology emerges, the widespread adoption of digital twin technology depends on standards development efforts.

This study focuses on the digital twin standards developed to broaden and facilitate the usage of digital twins. Digital twins are a significant component for the realization of smart manufacturing. However, due to their multidisciplinary nature, scarcity of applications, and the necessity of integrating emerging technologies, they possess a complex and at times challenging structure. Reducing this complexity, establishing standardized terminology, and having a common architectural framework that promotes interoperability will enhance the application success of digital twins. In this regard, the existence of the ISO 23247 standard, put forward by the International Standards Organization, fills a crucial gap for practitioners. The number of studies related to digital twins in the literature, especially concerning the ISO 23247 standard, is quite limited. This study aims to examine these standards to fill this gap in the literature and contribute significantly to the field for both practitioners and researchers. Additionally, given the scarce number of studies aligned with this standard, there is a need to identify aspects of the standard that are open to improvement.

Peer Review: Externally peer-reviewed.

Conflict of Interest: The author has no conflict of interest to declare.

Grant Support: The author declared that this study has received no financial support.

ORCID ID of the author

Esra Kumaş 0000-0003-3792-1295

REFERENCES

Accenture, (2021). Accenture Federal Technology Vision 2021: Five Trends for Post-pandemic leadership. U.S.

American National Standards Institute (ANSI). (2020). IPC-2551, International Standard for Digital Twins. https://webstore.ansi.org/preview-pages/IPC/preview_IPC-2551.pdf

Cabral, J. V. A., Gasca, E. A. R., & Alvares, A. J. (2023). Digital twin implementation for machining center based on ISO 23247 standard. *IEEE Latin America Transactions, 21*(5), 628-635.

Caiza, G., & Sanz, R. (2024a). An Immersive Digital Twin Applied to a Manufacturing Execution System for the Monitoring and Control of Industry 4.0 Processes. *Applied Sciences, 14*(10), 4125.

Caiza, G., & Sanz, R. (2024b). Immersive Digital Twin under ISO 23247 Applied to Flexible Manufacturing Processes. *Applied Sciences*, 14(10), 4204.

Eirinakis, P., Lounis, S., Plitsos, S., Arampatzis, G., Kalaboukas, K., Kenda, K., ... & Stojanovic, N. (2022). Cognitive digital twins for resilience in production: A conceptual framework. *Information*, 13(1), 33.

Ferrero, V., Alqseer, N., Reslan, M., & Morris, K. C. (2023). Adapting Digital Twin Frameworks Toward Lean Manufacturing for the Circular Economy. In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Vol. 87332, American Society of Mechanical Engineers.

Ferko, E., Bucaioni, A., Pelliccione, P., & Behnam, M. (2023, March). Standardisation in digital twin architectures in manufacturing. In 2023 IEEE 20th International Conference on Software Architecture (ICSA) (pp. 70-81). IEEE.

Huang, H., & Xu, X. (2021). Edge computing enhanced digital twins for smart manufacturing. In International Manufacturing Science and Engineering Conference (Vol. 85079, p. V002T07A020). American Society of Mechanical Engineers.Huang,

H., Ji, T., & Xu, X. (2022). Digital Twin Platforms: Architectures and Functions. *International Manufacturing Science* and *Engineering Conference* (Vol. 85819, p. V002T06A008). American Society of Mechanical Engineers.

International Electrotechnical Commision-IEC. (2023). "ISO/IEC JTC 1/SC 41 Internet of Things and Digital Twin".

IEEE Standards Association-IEEE SA. (2022). "P3144 Standard for Digital Twin Maturity Model and Assessment Methodology in Industry".

ISO. 2021. "ISO (DIS) 23247-21 Automation Systems and Integration-Digital Twin Framework for Manufacturing-Part 1: Overview and general principles". ISO/TC 184/SC4/WG15.

ISO. 2021. "ISO (DIS) 23247-2: Automation Systems and Integration-Digital Twin Framework for Manufacturing-Part 2: Reference Architecture". ISO/TC 184/SC4/WG15.

ISO. 2021. "ISO (DIS) 23247-3: Automation Systems and Integration-Digital Twin Framework for Manufacturing-Part 3: Digital Representation". ISO/TC 184/SC4/WG15.

ISO. 2021. "ISO (DIS) 23247-4: Automation Systems and Integration-Digital Twin Framework for Manufacturing-Part 4: Information Exchange". ISO/TC 184/SC4/WG15.

International Telecommunication Union-ITU. (2022). "Y.3090 : Digital twin network-Requirements and architecture".

Jacoby, M., Volz, F., Weißenbacher, C., Stojanovic, L., & Usländer, T. (2021). An approach for Industrie 4.0-compliant and data-sovereign Digital Twins: Realization of the Industrie 4.0 Asset Administration Shell with a data-sovereignty extension. *Automatisierungstechnik*, 69(12), 1051-1061.

Kim, D. B., Shao, G., & Jo, G. (2022). A digital twin implementation architecture for wire+ arc additive manufacturing based on ISO 23247. *Manufacturing Letters*, 34, 1-5.

Kibira, D., Shao, G., & Venketesh, R. (2023, May). Building A Digital Twin of AN Automated Robot Workcell. In 2023 Annual Modeling and Simulation Conference (ANNSIM) (pp. 196-207). IEEE.

Lidell, A., Ericson, S., & Ng, A. H. (2022). The current and future challenges for virtual commissioning and digital twins of production lines. In SPS2022 (pp. 508-519). IOS Press.

Shao, G., & Helu, M. (2020). Framework for a digital twin in manufacturing: Scope and requirements. *Manufacturing Letters, 24*, 105-107.

Shao, G. (2021). Use case scenarios for digital twin implementation based on ISO 23247. *National institute of standards: Gaithersburg, MD, USA*.

Shao, G., Frechette, S. P., & Srinivasan, V. (2023, June). An Analysis of the New ISO 23247 Series of Standards on Digital Twin Framework for Manufacturing. 2023 *MSEC Manufacturing Science & Engineering Conference*, New Brunswick, NJ, US.

Spaney, P., Becker, S., Ströbel, R., Fleischer, J., Zenhari, S., Möhring, H. C., ... & Portmann, A. (2023, October). Model-driven Digital Twin for Manufacturing Process Adaptation. In 2023 ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C) (pp. 465-469). IEEE.

Tripathi, N., Hietala, H., Xu Y, Liyanage R. (2024). Stakeholder collaborations, challenges, and emerging concepts in digital twin ecosystems. Information and Software Technology, 107424.

Türkiye Bilimler Akademisi- Tüba. (2023). *Turkish Dictionary of Science Terms.* http://terim.tuba.gov.tr/

Voas, J., Mell, P., & Piroumian, V. (2021). Considerations for digital twin technology and emerging standards (No. NIST Internal or Interagency Report (NISTIR) 8356 (Draft)). National Institute of Standards and Technology.

Wang, K., Wang, Y., Li, Y., Fan, X., Xiao, S., & Hu, L. (2022). A review of the technology standards for enabling digital twin. Digital Twin, 2, 4.

How cite this article

Kumaş, E. (2024). International standards for digital twins. *Acta Infologica, 8*(2), 274-286. https://doi.org/10.26650/acin.1395045