

JSR A

E-ISSN: 2687-6167
NUMBER 56
MARCH 2024

JOURNAL OF
SCIENTIFIC REPORTS A

Journal of Scientific Reports-A, March 2024, Number 56



Kutatya Durgam Cheruvu University Scientific Reports A
Kutatya Center Campus, Davangere Road, 10 KM, 432 70 Kutatya
Phone : (0214) 441 11-62
E-mail : journal@gmail.com
jgsra.com
Durgam Cheruvu University Press





Owner

On Behalf of Kütahya Dumlupınar University
Prof. Dr. Süleyman KIZILTOPRAK (Rector),
On Behalf of Institute of Graduate Studies
Assoc. Prof. Dr. Eray ACAR (Director)

Editorial Board

Önder UYSAL	Kütahya Dumlupınar University/ Mining Engineering
Fatih ŞEN	Kütahya Dumlupınar University / Biochemistry
Oktay ŞAHBAZ	Kütahya Dumlupınar University/ Mining Engineering
Nevzat BEYAZIT	Ondokuz Mayıs University / Enviromental Engineering
Onur KARAMAN	Akdeniz University / Medical Services and Tech.
Cafer ÖZKUL	Kütahya Dumlupınar University / Geological Engineering
Levent URTEKİN	Ahi Evran University / Mechanical Engineering
Ümran ERÇETİN	Kütahya Dumlupınar University / Mechanical Engineering
Ceren KARAMAN	Akdeniz University / Electrical and Energy
Durmuş ÖZDEMİR	Kütahya Dumlupınar University / Computer Engineering
Güray KAYA	Kütahya Dumlupınar University/ Metallurgical and Materials Eng.
Pelin Çağım TOKAT BİRGİN	Kütahya Dumlupınar University/ Metallurgical and Materials Eng.
Sait ALTUN	Kütahya Dumlupınar University/ Metallurgical and Materials Eng.
Sevgi KARACA	Kütahya Dumlupınar University/ Mining Engineering
Ramazan BAYAT	Kütahya Dumlupınar University / Biochemistry
Muhammed BEKMEZCİ	Kütahya Dumlupınar University / Biochemistry
Ayşenur AYGÜN	Kütahya Dumlupınar University / Biochemistry
Safa DÖRTERLER	Kütahya Dumlupınar University / Computer Engineering
Seyfullah ARSLAN	Kütahya Dumlupınar University / Computer Engineering
Büşra TUTUMLU	Kütahya Dumlupınar University / Industrial Engineering
Merve ARSLAN	Kütahya Dumlupınar University / Software Engineering
Bahadır YÖRÜR	Kütahya Dumlupınar University / Industrial Engineering
Naciye Nur ARSLAN	Kütahya Dumlupınar University / Software Engineering

Journal of Scientific Reports-A started its publication life in 2000 as name of Journal of Science and Technology of Dumlupınar University and is a national peer-reviewed journal published regularly twice a year in June and December. The language of the journal is English. Articles submitted to the journal are evaluated by at least two referees who are experts in the subject and selected by the editorial board. All articles submitted to the journal are evaluated by the double-blind method. Articles submitted to our journal for review should not be previously published, accepted for publication and in the process of being evaluated for publication in another journal. All responsibility for the articles published in the journal belongs to the author(s).

The journal aims to share scientific studies carried out in the fields of science and engineering at national and international level with scientists and the public. Original research articles, review articles and short notes in science and engineering disciplines are accepted for the journal. Original research articles are expected to contain

theoretical and experimental results and should not be published in other journals. In the review articles, it is expected that scientific, technological and current developments on a specific subject are reflected by using an extensive bibliography and made a satisfying evaluation of these. Short notes should be brief writings prepared to announce the first findings of an original study.

Editorial Policy

The journal is open access and the article evaluation period is between 1-2 months.

Correspondence Address: Kütahya Dumlupınar Üniversitesi Evliya Çelebi Yerleşkesi Fen Bilimleri Enstitüsü
43270 KÜTAHYA

E-mail: joursra@gmail.com

Phone: 0 274 443 19 42

Webpage: gsjsra.com

Fax: 0 274 265 20 60

Section Editors

Civil Engineering Prof. Dr. M. Çağatay KARABÖRK	Kütahya Dumlupınar University
Mechanical Engineering Prof. Dr. Ramazan KÖSE	Kütahya Dumlupınar University
Electrical-Electronics Engineering Assist. Prof. Kadir VARDAR	Kütahya Dumlupınar University
Computer Engineering Assoc. Prof. Doğan AYDIN	Kütahya Dumlupınar University
Industrial Engineering Assist. Prof. Üyesi Kerem CİDDİ	Kütahya Dumlupınar University
Mining Engineering Assist. Prof. Uğur DEMİR	Kütahya Dumlupınar University
Geology Engineering Assist. Prof. Muzaffer ÖZBURAN	Kütahya Dumlupınar University
Metallurgical and Materials Engineering Prof. Dr. İskender IŞIK	Kütahya Dumlupınar University
Food Engineering Prof. Dr. Muhammet DÖNMEZ	Kütahya Dumlupınar University
Environmental Engineering Assoc. Prof. Dr. Nevzat BEYAZIT	Ondokuz Mayıs University
Mathematics Assist. Prof. Cansu KESKİN	Kütahya Dumlupınar University
Physics Assoc. Prof. Huriye Sanem AYDOĞU	Kütahya Dumlupınar University
Chemistry Assoc. Prof. Bülent ZEYBEK	Kütahya Dumlupınar University
Biology Assist. Prof. Nüket Akalın BİNGÖL	Kütahya Dumlupınar University
Biochemistry Assoc. Prof. Derya KOYUNCU ZEYBEK	Kütahya Dumlupınar University
Occupational Health and Safety Prof. Dr. Cem ŞENSÖĞÜT	Kütahya Dumlupınar University
Software Engineering Assist. Prof. Şerif Ali SADIK	Kütahya Dumlupınar University

Advisory Board

Prof. Dr. Sibel AKAR	Eskişehir Osmangazi University / Chemistry
Prof. Dr. Abdurrahman AKTÜMSEK	Selçuk University/ Bialogy
Prof. Dr. Mustafa ALTUNOK	Gazi University / Tree-Jobs Industrial Engineering
Prof. Dr. Uğur ARİFOĞLU	Sakarya University / Electirical and Electr. Engineering
Prof. Dr. Oktay ARSLAN	Balıkesir University / Chemistry
Prof. Dr. Şükrü ASLAN	Sivas Cumhuriyet University / Enviromental Engineering
Prof. Dr. Ülfet ATAV	Selçuk University / Pyhsics
Prof. Dr. Mustafa BAYRAKTAR	TOBB Ekonomi ve Teknoloji University / Mathamathics
Prof. Dr. Niyazi BİLİM	Konya Technical University / Mining Engineering
Prof. Dr. İsmail BOZTOSUN	Akdeniz University / Pyhsics
Prof. Dr. Erdal ÇELİK	Dokuz Eylül University / Metalurgical and Material Eng.
Prof. Dr. Hayri DAYIOĞLU	Kütahya Dumlupınar University / Bialogy
Prof. Dr. Muhammet DÖNMEZ	Kütahya Dumlupınar University / Food Engineering
Prof. Dr. Mehmet Ali EBEOĞLU	Kütahya Dumlupınar University / Elec.and Electr. Eng.
Prof. Dr. İsmail Göktay EDİZ	Kütahya Dumlupınar University / Mining Engineering
Prof. Dr. İsmail EKİNCİOĞLU	Kütahya Dumlupınar University / Mathematics
Prof. Dr. Kaan ERARSLAN	Kütahya Dumlupınar University / Mining Engineering
Prof. Dr. Zeynal Abiddin ERGÜLER	Kütahya Dumlupınar University / Geological Eng.
Prof. Dr. Seyhan FIRAT	Gazi University / Civil Engineering
Prof. Dr. Remzi GÖREN	Sakarya University / Metalurgical and Material Eng.
Prof. Dr. Rasim İPEK	Ege University / Mechanical Engineering
Prof. Dr. Refail KASIMBEYLİ	Eskişehir Technical University / Industrial Engineering
Prof. Dr. Hamdi Şükür KILIÇ	Selçuk University / Physics
Prof. Dr. Yaşar KİBİCİ	Bilecik Şeyh Edebali University / Geological Eng.
Prof. Dr. İsmail KOCAÇALIŞKAN	Yıldız Technical University / Molecular Bio. and Gen.
Prof. Dr. Mahmut KOÇAK	Eskişehir Osmangazi University / Math-Computer
Prof. Dr. Muhsin KONUK	Üsküdar University / Molecular Biology and Gen.
Prof. Dr. Mustafa KURU	Başkent University / Molecular Biology and Gen.
Prof. Dr. Ömer İrfan KÜFREVİOĞLU	Atatürk University / Biochemistry
Prof. Dr. Halim MUTLU	Ankara University / Geological Engineering
Prof. Dr. Ekrem SAVAŞ	İstanbul Ticaret University / Mathematics
Prof. Dr. Murat TANIŞLI	Eskişehir Technical University / Physics
Prof. Dr. Ali Rehber TÜRKER	Gazi University / Chemistry
Prof. Dr. Mustafa TÜRKMEN	Giresun University / Biology
Prof. Dr. Abdülmecit TÜRÜT	İstanbul Medeniyet University / Physics Engineering
Prof. Dr. Eşref ÜNLÜOĞLU	Eskişehir Osmangazi University / Civil Engineering
Prof. Dr. Nurettin YAYLI	Karadeniz Technical University / Pharmacy
Prof. Dr. Yusuf YAYLI	Ankara University / Mathematics
Prof. Dr. Elçin YUSUFOĞLU	Uşak University / Mathematics
Prof. Dr. Hüseyin Serdar YÜCESU	Gazi University / Automotive Engineering
Prof. Dr. Mehmet Tevfik ZEYREK	Middle East Technical University / Pyhsics

JOURNAL OF SCIENTIFIC REPORTS-A
E-ISSN: 2687-6167

CONTENTS

CORRIGENDUM ARTICLE

- Corrigendum to “betaine supplementation protects rats against alcohol-induced hepatic and duodenal injury: an histopathological study” [Journal of Scientific Reports-A, Number 52, 220-233, March 2023]* 1-2
Ayşe Çakır Gündoğdu*, Fatih Kar, Cansu Özbayer

RESEARCH ARTICLES

- Multi-objective genetic algorithm for the assembly line worker assignment and balancing problem: A case study in the automotive supply industry* 3-22
Gözde Kurada*, Derya Deliktaş
- Determination of the species boundaries of genus Dolerus (Hymenoptera: Tenthredinidae) using the COI gene* 23-35
Mehmet Gülmez*, Ertan Mahir Korkmaz, Mahir Budak
- Hyperspectral anomaly detection with an improved approach: integration of go decomposition algorithm and laplacian matrix modifier* 36-44
Fatma Küçük*
- Evaluation of overall equipment effectiveness (OEE) for mining equipment (shovel-truck): A case study Manisa-Soma (Turkey) open pit mine* 45-55
Sedat Toraman*
- Data correlation matrix-based spam URL detection using machine learning algorithms* 56-69
Funda Akar*
- Drug repurposing analysis with co-expressed genes identifies novel drugs and small molecules for bladder cancer* 70-81
Esra Gov*, Gökçe Kaynak Bayrak,
- A comparative study on the estimation of ultimate bearing capacity of rock masses using finite element and limit equilibrium methods* 82-93
Serdar Koltuk*
- Synthesis, characterization, anti-microbial activity studies of salicylic acid and 2-* 94-104

aminopyridine derivatives salts and their Cu(II) complexes

Halil İlkimen*, Aysel Gülbandır

The effect of fingerprint enhancement methods applied on adhesive surfaces on DNA recovery: a preliminary study 105-115

Fatma Cavus Yonar, Yakup Gulekci*

Supplier selection using the integrated MEREC – CoCoSo methods in a medical device company 116-133

Gülnehal Özel Sönmez, Pelin Toktaş*

Calculations of $Ba_{(1-x)}Sr_xTiO_3$ structure and band gap properties by using density functional theory 134-144

Sinem Aksan*

REVIEW ARTICLES

Urban planning and development in harmony with the geosciences 145-154

Ümit Yıldız



E-ISSN: 2687-6167

Number 56, March 2024

CORRIGENDUM ARTICLE

Receive Date: 15.03.2023

Accepted Date: 22.03.2023

Corrigendum to “betaine supplementation protects rats against alcohol-induced hepatic and duodenal injury: an histopathological study” [Journal of Scientific Reports-A, Number 52, 220-233, March 2023]

Ayşe Çakır Gündoğdu^{a*}, Fatih Kar^b, Cansu Özbayer^c

^a*Kütahya Health Science University, Faculty of Medicine, Department of Histology and Embryology, Kütahya, ORCID: 0000-0002-2466-9417*

^b*Kütahya Health Science University, Faculty of Medicine, Department of Medical Biochemistry, Kütahya, ORCID: 0000-0001-8356-9806*

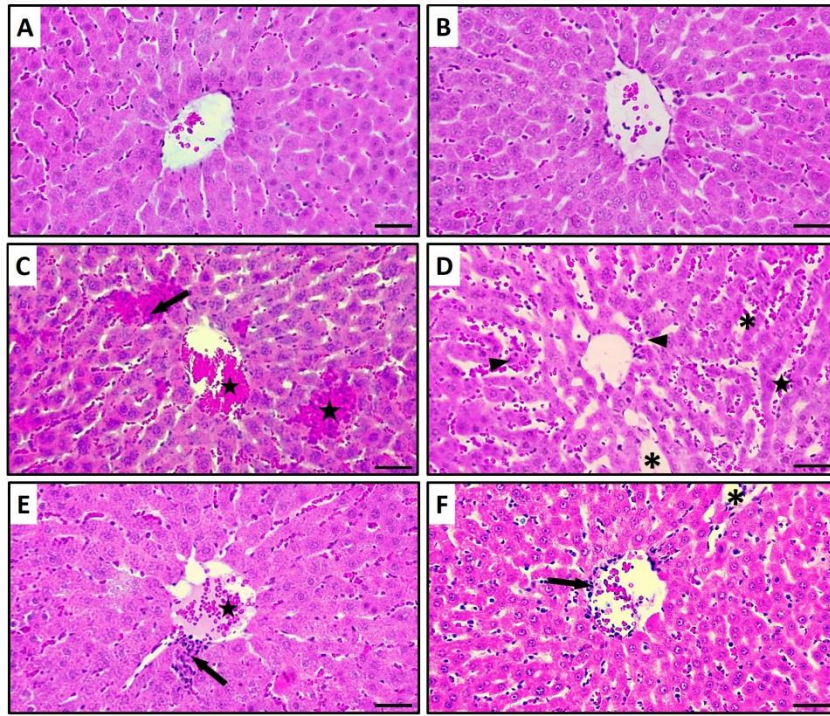
^c*Kütahya Health Science University, Faculty of Medicine, Department of Medical Biology, Kütahya, ORCID: 0000-0002-1120-1874*

© 2023 DPU All rights reserved.

Keywords: Betaine, Ethanol, Liver, Duodenum, Histopathological changes.

The authors regret that the printed version of the article contained some errors. The article title contains a grammatical error. It is corrected as "BETAINE SUPPLEMENTATION PROTECTS RATS AGAINST ALCOHOL-INDUCED HEPATIC AND DUODENAL INJURY: A HISTOPATHOLOGICAL STUDY". The microscopic image of the Control group in Figure 1A was unintentionally presented as a different image by the authors. There is also a mistake in the scale bars of the images in Figure 1. The corrected version is shown below. The correction does not affect the results and conclusion of this article. The authors would like to apologise for any inconvenience caused.

* Corresponding author. *E-mail address:* ayse.cakirgundogdu@ksbu.edu.tr



REFERENCES

- [1] akir Gündođdu, A., Kar, F. and zbayer, C., BETAINE SUPPLEMENTATION PROTECTS RATS AGAINST ALCOHOL-INDUCED HEPATIC AND DUODENAL INJURY: AN HISTOPATHOLOGICAL STUDY, *Journal of Scientific Reports-A*, 52, 220-233, 2023



E-ISSN: 2687-6167

Number 56, March 2024

RESEARCH ARTICLE

Receive Date: 01.09.2023

Accepted Date: 10.10.2023

Multi-objective genetic algorithm for the assembly line worker assignment and balancing problem: A case study in the automotive supply industry

Gözde Kurada^{a*}, Derya Deliktaş^b

^aKütahya Dumlupınar University, Faculty of Engineering, Department of Industrial Engineering, 43000, Kütahya, Turkey, ORCID: 0000-0003-0757-3269

^bKütahya Dumlupınar University, Faculty of Engineering, Department of Industrial Engineering, 43000, Kütahya, Turkey, ORCID:0000-0003-2676-1628

Abstract

The primary challenge in assembly line design is the need for more appropriately allocating tasks and workers to workstations. This study addresses the problem of line balancing and worker assignments, considering the performance disparities among workers during the line balancing process. In the relevant literature, this problem is known as the Assembly Line Worker Assignment and Balancing (ALWAB) problem. This research examines a multi-objective ALWAB Type-2 problem, simultaneously evaluating cycle time and squared load assignment objectives. The study is conducted based on a real-life scenario in a sub-industry automotive industry that manufactures cable equipment. To solve this problem, a multi-objective genetic algorithm approach is proposed. Recognising that the selection of parameter values will influence the algorithm's performance, parameter calibration has been performed. A full factorial experimental design and the **irace** method have been utilised for this purpose. The results are compared with those using parameter values utilised for similar problems in the literature. Furthermore, a sensitivity analysis has been carried out to examine the impact of various relative weight values of the objectives on the result. The results indicate that the experimental design generally yields superior results compared to other methods.

© 2023 DPU All rights reserved.

Keywords: Genetic Algorithm; Assembly Line Worker Assignment and Balancing; **irace**; Design of Experimental Design; Type-2

* Corresponding author. Address.: Kütahya Dumlupınar University, Faculty of Engineering, Department of Industrial Engineering, 43000, Kütahya, Turkey. E-mail address: gozdekurada97@gmail.com

1. Introduction

Production is a primary activity for enterprises producing goods and services [1]. Flow-line production systems have become increasingly crucial for enterprises aiming to increase production quantities, enhance productivity, and reduce costs. Assembly lines are specialised flow production systems crucial in the industrial manufacture of highly standardised products [2]. Designed to meet large production quantities, assembly lines have become foundational to production systems. The workload between workstations must be balanced to achieve more efficient and rapid production on assembly lines. This balanced distribution of tasks across stations is termed assembly line balancing (ALB). The challenge of determining which individual will perform a given task, taking into account performance differences among workers in addition to line balancing, is known as the assembly line worker assignment and balancing (ALWAB) problem [3]. This type of problem is viewed as an extension of the basic ALB problem [4]. The introduction of worker-dependent durations for defined tasks and the integration of worker assignment decisions with workstations in the ALWAB approach enhance the applicability of ALB problem-solving in the manufacturing sectors [5]. ALB problems are categorised into four groups based on their objective functions. This classification can be briefly described as follows:

- (i) Type-1: In Type-1 problems, the cycle time is fixed and known. The objective is to minimise the number of workstations.
- (ii) Type-2: The number of workstations is determined. The objective is to minimise the cycle time according to this number. By minimising the cycle time, it is aimed to increase the production amount per unit of time.
- (iii) Type-E: It is a problem type that tries to simultaneously minimise the cycle time and the number of workstations.
- (iv) Type-F: It is a feasibility problem type which is to determine whether a feasible line balance exists for a given number of workstations and cycle time [6].

In this study, Type-2 was utilised. The proposed algorithm in this study seeks to minimise cycle time and enable quadratic load assignment. This ensures a balanced distribution of tasks to the workstation while taking into account precedence relationships among the jobs, worker-specific task durations, and worker walking times. The ALWAB problem is known as the NP-hard problem [7,8]. Hence, a multi-objective genetic algorithm (GA) has been proposed to solve a real-life problem faced by an automotive sub-industry that produces cable equipment. A multi-objective mathematical model was presented to define this problem.

Parameters used in algorithms are crucial to attaining optimal solutions. In this study, a full factorial experimental design was employed to calibrate and control these parameters. Additionally, using **irace**—a method for parameter calibration—the optimal parameters were identified, allowing the problem to be resolved. Moreover, the algorithm proposed by Mutlu et al. [9] for the ALWAB problem has been employed in the literature. This study also compares the results derived from it. Sensitivity analysis is applied to check the outcome of a decision-maker, to see the risks involved and to analyse the values of parameters [10]. Here, sensitivity analysis was conducted to study the impact of varying relative weight values of objectives on the outcomes.

The rest of this study is organised as follows: Section 1 offers an in-depth review of the ALWAB problem. Section 2 introduces a detailed explanation of the problem, the mathematical model, the proposed GA algorithm, and the calibration of the GA algorithm's parameters and presents both experimental and comparative results. The final section encapsulates the study's concluding observations.

2. Experimental method

2.1. Literature review

The study of the ALB problem in the existing literature was first addressed by Bryton in 1954 and many studies on ALB have been carried out until today [8]. A detailed literature review has been carried out to understand the depth of the work discussed in this study and identify the missing points. This is summarised in Table 1.

In the first stage, ALB studies in the literature were analysed. In ALB studies, task processing times do not vary depending on worker performance. In this problem type, the performance of each worker is neglected. However, this situation does not reflect real life. Since the task processing times vary depending on the worker performances in the real-life problem considered in this study, in the second stage, the studies on ALWAB in the literature were examined. Analysing the accessed studies, it was observed that the setup/transport (walking) times are generally ignored or included in the processing times since they are much smaller than the processing times. Considering that ignoring walking and transport times in problems does not reflect the real situation, walking times are handled separately from work times in this study. In the study of Karsu & Azizoğlu [16], the objective function of minimising the squared load assignment was developed and the walking time was added, and a multi-objective mathematical model was obtained by aiming at workload balancing. This study aimed to contribute to the literature by addressing a real-life problem in the automotive sub-industry that produces cable equipment.

Table 1. Related literature on the ALB and ALWAB problems.

Category	References	Assembly line type	Used objective functions	Solution methods	TP/RLP
ALB problems	Karsu & Azizoğlu [11]	Simple	Workload balancing (Min)	B&B, TS	TP
	Altunay et al. [12]	Parallel	Cycle time (Min)	MM	TP
	Arikan [13]	Simple	Workload balancing (Min)	MM, TS	TP
	Delice et al. [14]	Two-sided U-type	Number of mated-stations (Min) Number of workstations (Min)	ACO	TP
	Süer & Sadeghi [15]	Parallel	Assembly rate (Max) Number of operators (Min)	MM	TP
	Karsu & Azizoğlu [16]	Simple	Total squared load (Min)	B&B	TP
	Meng et al. [17]	Simple	Cycle time (Min) Task alteration (Min)	MM, WOA	RLP
	Erten [18]	Simple	Number of agents (Min) Workload balancing (Min)	MM, SA	TP
	Tang et al. [19]	Simple	Cycle time (Min) Task alteration (Min)	MM, MFEA	TP
	Petroodia et al. [20]	Mixed-model	Cost (Min)	MM, FOH, CM	TP
	Yin et al. [21]	Partial disassembly	Cycle time (Min) Peak Energy Consumption (Min) Total Energy Consumption (Min) Hazardous index (Min)	MM, HDA	TP
	Meng et al. [22]	Mixed-model	Cycle time (Min) Task alteration (Min)	MM, CCEA	TP
	Zhao & Zhang [23]	-	Cycle time (Min) Task adjustments (Min)	IVNS	TP
	Deliktaş & Aydın [54]	Simple	Smoothness index (Min) Line efficiency (Max)	IABC, HH	TP
ALWAB problems	Aryanezhad et al. [24]	Cellular	Cost (Min)	MM	TP
	Blum & Miralles [25]	Simple	Cycle time (Min)	MM, BS	TP
	Sungur & Yavuz [26]	Simple	Cost (Min)	MM	TP
	Borba & Ritt [27]	Simple	Cycle time (Min)	B&B	TP
	Vilá & Pereira [28]	Simple	Cycle time (Min)	B&B	TP

		Number of workstations (Min)		
Ritt & Miralles [29]	Simple	Cycle time (Min)	MM, SA	TP
Polat et al. [30]	Simple	Cycle time (Min)	VNS	RLP
Zacharia & Nearchou [4]	Simple	Cycle time (Min) Smoothness index (Min)	EA	TP
Janardhanan & Nielsen [31]	Two-sided	Cycle time (Min)	MM, MBO	TP
Yılmaz & Demir [32]	Simple	Cycle time (Min)	MM	TP
Janardhanan et al. [33]	Two-sided	Cycle time (Min)	ABC	TP
Yıldız et al. [34]	Simple	Cycle time (Min) Number of workers (Min)	MM, Arena Simulation	RLP
Zhang et al. [35]	U-shaped	Cycle time (Min) Ergonomic risks (Min)	MM, RIPG, OCRA	TP
Karaş & Özçelik [36]	Simple	Weighted sum of the relative percent deviation from the lower bounds (Min)	MM, ABC	TP
Campanaa et al. [37]	Simple	Cost (Min)	MM, H, VNS	TP

Table 1. (Cont.) Related literature on the ALB and ALWAB problems.

Category	References	Assembly line type	Used objective functions	Solution methods	TP/RLP
ALWAB problems (Cont.)	Gräßler et al. [38]	Manual	Skill improvement (Max) Cycle time (Min)	MM	TP
	Katirae et al. [39]	Simple	Cycle time (Min) Physical effort (Min)	ϵ -constraint	RLP
	Küçükkoç [40]	Simple	Number of workstations (Min) Disparity between workstations (Min)	MM, Hybrid GA	TP
	Akpınar & Bayhan [41]	Parallel	Number of workstations (Min)	MM, GA	TP
	Moreira et al. [42]	Simple	Cycle time (Min)	MM, Hybrid GA	TP
	Mutlu et al. [9]	Simple	Cycle time (Min)	MM, IGA	TP
	Oksuz et al. [43]	U-type	Line efficiency (Max)	MM, GA, ABC	TP
	Fathi et al. [44]	Simple	Number of workstations (Min) Workload balancing (Min)	MM, GA, VNS	TP
Liu et al. [45]	Simple	Cost (Min) Energy consumption (Min)	MM, NSGA- II, SA	TP	
Kılınççı [46]	Simple	Number of workstations (Min)	MM, GA	TP	

Solution methods:

ABC: Artificial Bee Colony Algorithm, **ACO:** Ant Colony Optimization, **B&B:** Branch and Bound Algorithm, **CCEA:** Cooperative Co-evolutionary Algorithm, **CM:** Constructive Matheuristic, **EA:** Evolutionary Algorithm, **FOH:** Fix-and-optimize Heuristic, **H:** Heuristic, **HDA:** Hybrid Driving Algorithm, **HH:** Hyper-heuristic, **IABC:** Improved Artificial Bee Colony Algorithm, **IGA:** Iterative Enhanced Genetic Algorithm, **IVNS:** Improved Variable Neighborhood Search, **MFEA:** Multifactorial Evolutionary Algorithm, **MM:** Mathematical Model, **NSGA- II:** Non-dominated Sorting Genetic Algorithm-II, **OCRA:** Occupational Repetitive Action Tool, **RIPG:** Restarted Iterated Pareto Greedy Algorithm, **SA:** Simulated Annealing, **TS:** Tabu Search, **VNS:** Variable Neighborhood Search, **WOA:** Whale Optimization Algorithm
TP: Test Problem, **RLP:** Real-life problem

2.2. Problem description

This study was conducted at an automotive company that produces electrical equipment for passenger vehicles. Each passenger vehicle has its own electrical equipment. The company produces these electrical components for vehicles. The assembly area of the company consists of a line system, with each line composed of 20 conveyors and an assembly table for each conveyor based on the production quantity. Fig. 1 shows the conveyor system for a vehicle.

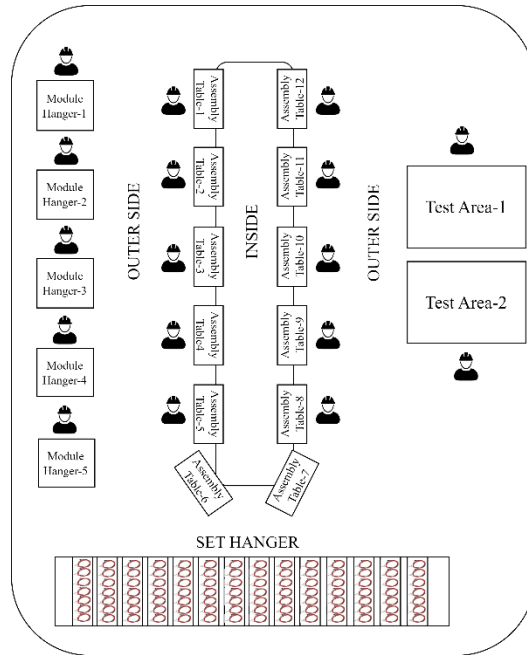


Fig. 1. An example of the assembly line in the cable industry.

The assembly tables on the conveyor are set to face outwards. Therefore, as can be seen in Fig. 1, they work on the outer side of the assembly tables and cannot enter the inner side. The conveyor system works in such a way that there is one operator at each assembly table. Each assembly table corresponds to a licence plate where the wiring of the respective car is addressed. Each vehicle has its own electrical equipment. This equipment includes different cable colours, cross-sections, lengths and routes. Accordingly, each worker has to perform the cable addressing on the assembly table, taking into account the tasks assigned to her/him. A sample section of the assembly table Fig. 2 shows a sample assembly table of the line where the ceiling installation equipment from the project groups is produced. In this line, there are cable addressing and harness taping processes on the conveyor. In the cable addressing process, each cable has an address (A1-A6, B1-B6 and C1-C5) and the cable routes to these addresses (purple and blue solid lines, red dashed lines) on the assembly table. In Fig. 2, the blue and purple solid lines and the red dashed line are the paths that the cables must follow. In order to ensure correct electrical conductivity, each cable must be taken from its own defined places.

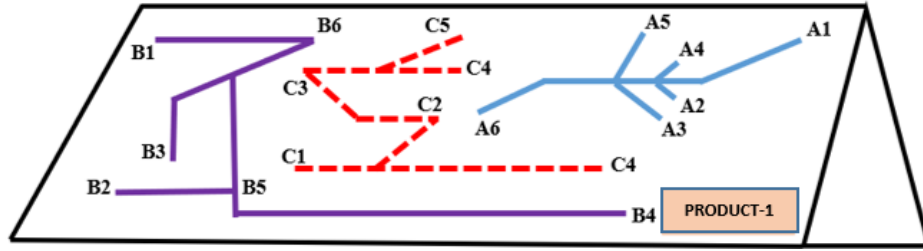


Fig. 2. An example of the assembly table in the cable industry.

In simple assembly line balancing problems, setup/transport (walking) times are usually ignored in studies since they are much smaller than the processing times, and in some cases, they are added to the processing times. Andres et al. [47] separately evaluated the setup times in the simple ALB problem and defined a new sequence-dependent assembly line balancing problem with setup times. No studies consider walking times in the ALWAB Type-2 problem type. In the considered enterprise, in order for each workstation to start working, the cable bundles in the set hangers must be transported to the workstations. The distance of each station to the set hanger where the cable bundles are located is different. The walking times here were not ignored, and a line balancing study was carried out by evaluating the station-based preparation times.

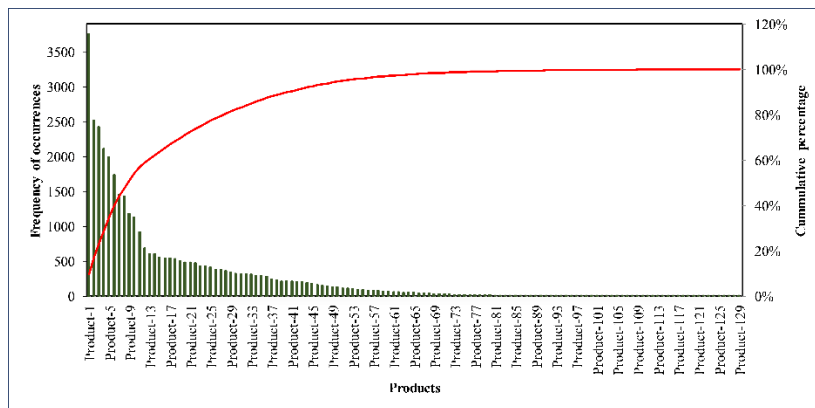


Fig. 3. Pareto analysis for the produced products in the last 6 months.

Pareto analysis, as a decision-making technique, statistically isolates a limited set of input factors, whether they are desirable or undesirable, that exert the most significant influence on the overall outcome. In order to determine the model type that will be the most produced according to the order rates received from the customer, a Pareto analysis was performed on the order quantity of 129 products for the last 6 months. As seen in Fig. 3, Product-1 is the product with the highest number of orders. The number of workers and workstations is determined for the problem to be balanced in the enterprise. The performance of all workers is different from each other. There are priority relations between the tasks. Product-1 consists of 48 jobs, and 8 workers work on the line to be assembled.

2.3. Mathematical model

The assumptions of the proposed mathematical model for the assembly line worker assignment and balancing problem are as follows:

- The processing times of each task of each worker are known deterministically.
- The processing times of each job vary depending on the worker.
- Each worker can do every job.
- The setup time of the worker at each station is included in the processing times.
- A worker cannot be assigned to more than one workstation.
- A task can be assigned to one and only one worker.
- Assignment of more than one worker to a workstation is not allowed.
- Precedence relationships between tasks exist, and precedence relationships are known.
- Each worker can perform more than one task.
- Stations can be assigned more than one task.
- There is a single type of product production.
- Tasks are indivisible.

Notations:

- J : Set of the workers, $j \in J = \{1, 2, \dots, m\}$
 I : Set of the jobs, $i, k \in I = \{1, 2, \dots, n\}$
 S : Set of the workstations, $s \in S = \{1, 2, \dots, p\}$
 m : The number of workers
 n : The number of jobs
 p : The number of workstations in the initial state
 D_i : The set of immediate predecessors of job i in the precedence network
 P_{ij} : Processing time of job i when worker j executes it
 R_s : Cable transport time for workstation s
 M : A large positive number

Decision variables:

- CT : Cycle time
 X_{sij} : 1, if the job i is assigned to the worker j in the workstation s ; 0, otherwise
 Y_{sj} : 1, if the worker j is assigned to the workstation s ; 0, otherwise

Multi-objective mathematical model:

$$\text{Min } f_1(x) = CT \tag{1}$$

$$\text{Min } f_2(x) = \sum_{s \in S} \sum_{j \in J} \left(\sum_{i \in I} ((P_{ij} \times X_{sij}) + R_s)^2 \right) \tag{2}$$

$$\sum_{s \in S} \sum_{j \in J} X_{sij} = 1, \quad \forall i \tag{3}$$

$$\sum_{j \in J} Y_{sj} \leq 1, \quad \forall s \tag{4}$$

$$\sum Y_{sj} \leq 1, \quad \forall j \tag{5}$$

$$\sum_{s \in S} \sum_{j \in J} s \times X_{sij} \leq \sum_{s \in S} \sum_{j \in J} s \times X_{skj}, \quad \forall i, k/i \in D_i \tag{6}$$

$$\sum_{i \in I} \left((P_{ij} \times X_{sij}) + R_s \right) \leq CT, \quad \forall s, j \tag{7}$$

$$\sum_{i \in I} X_{sij} \leq M \times Y_{sj}, \quad \forall j \tag{8}$$

$$M > \sum_{s \in S} \sum_{j \in J} \left((P_{ij} \times X_{sij}) + R_s \right), \quad \forall i \tag{9}$$

$$X_{sij}, Y_{sj} \in \{0,1\}, CT \geq 0, \forall i, j, i \neq j, s \tag{10}$$

While Eq. (1) aims to minimise the cycle time, Eq. (2) aims to balance the task loads by quadratic load assignment. Eq. (3) ensures that each operation is assigned to a worker and a workstation. Eq. (4) guarantees that each workstation has only one worker. Eq. (5) ensures that each worker is assigned to only one workstation, respectively. Eq. (6) defines the priority relationships between tasks. Eqs. (7)-(8) ensure that the sum of the assigned task processing and walking times do not exceed the cycle time. Eq. (9) states that each worker assigned to a workstation can perform more than one task as long as the cycle time is not exceeded, where M is a large enough constant. Eq. (10) specifies the decision variables. Since the problem considered has two objective functions, the objective functions are combined using the weighted-sum method (WSM). WSM is one of the most well-known methods for obtaining Pareto efficient solutions [48]. The mathematical model of WSM is formulated in Eq. (11).

$$Min WSM(x) = [w_1 \times CT] + \left[w_2 \times \sum_{s \in S} \sum_{j \in J} \left(\sum_{i \in I} \left((P_{ij} \times X_{sij}) + R_s \right)^2 \right) \right] \tag{11}$$

Here w_1 and w_2 represent the importance weights for CT and $\sum_{s \in S} \sum_{j \in J} \left(\sum_{i \in I} \left((P_{ij} \times X_{sij}) + R_s \right)^2 \right)$. Assuming $w_1 + w_2 = 1$, it will be in the form of $w_1, w_2 \geq 0$ and all the constraints in Eqs. (3)-(10) must be satisfied.

The average walking distances of each worker from the set hanger where the harnesses are located to the workstations are given in Table 2. These times depend on the distance between the workstation and the set hanger and do not change according to the job. The precedence relationship of each job is given in Fig. 4.

Table 2. Average walking distance from set hanger to workstations.

Workstation no	R_s	Workstation no	R_s
1	39.8	5	18.5
2	32.9	6	15.3
3	27.2	7	12.0
4	22.6	8	10.1

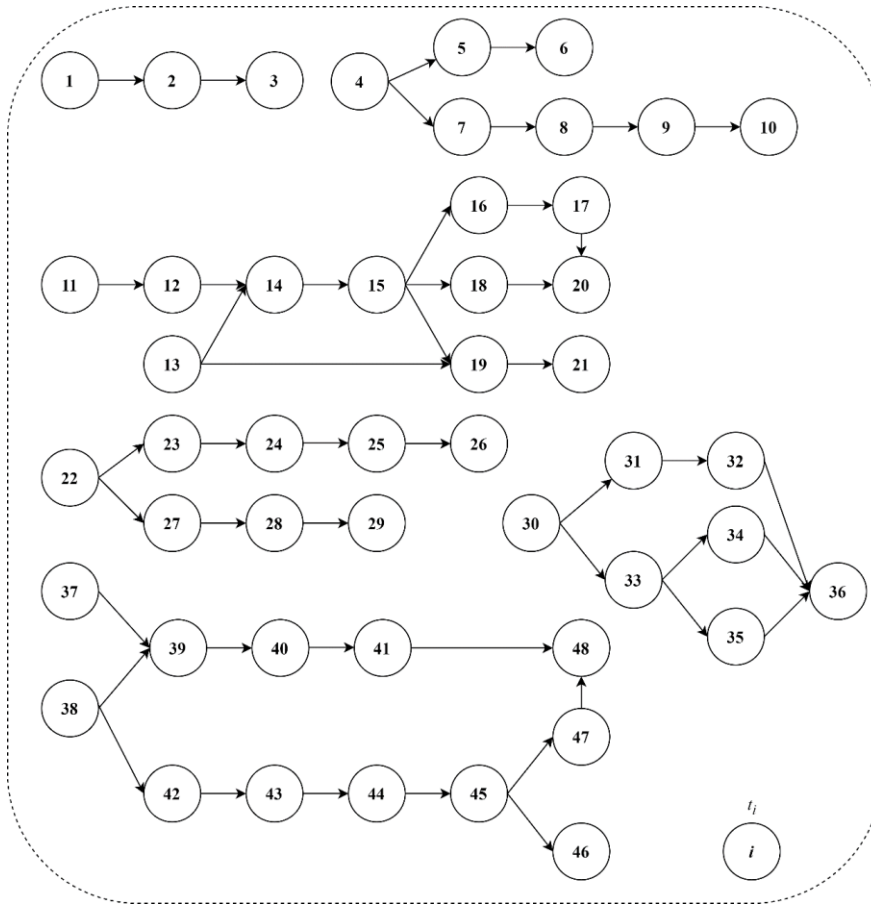


Fig. 4. Precedence relationship diagram for ceiling installation harness of Product-1.

2.4. Multi-objective genetic algorithm

Due to the NP-hardness of the ALWAB Type-2 problem and the problem size limitation of exact solution methods, approximation procedures are needed to solve the problem [49]. Therefore, (meta)heuristic procedures have been developed for solving the ALWAB Type-2 problem. The pseudo-code of the proposed algorithm is given in Table 3.

Table 3. Pseudo-code of the proposed algorithm.

Algorithm 1: Pseudo-code of the proposed algorithm

Input: $I, J, S, D_j, P_{ij}, R_s, w_j, w_2, Tournament_{size}, C_{rate}, M_{rate}, UD_{Max}$

Output: Job-workstation-worker matrix, $f_1(x), f_2(x), WSM^{Norm}(x)$

```

1: begin
2: Randomly generate an initial population ( $Pop_{init}$ )
3: Compute of the fitness values ( $Pop_{init}$ ) ▷ see Eq. (17)
4:  $Pop_{current} \leftarrow Pop_{init}$ 
5: while termination criteria not satisfied do
6:   Parents  $\leftarrow$  Tournament selection ( $Pop_{current}, Tournament_{size}$ )
7:   Childs  $\leftarrow$  WMX (Parents)

```

```

7:      Popcurrent ← Swap (Childs)
8:      Calculate of the fitness values ← (Popcurrent) ▷ see Eq. (17)
11:    end while
12:    return the best individual in Popcurrent
13:  end

```

This study computes the cycle time by using the modified bisection search method proposed by Mutlu et al. [9]. Unlike the studies in the literature, walking times differ according to the distance between workstations. In order to achieve a balanced line distribution, walking times according to the station distance were added to the method. As a result, the bisection search method was modified, and the temporary cycle time was calculated.

The lower bound (C_{LB}) is calculated as given in Eq. (12), and the upper bound (C_{UB}) as given in Eq. (13) [50]:

$$C_{LB} = \text{Max} \left\{ \left(\frac{1}{S} \sum_{i \in I} \text{Min}(P_{ij}), \forall j \in J \right), \left(\text{Min}(P_{ij}), \forall j \in J \vee \forall i \in I \right) \right\} \quad (12)$$

$$C_{UB} = \text{Max} \left\{ \left(\frac{1}{S} \sum_{i \in I} \text{Max}(P_{ij}), \forall j \in J \right), \left(\text{Max}(P_{ij}), \forall j \in J \vee \forall i \in I \right) \right\} \quad (13)$$

In this study, the modified bisection search method used the following equations to calculate the expected cycle time (see Eqs. (14)-(15)):

$$C_m = \frac{1}{S} \sum_{i \in I} \text{avg}(P_{ij}), \forall j \in J \quad (14)$$

$$C_e = \frac{1}{6} (C_{LB} + 4 \times C_m + C_{UB}) + \text{avg}(R_s) \quad (15)$$

where $\text{avg}(P_{ij})$ and $\text{avg}(R_s)$ represent the average of processing and walking times, respectively.

The initial population is created in the initialization phase. The chromosomes in the population are randomly generated between 1 and the number of jobs (n). Every job is denoted by a numerical value, which is positioned within a sequence of numbers referred to as chromosomes. The length of each chromosome corresponds to the total number of jobs, and the value of each gene within the chromosome signifies the specific job's position in the assembly line. The jobs are listed sequentially in accordance with their assigned order. Fig. 5 represents an example of solution string corresponding to a data set consisting 11 jobs.

Job Number	1	2	3	4	5	6	7	8	9	10	11
Job Priority	8	10	7	5	2	6	11	3	9	1	4

Fig. 5. Chromosome representation.

Workers are also assigned to workstations. The worker numbers given in the representation are assigned to workstations, respectively. An example representation is given in Fig. 6.

Workstation Number	1	2	3
Worker Number	2	3	1

Fig. 6. The initial worker assignment of the example problem.

In the proposed multi-objective genetic algorithm (MOGA), the job assignment is made by considering the precedence diagram between jobs according to the determined job priority values, initial chromosome and worker assignment results. In order for the worker to start working, s/he needs to fetch the cable bundles from the set hanger. The cable retrieval time defined for the first station is added, and jobs are started to be assigned to the relevant station, taking into account the ability of the assigned worker to do the jobs and the duration of the job. If there is a possibility of selecting more than one job according to the priority status, the job with the highest priority number initially assigned by the algorithm is selected. Jobs continue to be assigned to the first workstation until the specified cycle time is reached. In the solution of proposed algorithm, to perform the optimum assignment process, only the last job of the assigned workstation is ignored, and if other jobs can be assigned, they are assigned. After the jobs that can be assigned to the first workstation are finished, we move to the next workstation and repeat the process described above.

In this study, the proposed MOGA aims to minimise the cycle time and the squared load assignment. When assigning the jobs to the workstation according to the specified cycle time (CT_{opt}), all the remaining jobs are assigned to the last station according to the order of priority since each task needs to be done. This may not be a feasible solution as the last station is assigned without considering the cycle time constraint. The fitness function applied in study of Mutlu et al. [9] prevents the station time of the last station from exceeding the cycle time via a high penalty in given in Eq. (16). In this equation, FS denotes the time by which the station time of the last station exceeds the cycle time.

$$Fitness\ value = CT_{opt}^2 \times FS \quad (16)$$

A group of individuals is randomly selected from the population, and the individual with the highest fitness value is selected to be the parent of the next generation population, and the tournament size is taken as 2. In each generation, the selection method, crossover and mutation operators are applied to ensure genetic diversity in the population. In the proposed genetic algorithm, a two-point-based weight mapping crossover (WMX) is applied as a crossover operator. The applied WMX operator consists of four primary steps: (i) the selection of a random sub-vector from two randomly selected parents (see Step 1), (ii) the prioritization of jobs in ascending order based on their assigned priorities, with lower numbers indicating higher-priority jobs (see Step 2), (iii) the exchange of ranks between the selected sub-vectors and the subsequent rearrangement of priorities according to the new rankings (see Step 3), and (iv) the generation of offspring using the newly determined job priorities from Step 3 (see Step 4). An illustrative example for the applied WMX operator is shown in Fig. 7 [44].

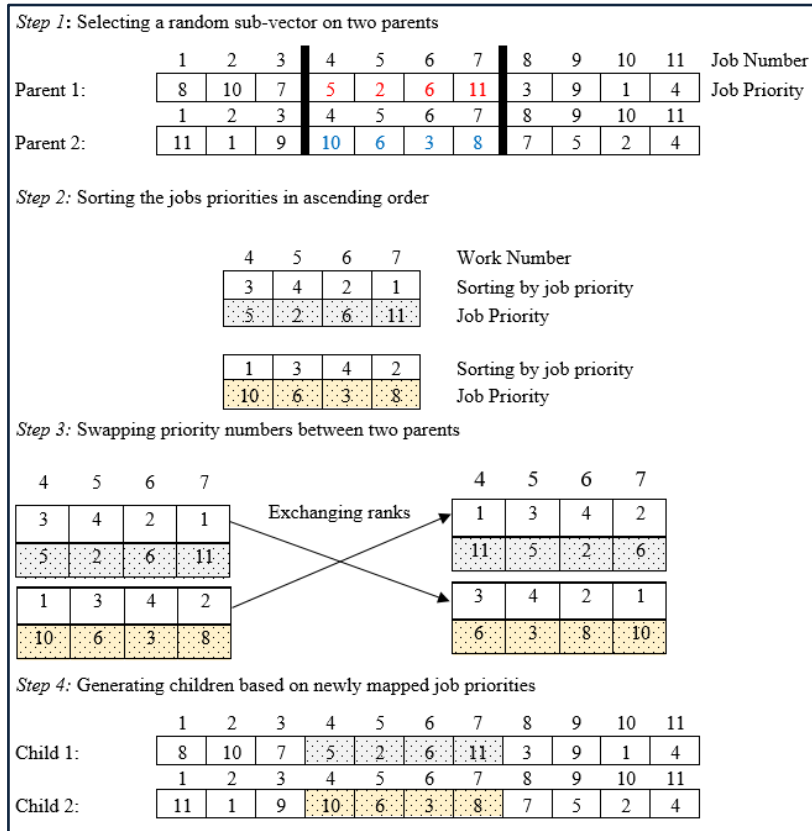


Fig. 7. Representation of the proposed WMX operator.

Swap-mutation (S_{wap}) operator was applied in the proposed genetic algorithm to ensure diversity. For the crossover rate (C_{rate}) and mutation rate (M_{rate}), the values determined from parameter tuning are used. The fitness value of the new population ($Pop_{current}$) is calculated. After the stopping condition is met, the best individual in the current population constitutes the output of the algorithm.

2.5. Parameter tuning

The selection of appropriate parameters is pivotal for genetic algorithms to achieve optimal solutions. Ensuring the correct parameters for a given problem can profoundly influence the results. By making fine adjustments to these parameters, algorithms often demonstrate enhanced performance, leading to improved results [51]. These parameters are preset at the algorithm’s onset based on the specifics of the problem at hand. The critical parameters to be determined include the population size, the crossover rate, and the mutation rate. In this study, the parameters for the proposed MOGA were ascertained using both the **irace** and design of experimental (DoE) methods. The real-life problem was then addressed using the suitable parameters derived from each approach.

The algorithm was executed for the problem under consideration with objective weights $w_1=0.5$ (representing the weight of the cycle time objective function) and $w_2=0.5$ (for the weight of the squared load assignment objective function). The algorithm was run 31 times with objective weights set at $w_1=0.0$ ($w_2=1.0$) to obtain the nadir value associated with the cycle time objective. Similarly, the algorithm was initiated 31 times with objective weights configured at $w_1=1.0$ ($w_2=0.0$) to pinpoint the nadir value for the squared load assignment objective function.

Given the diverse magnitudes present among the objective values, it is essential to normalise these values by dividing the corresponding objective function by its nadir value [52]. When scaled using the weighted sum method, the fitness function is normalised as outlined in Eq. (17). In this equation, N_i denotes the nadir point of the i th objective function ($i = 1,2$).

$$Min WSM^{Norm}(x) = \left[w_1 \times \frac{CT}{N_1} \right] + \left[w_2 \times \frac{\sum_{s \in S} \sum_{j \in J} \left(\sum_{i \in I} \left((P_{ij} \times X_{sij}) + R_s \right)^2 \right)}{N_2} \right] \tag{17}$$

Additionally, the obtained results were executed using the determined parameters for the genetic algorithm used in the study of Mutlu et al. [9] and the obtained results were compared with the obtained results of the **irace** and DoE methods. For the **irace** method, the parameter ranges used for each genetic operator (population size, crossover rate, and mutation rate) are presented in Table 4.

Table 4. Ranges of the parameters for the proposed MOGA and the best parameter configuration obtained with the **irace** method.

Factors	Parameter Values	Selected Parameter ($w_1=0.5 ; w_2=0.5$)
Population Size	U ~ [20 ; 200]	35
Crossover Rate	U ~ [0.7 ; 1]	0.87
Mutation Rate	U ~ [0.05 ; 0.4]	0.06

The stopping criterion for the algorithm has been considered based on the maximum fitness value (UD_{Max}). This value is obtained by multiplying the number of tasks by a fixed number. As seen the convergence graph given in Fig. 8, this constant number is determined to be 10,000 after various trials.

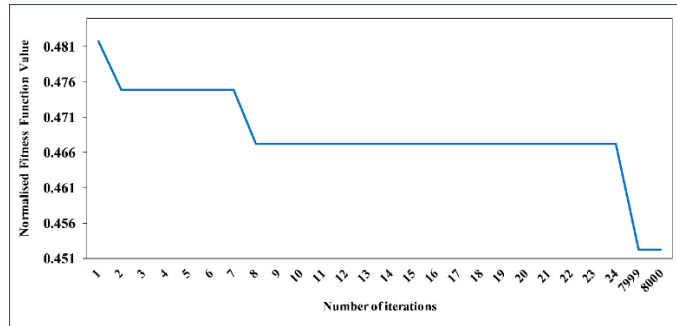


Fig. 8. Convergence graph for the normalized objective function values.

A design of experimental was implemented to determine the appropriate parameter values for genetic operators. Accordingly, the levels associated with the considered operators are given in Table 5.

Table 5. Factors and levels of factors.

Factors	Levels		
	Level 1	Level 2	Level 3
Population Size	20	60	100
Crossover Rate	0.70	0.85	1.00
Mutation Rate	0.05	0.10	0.40

The proposed MOGA, with a stopping condition and tournament size similar to those considered in the **irace** method, was analysed using Minitab 21.0 software for each determined factor level. In the DoE, due to 27 different combinations and each experiment being run 31 times, a total of $27 \times 31 = 837$ experiments were conducted.

Accordingly, an analysis of variance (ANOVA) was carried out to determine the critical parameters and their interactions. In the variance analysis, the main effect is a value that indicates the degree of a factor’s impact on the result. The main effects plot is used to examine the difference between the level averages of the factors [53]. According to the analysis results shown in Fig. 9, the *p*-values related to population size, crossover rate, and mutation rate are below 0.05 with a 95% confidence interval. Hence, these factors have been selected as critical factors in minimising the average normalised results.

Analysis of Variance					
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	10	0,000307	0,000031	111,86	0,000
Linear	6	0,000289	0,000048	175,71	0,000
Population_size	2	0,000220	0,000110	400,72	0,000
Crossover_rate	2	0,000003	0,000001	5,01	0,020
Mutation_rate	2	0,000067	0,000033	121,40	0,000
2-Way Interactions	4	0,000018	0,000004	16,08	0,000
Population_size*Mutation_rate	4	0,000018	0,000004	16,08	0,000
Error	16	0,000004	0,000000		
Total	26	0,000311			

Model Summary			
S	R-sq	R-sq(adj)	R-sq(pred)
0,0005239	98,59%	97,71%	95,98%

Fig. 9. The ANOVA result based on coded values.

Upon examining the main effect graph, the levels of the critical factors are determined. The number of experiments conducted is reflected in Figs. 10(a) and 10(b), where both the main effects and the interaction plots are displayed.

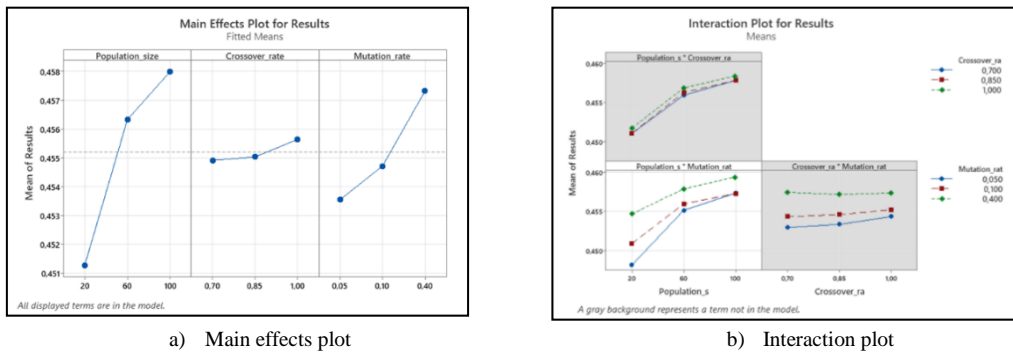


Fig. 10. The main effects and interaction plots for the proposed MOGA parameters.

According to the results obtained, the best values for the critical factors of population size, crossover rate, and mutation rate have been determined as 20 (Level 1), 0.7 (Level 1), and 0.05 (Level 1), respectively.

2.6. Comparative analysis

In this subsection, for $w_1=0.5$ and $w_2=0.5$ objective weights, the results obtained from the appropriate parameters via the DoE are compared with the results obtained from the appropriate parameter values via the **irace** method, and the results obtained using the appropriate parameters suggested by Mutlu et al. [9]. The results obtained are

displayed in a box plot in Fig. 11.

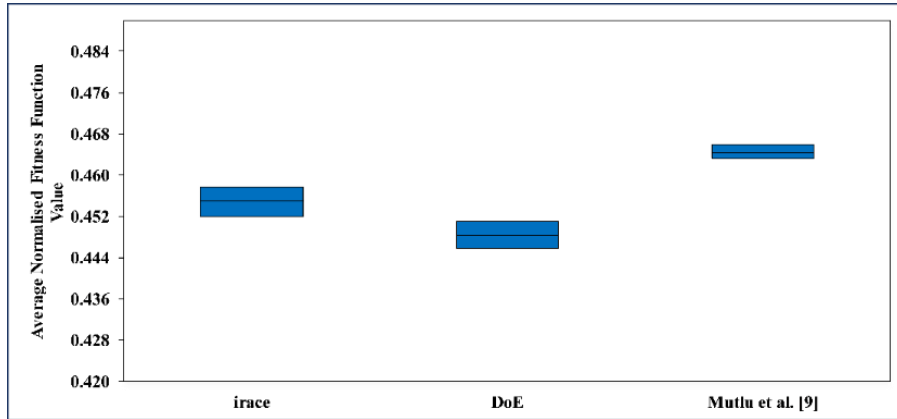


Fig. 11. Box plots of average normalised fitness values obtained from 31 runs of the MOGA with weighted-sum scalarisation for three different parameter determination methods.

The box plot shows that the average normalised fitness values obtained through the DoE have yielded better results than others. The results obtained with the **irace** method are more successful than the results achieved with the parameters proposed by Mutlu et al. [9].

3. Conclusions

Taking into account the task times specific to workers and the walking times of workers, the cycle time was intended to be optimised along with aiming for a balanced distribution of tasks assigned to workstations. For our proposed multi-objective genetic algorithm, the parameter selection and calibration were carried out using the full factorial experimental design and the **irace** method, one of the automatic parameter determination methods, considering the objective weights of $w_1=0.5$ and $w_2=0.5$. The real-life problem was resolved using the suggested MOGA approach with the identified suitable parameters. The worker assignments, task allocations, and station times for the best result among the 31 repetitions are presented in Table 6.

Table 6. Comparison of results of the proposed MOGA using appropriate parameters obtained from three different ways with the current situation of the cable industry.

WS No	Parameter Method	Worker No	Assigned Jobs								WS Time
WS 1	DoE	Worker-7	30	4	33	11	13	34	1	37	159.0
	Mutlu et al. [9]	Worker-6	13	4	7	8	30	33	35	38	174.3
	irace	Worker-3	38	11	4	7	37				147.3
	CS	Worker-1	1	2	3	4	5	6	7	30	175.3
WS 2	DoE	Worker-6	38	42	43	44	45	35	5		177.2
	Mutlu et al. [9]	Worker-7	42	43	11	34	1	37			178.1
	irace	Worker-7	8	42	1	30	33	34	31	13	178.5
	CS	Worker-2	8	9	10	31	33	35	37		163.5
WS 3	DoE	Worker-5	12	14	22	23	24	47			174.4
	Mutlu et al. [9]	Worker-8	39	22	23	24	40				175.7
	irace	Worker-6	32	35	36	22	27	28			173.6
	CS	Worker-3	11	12	13	14	15	16	18	19	22
WS 4	DoE	Worker-4	25	7	8	46	2	3	15		178.8
	Mutlu et al. [9]	Worker-3	41	25	44						139.2
	irace	Worker-4	29	23	24	25	9	43			179.0

	CS	Worker-4	17	20	23						169.9
WS 5	DoE	Worker-2	31	18	26	27	28	29	16	19	150.6
	Mutlu et al. [9]	Worker-2	45	47	46	26	48	2	31	5	175.6
	irace	Worker-5	44	45	47	46	10	12			154.7
	CS	Worker-5	21	24	27	25					182.2
WS 6	DoE	Worker-3	21	39	40	41					177.3
	Mutlu et al. [9]	Worker-5	32	36	6	27	28	29	12		178.3
	irace	Worker-2	14	15	18	19	21	26	5	16	170.3
	CS	Worker-6	2	34	28	29	38	36			177.6
WS 7	DoE	Worker-1	17	20	9	48					162.1
	Mutlu et al. [9]	Worker-4	14	15	16	19	21	18	17	9	174.4
	irace	Worker-8	17	6	20	2					170.2
	CS	Worker-7	26	39	42	43					182.9
WS 8	DoE	Worker-8	6	10	32	36					135.8
	Mutlu et al. [9]	Worker-1	3	10	20						140.5
	irace	Worker-1	3	39	40	41	48				156.4
	CS	Worker-8	40	41	44	45	45	47	48		174.7

Abbreviation(s):

CS: Current Situation, WS: Workstation

The objective values obtained based on the parameter determination methods used in Table 7 and the current situation are provided. Accordingly, the cycle time obtained from the parameters of Mutlu et al. [9] performs better than others. However, the squared load assignment obtained from the parameters of the DoE provides better performance than others.

Table 7. Comparison of the results of the proposed MOGA run using parameter values obtained in three different ways with the current situation in the cable industry.

Parameters	Cycle Time	Squared Load Assignment
DoE	178.8	217899.34
Mutlu et al. [9]	178.3	225129.09
irace	179.0	222100.68
Current Situation	182.9	245959.40

In this study, the effect of parameterisation methods on the results when run with different objective weights is analysed. The sensitivity analysis, as depicted in Fig. 12, indicates that the experimental design method for parameter determination with objective weights of $w_1=0.7$ and $w_2=0.3$ yields better results for all objective weights. However, it exhibits poorer performance when compared with the parameters proposed by Mutlu et al. [9]. Additionally, results obtained using the **irace** method for $w_1=0.7$ and $w_2=0.3$ objective weights have proven to be better. These findings show that the experimental design generally yields better outcomes across various objective weights when compared to other methods.

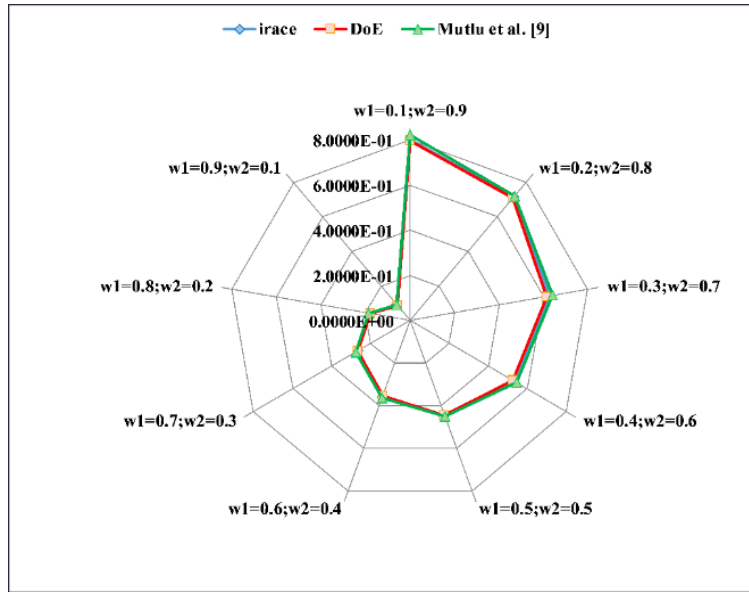


Fig. 12. Sensitivity analysis for the different objective weights.

While performing assembly line balancing, it should not be ignored that the characteristics and performances of the workers are different. It is accepted that each worker has different capabilities for the same task, and it is decided that worker assignments should also be made while performing line balancing.

This study considers the multi-objective assembly line worker assignment and balancing problem (ALWAB). The algorithm aims to minimise both the cycle time and the squared load assignment, striving for a balanced distribution of jobs assigned to the workstation. This is achieved by considering the precedence diagrams between jobs, the times of jobs for each worker, and the walking times of the workers. The proposed multi-objective algorithm is applied to a real-life problem in an automotive sub-industry that manufactures wiring harnesses. The product to be assembled in the plant belongs to the roof wiring harness of a vehicle, which consists of 48 jobs and 8 workers are assigned to this process. Since the problem is multi-objective and NP-hard, a multi-objective genetic algorithm (MOGA) approach is proposed to solve the problem. Since the parameter values used in the algorithm directly affect the result, it is of great importance to use appropriate values. A full factorial experimental design and the **irace** method, one of the automatic parameter determination methods, were proposed for the parameter calibration of the proposed MOGA. With the appropriate parameters found as a result of these two methods, the proposed MOGA was run for objective weights $w_1=0.5$ and $w_2=0.5$. In addition, the algorithm was also run with the parameters used in the study of Mutlu et al. [9], and the results obtained were compared. It was found that the full factorial experimental design performed better than the others. In addition, the performance of the parameterisation methods with different objective weights was tested by sensitivity analysis. According to the results, it was observed that the experimental design generally performed better than the others at different objective weights.

In future directions, the proposed algorithm can be solved by taking into account that not every worker is capable of every job. The performance of the proposed algorithm can be tested with benchmark problems in the literature, and the results can be compared. The problem can be extended by adding different objective functions. Also, ergonomic risk factors can be added to the ALWAB Type-2 problem. A decision support system with a user-friendly interface that can be used by the employees of the organisation can be designed.

Acknowledgement

The authors thank the reviewers for their valuable comments and suggestions, which improved the clarity and scope of the article.

Ethical Approval

This study was carried out as part of the master thesis [ID: 780444] prepared by Gözde KURADA under the supervision of Assoc. Prof. Dr. Derya DELİKTAŞ.

References

- [1] Kobu, B., Üretim yönetimi (17. baskı). İstanbul: Beta Yayınevi, 2014
- [2] Boysen, N., Flidner, M. and Scholl, A., A classification of assembly line balancing problems. *European journal of operational research*, 183(2), 674-693. URL: <https://doi.org/10.1016/j.ejor.2006.10.010>, 2007.
- [3] Okyay, Y. R., Solving assembly line balancing problem with positional constraints and worker assignments using mathematical programming and heuristic solution approaches. Yayınlanmamış Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi Fen Bilimleri Enstitüsü, İzmir, 2018
- [4] Zacharia, P. T. and Nearchou, A. C., A population-based algorithm for the bi-objective assembly line worker assignment and balancing problem. *Engineering Applications of Artificial Intelligence*, 49, 1-9. URL: <https://doi.org/10.1016/j.engappai.2015.11.007>, 2016
- [5] Çengil, M.F., Heuristic approaches for assembly line balancing and competent worker assignment problem (ALWABP). Yayınlanmamış Yüksek Lisans Tezi, Özyeğin Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 2018
- [6] Becker, C. and Scholl, A., A survey on problems and methods in generalized assembly line balancing. *European Journal of Operational Research*, 168(3), 694-715. URL: <https://doi.org/10.1016/j.ejor.2004.07.023>, 2006
- [7] Karp, R. M., Reducibility among combinatorial problems. In R. E. Miller, J. W. Thatcher and J. D. Bohlinger (Eds.), *Complexity of computer computations*. New York: Plenum Press, pp. 85-103, 1972
- [8] Doğan, A. and Sakallı, Ü.S., Bulanık işlem zamanlı geleneksel montaj hattı dengeleme problemi için yeni bir yaklaşım: Savunma sanayii uygulaması. *Kırıkkale Üniversitesi Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi*, 8(1), 31-50, 2016
- [9] Mutlu, Ö., Polat, O. and Supçiller, A.A., An iterative genetic algorithm for the assembly line worker assignment and balancing problem of type-II. *Computers & Operations Research*, 40(1), 418-426. URL: <https://doi.org/10.1016/j.cor.2012.07.010>, 2013
- [10] Tinç, H.K., An Algebraic Approach to Sensitivity Analysis in Linear Programming. Unpublished Master's Thesis, İstanbul Technical University Institute of Science and Technology, İstanbul, 2007
- [11] Karsu, Ö. and Azizoğlu, M., Bicriteria multiresource generalized assignment problem. *Naval Research Logistics (NRL)*, 61(8), 621-636. URL: <https://doi.org/10.1002/nav.21607>, 2014a
- [12] Altunay, H., Özmütlu, H.C. and Özmütlu, S., Paralel görev atamalı montaj hattı dengeleme problemi için yeni bir matematiksel model önerisi. *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 18(1), 15-33, 2017
- [13] Arıkan, M., İş yükü dengelemeli ikinci tip basit montaj hattı dengeleme problemi için bir tabu arama algoritması. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 32(4), 1169-1179. URL: <https://doi.org/10.17341/gazimmfd.369529>, 2017
- [14] Delice, Y., Aydoğan, E. K., Söylemez, I. and Özcan, U., An ant colony optimisation algorithm for balancing two-sided U-type assembly lines with sequence-dependent set-up times. *Computer Science, Sādhanā*, 43, 1-15. URL: <https://doi.org/10.1007/s12046-018-0969-9>, 2018
- [15] Stür, G.A., Sinaki, R.Y. and Sadeghi, A., A hierarchical hybrid heuristic optimization approach for multi-product assembly line design problem. *Procedia Manufacturing*, 39, 1067-1075. URL: <https://doi.org/10.1016/j.promfg.2020.01.368>, 2019
- [16] Karsu, Ö. and Azizoğlu, M., An exact algorithm for the minimum squared load assignment problem. *Computers & Operations Research*, 106, 76-90. URL: <https://doi.org/10.1016/j.cor.2019.02.011>, 2019b
- [17] Meng, K., Tang, Q., Zhang, Z. and Qian, X., An improved lexicographical whale optimization algorithm for the type-II assembly line balancing problem considering preventive maintenance scenarios. *IEEE Access*, 8, 30421-30435. URL: 2020
- [18] Erten, K., Darboğaz geliştirilmiş atama problemi için bir çözüm yaklaşımı. Yayınlanmamış Yüksek Lisans Tezi, Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü, Eskişehir, 2021
- [19] Tang, Q., Meng, K., Cheng, L. and Zhang, Z., An improved multi-objective multifactorial evolutionary algorithm for assembly line balancing problem considering regular production and preventive maintenance scenarios. *Swarm and Evolutionary Computation*, 68(3), 101021. URL: <https://doi.org/10.1016/j.swevo.2021.101021>, 2022
- [20] Hashemi-Petroodi, S. E., Thevenin, S., Kovalev, S. and Dolgui, A., Model-dependent task assignment in multi-manned mixed model assembly lines with moving workers. *Omega*, 113(12), 1-46. URL: <https://doi.org/10.1016/j.omega.2022.102688>, 2021
- [21] Yin, T., Zhang Z., Zhang, Y., Wu, T. & Liang, W., Mixed-integer programming model and hybrid driving algorithm for multi-product partial disassembly line balancing problem with multi-robot workstations. *Robotics and Computer-Integrated Manufacturing*, 73, 102251. URL: <https://doi.org/10.1016/j.rcim.2021.102251>, 2022
- [22] Meng, K., Thang, Q., Cheng, L. and Zhang, Z., Mixed-model assembly line balancing problem considering preventive maintenance scenarios: MILP model and cooperative co-evolutionary algorithm. *Applied Soft Computing*, 127, 109341. URL: <https://doi.org/10.1016/j.asoc.2022.109341>, 2022

- [23] Zhao, L., Tang Q. and Zhang, Z., An improved preference-based variable neighborhood search algorithm with *ar*-dominance for assembly line balancing considering preventive maintenance scenarios. *Engineering Applications of Artificial Intelligence*, 109, 104593. URL: <https://doi.org/10.1016/j.engappai.2021.104593>, 2022
- [24] Aryanezhad, M.B., Deljoo, V. and Mirzapour Al-e-hashem S.M.J., Dynamic cell formation and the worker assignment problem: a new model. *The International Journal of Advanced Manufacturing Technology*, 41(3), 329-342. URL: <https://doi.org/10.1007/s00170-008-1479-4>, 2009
- [25] Blum, C. and Miralles, C., On solving the assembly line worker assignment and balancing problem via beam search. *Computers & Operations Research*, 38(1), 328-339. URL: <https://doi.org/10.1016/j.cor.2010.05.008>, 2011
- [26] Sungur, B. and Yavuz, Y., Assembly line balancing with hierarchical worker assignment. *Journal of Manufacturing Systems*, 37(1), 290-298. URL: <https://doi.org/10.1016/j.jmsy.2014.08.004>, 2014
- [27] Borba, L. and Ritt, M., A heuristic and a branch-and-bound algorithm for the assembly line worker assignment and balancing problem. *Computers & Operations Research*, 45, 87-96. URL: <https://doi.org/10.1016/j.cor.2013.12.002>, 2014
- [28] Vilà, M. and Pereira, J., A branch-and-bound algorithm for assembly line worker assignment and balancing problems. *Computers & Operations Research*, 44, 105-114. URL: <https://doi.org/10.1016/j.cor.2013.10.016>, 2014
- [29] Ritt, M., Costa, A.M. and Miralles C., The assembly line worker assignment and balancing problem with stochastic worker availability. *International Journal of Production Research*, 54(3), 907-922. URL: <https://doi.org/10.1080/00207543.2015.1108534>, 2015
- [30] Polat, O., Kalayci, C.B., Mutlu, Ö. and Gupta, S. A two-phase variable neighbourhood search algorithm for assembly line worker assignment and balancing problem type-II: An industrial case study. *International Journal of Production Research*, 54(3), 722-741. URL: <https://doi.org/10.1080/00207543.2015.1055344>, 2016
- [31] Janardhanan, M.N., Li, Z. and Nielsen, P., Model and migrating birds optimization algorithm for two-sided assembly line worker assignment and balancing problem. *Soft Computing*, 23(21), 11263-11276. URL: <https://doi.org/10.1007/s00500-018-03684-8>, 2019
- [32] Yılmaz, H. and Demir, Y., A new mathematical model for assembly line worker assignment and balancing. *Journal of the Institute of Science and Technology*, 9(4), 2002-2008. URL: <https://doi.org/10.21597/jist.579958>, 2019
- [33] Janardhanan, M.N., Li, Z., Nielsen, P. and Tang, Q., Artificial bee colony algorithms for two-sided assembly line worker assignment and balancing problem. In *Distributed Computing and Artificial Intelligence*, 14th International Conference, *Advances in Intelligent Systems and Computing* 620, Springer, Cham. URL: https://doi.org/10.1007/978-3-319-62410-5_2, 2018
- [34] Topaloğlu, Yıldız, Ş., Yıldız, G. & Cin, E., Bir elektronik firmasındaki işçi atamalı montaj hattı dengeleme problemine matematiksel programlama ve benzetim modelleme tabanlı bir çözüm yaklaşımı. *Afyon Kocatepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 22(1), 57-73. URL: <https://doi.org/10.33707/akuiibfd.645402>, 2020
- [35] Zhang, Z., Tang, Q., Ruiz, R. and Zhang, L., Ergonomic risk and cycle time minimization for the U-shaped worker assignment assembly line balancing problem: A multi-objective approach. *Computers and Operations Research*, 118(2), 104905. URL: <https://doi.org/10.1016/j.cor.2020.104905>, 2020
- [36] Karas, A. and Ozcelik, F., Assembly line worker assignment and rebalancing problem: A mathematical model and an artificial bee colony algorithm. *Computers & Industrial Engineering*, 156, 107195. URL: <https://doi.org/10.1016/j.cie.2021.107195>, 2021
- [37] Campana, N.P., Iori, M. and Moreira, C.O., Mathematical models and heuristic methods for the assembly line balancing problem with hierarchical worker assignment. *International Journal of Production Research*, 60(7), 2193-2211. URL: <https://doi.org/10.1080/00207543.2021.1884767>, 2021
- [38] Gräßler, I., Roeasmann, D., Cappelo, C. and Steffen, E., Skill-based worker assignment in a manual assembly line. *Procedia CIRP*, 100, 433-438. URL: <https://doi.org/10.1016/j.procir.2021.05.100>, 2021
- [39] Katiraei, N., Calzavara, M., Finco, S. and Battini, D., Consideration of workforce differences in assembly line balancing and worker assignment problem. *IFAC PapersOnLine*, 54(1), 13-18. URL: <https://doi.org/10.1016/j.ifacol.2021.08.002>, 2021
- [40] Küçükkoç, İ., Karşık modelli montaj hattı dengeleme problemleri ve genetik algoritmalar ile bir uygulama. *Yayımlanmamış Yüksek Lisans Tezi, Balıkesir Üniversitesi Fen Bilimleri Enstitüsü, Balıkesir*. URL: <https://hdl.handle.net/20.500.12462/2560>, 2011
- [41] Akpınar, S. and Bayhan, G.M., A hybrid genetic algorithm for mixed model assembly line balancing problem with parallel workstations and zoning constraints. *Engineering Applications of Artificial Intelligence*, 24(3), 449-457. URL: <https://doi.org/10.1016/j.engappai.2010.08.006>, 2011
- [42] Moreira, M.C.O., Ritt, M., Costa, A.M. and Chaves, A.A., Simple heuristics for the assembly line worker assignment and balancing problem. *Journal of Heuristics*, 18, 505-524. URL: <https://doi.org/10.1007/s10732-012-9195-5>, 2012
- [43] Oksuz, M. K., Buyukozkan, K. and Satoglu, S. I., U-shaped assembly line worker assignment and balancing problem: A mathematical model and two metaheuristics. *Computers & Industrial Engineering*, 112, 246-263. URL: <https://doi.org/10.1016/j.cie.2017.08.030>, 2017
- [44] Fathi, M., Nourmohammadi, A., HC Ng, A., Syberfeldt, A. and Eskandari, H., An improved genetic algorithm with variable neighborhood search to solve the assembly line balancing problem. *Engineering Computations*, 37(2), 501-521. URL: <https://doi.org/10.1108/EC-02-2019-0053>, 2019
- [45] Liu, R., Liu, M., Chu, F., Zheng, F. and Chu, C., Eco-friendly multi-skilled worker assignment and assembly line balancing problem. *Computers & Industrial Engineering*, 151(2), 106944. URL: <https://doi.org/10.1016/j.cie.2020.106944>, 2021
- [46] Kılınççı, Ö., Assembly line balancing problem with resource and sequence-dependent setup times (ALBPRS). *Journal of the Faculty of Engineering and Architecture of Gazi University*, 38(1), 557-570. URL: <https://doi.org/10.17341/gazimmfd.757276>, 2022
- [47] Andres, C., Miralles C. and Pastor, R., Balancing and scheduling tasks in assembly lines with sequence-dependent setup times. *European Journal of Operational Research*, 187 (3), 1212-1223. URL: <https://doi.org/10.1016/j.ejor.2006.07.044>, 2008
- [48] Miettinen, K., Survey of methods to visualize alternatives in multiple criteria decision-making problems. *OR Spectrum*. 36(1), 3-37. URL: <https://doi.org/10.1007/s00291-012-0297-0>, 2014
- [49] Miralles, C., Garcia-Sabater, J. P., Andrés, C. and Cardós, M., Branch and bound procedures for solving the assembly line worker assignment and balancing problem: application to sheltered work centers for disabled. *Discrete Applied Mathematics*, 156(3), 352-367. URL:

<https://doi.org/10.1016/j.dam.2005.12.012> , 2008

[50] Scholl, A and Voß, S., Simple assembly line balancing—heuristic approaches. *Journal of Heuristics*, 2(3), 217–244. URL: <https://doi.org/10.1007/BF00127358> , 1997

[51] Yavuz, G., L-shade algoritmasının otomatik parametre yapılandırma yöntemi ile iyileştirilmesi. *Bilişim Teknolojileri Dergisi*, 15(2), 189-197. URL: <https://doi.org/10.17671/gazibtd.1034921>, 2022

[52] Marler, RT. and Arora, J.S., Function-transformation methods for multi-objective optimization. *Engineering Optimization*, 37(6), 551–570. URL: <https://doi.org/10.1080/03052150500114289>, 2005

[53] Ross, P. J., Taguchi techniques for quality engineering: loss function, orthogonal experiments, parameter and tolerance design (1rd ed.). USA: The McGraw-Hill Companies Inc, 1988

[54] Deliktaş, D. and Aydın, D., An artificial bee colony based-hyper heuristic algorithm with local search for the assembly line balancing problems. *Engineering Computations*, In Press, URL: <https://doi.org/10.1108/EC-02-2023-0075>, 2023



RESEARCH ARTICLE

Receive Date: 24.10.2023

Accepted Date: 10.01.2024

Determination of the species boundaries of genus *Dolerus* (Hymenoptera: Tenthredinidae) using the *COI* gene

Mehmet Gülmez^{a*}, Ertan Mahir Korkmaz^b, Mahir Budak^c

¹*Sivas Cumhuriyet University, Institute of Science, Department of Molecular Biology and Genetics, 58140, Sivas, Turkey, ORCID: 0000-0001-6547-7190

²Sivas Cumhuriyet University, Faculty of Science, Department of Molecular Biology and Genetics, 58140, Sivas, Turkey, ORCID: 0000-0003-0699-1354

³Sivas Cumhuriyet University, Faculty of Science, Department of Molecular Biology and Genetics, 58140, Sivas, Turkey, ORCID: 0000-0001-5610-486X

Abstract

New generation molecular approaches and methods are being developed to identify species and determine species boundaries. There are many different approaches of species delimitation used to assess the species richness of poorly studied and highly diverse invertebrate taxa. The basis of these approach is DNA barcoding studies. DNA barcoding has been used as a powerful tool for species identification and delimitation. Although DNA barcoding studies have been carried out on the family Tenthredinidae, there are no studies on species delimitation. Herein, we compare species delimitation analyzes belong to *Dolerus* genus based on *cytochrome c oxidase I (COI)* region. In this context, it was used five species delimitation approaches (ABGD, ASAP, DNA Taxon, PTP and GMYC). Thirty-six morphotypes were used in the study. These morphotypes separated into six species (*Dolerus triplicatus*, *Dolerus germanicus*, *Dolerus puncticollis*, *Dolerus nigratus*, *Dolerus* sp1 and *Dolerus* sp2) in ABGD, ASAP and DNA Taxon approaches. Two additional species were introduced because of the tree-based PTP and GMYC approaches. These species were named as *Dolerus* sp3 and *Dolerus* sp4 which were separated from *Dolerus puncticollis* clade and *Dolerus nigratus* clade, respectively. These analyzes were supported by the phylogenetic tree and CBC entities that constitute the ITS2 data.

© 2023 DPU All rights reserved.

Keywords: COI; *Dolerus*; Hymenoptera; Species Delimitation; Tenthredinidae.

1. Introduction

Hymenoptera, one of the ‘big four’ megadiverse insect orders, has more than 153,000 described and one million estimated species [1, 2, 3]. Along with species richness, the lifestyles of Hymenoptera are extremely diverse, ranging from feeding on or in plants to a wide variety of parasitic and predatory species [4, 5]. Symphyta (Gerstaecker, 1867), commonly known as sawflies [6], is a small suborder of Hymenoptera represented with 4,396 species. [7]. Tenthredinidae is the largest of the nine families of Symphyta suborder and includes 415 genera comprising 5721 species [7]. *Dolerus* (Panzer, 1801), is a genus belonging to Tenthredinidae, has 259 species distributed in the Palearctic and Nearctic regions [7, 8]. Adults and larvae of *Dolerus* are found in different habitats: open vernal pools like sedges (Cyperaceae), horsetails (Equisetaceae), open, wet, grass communities (Poaceae)

and rushes (Juncaceae) [9].

The morphological identification problems and inadequate taxonomic studies of sawflies lead to difficulties in identification of these taxa. Although, there are many studies on the order Hymenoptera involving both DNA barcoding and species delimitation approaches [10-17], the number of studies on phylogenetic relationships of Symphyta is still limited [1]. Both conventional taxonomy and molecular marker investigations have been conducted on the Tenthredinidae [18-25], however none of them have employed species delimitation techniques.

The finding of unique morphological differences in identification keys was the foundation of traditional taxonomy, which is still in widespread use today. However, modern approaches are being developed every day to identify species and determine species boundaries [26]. Integrative taxonomy, which includes DNA data and morphology-dependent analyses, is now utilized for efficient taxonomic identification [27, 28]. DNA barcoding [29, 30] refers to the utilization of the *cytochrome c oxidase I (COI)* region, located on mitochondrial DNA (mtDNA), to efficiently and precisely identify species of taxa that are challenging to discern based on their morphology. These studies mostly use mitochondrial gene (*COI*) or nuclear region (ITS2) which known as molecular markers [31]. For insects, an approximately 650 bp fragment of the *COI* is used as the standard “barcode region” [32, 33]. The relatively high mutation rate of mitochondrial genes compared to nuclear genes allows us to reveal phylogenetic relationships and incompatibilities such as geographic variation [34, 35]. The *COI* gene has an important role in revealing the taxonomy and evolutionary relationships in the DNA barcoding studies, due to its comprising both highly conserved and variable regions [36]. Because of all these advantages, the *COI* gene is preferred in barcoding studies by many researchers. The *COI* gene has been used for species delimitation approaches also in many Hymenoptera families, including diverse groups such as Braconidae [37, 38], Formicidae [39], Gasteruptionidae [40, 41], Eurytomidae [42], Vespidae [43], Ichneumonidae [44]. The barcoding and species delimitation studies can also show unsolved diversity [45], reveal lineages or point out new species [46].

Contemporary molecular-based species delimitation analyses consist of procedures for classifying individuals as either members of an existing species or as new species [47]. These analyzes are now widely used in a variety of taxa to support traditional taxonomy [48, 49]. A single locus is considered ideal in these analyzes, while multiple loci may sometimes be preferred. Single-locus species delimitation methods are still widely applied in both DNA barcoding and species delimitation studies involving organisms like bacteria, fungi, vertebrates, and invertebrates [50, 51]. Species delimitation approaches can also use processed data such as distance or phylogenetic trees. The aim of the using different data is to verify consistency of results [52-54].

It is important to use different genes or additional data such as morphology in integrative taxonomic analyzes to delimit the species more accurately [27, 28]. Over the last 20 years, ITS2 region together with *COI*, has been the most popular marker for phylogeny and species identification from different perspectives [55, 56]. However, high variation of ITS2 prevent its safe use in species delimitation and it has been understood that the sequences of ITS2, are not informative enough for species-level comparisons for some insect genera [57]. Therefore, revealing the species diversity needs further use of DNA barcoding and species delimitation approaches with different gene or regions [6].

Many of the species groups in the genus of *Dolerus* have not been yet resolved taxonomically [58]. Therefore, we preferred species delimitation approaches used in taxonomic and molecular studies here. The aims of the present study: a) to compare distance- and tree-based species delimitation approaches on *Dolerus* genus, b) to compare the results of the species delimitation analyzes with those of our previous ITS2-based study [25]. For this purpose, we utilized the *COI* phylogenetic tree, genetic distances, and comparison of various species delimitation approaches. At the same time, our study represents the first evaluation of comparing species delimitation approaches based on partial *COI* data of the genus *Dolerus* using molecular data.

2. Materials and Methods

2.1. Molecular analysis

DNA extracts of 36 morphospecies from the genus *Dolerus* identified in our previous study, obtained by using

Salting out protocol [59], were preserved at -20°C in Entomological Collection of Cumhuriyet University, Sivas. These samples were used for the amplification of the *COI* region by using primer pairs s1859 (5'-GGA ACI GGA TGA ACW GTT TAY CCI CC -3') and a2590 (5'-GCT CCT ATT GAT ARW ACA TAR TGR AAA TG-3') [60]. PCR reactions and cycling conditions were taken from Gülmez et. al. 2022 [25] except for the annealing stage which is conducted at 46°C for 30 s. The obtained PCR products were visualized by electrophoresing on the 1% agarose gel. PCR products were then sequenced using Sanger technology (Macrogen Ltd., Seoul, Korea) in both directions.

2.2. Phylogenetics analysis

The raw sequences of 36 samples from the genus *Dolerus* were generated for this study and each sequence with the forward and reverse direction were assembled, edited, and manually checked by eye using Geneious R9 [61]. Each partial *COI* sequence was checked whether belonging to the genus *Dolerus* using “blastn” algorithm [62]. The sequences were deposited to GenBank under the accession numbers OR721886- OR721921. Alignments of partial *COI* sequences of the 36 samples of *Dolerus* were performed using the MAFFT algorithm [63]. Pairwise genetic distances of the partial *COI* dataset were determined using Kimura-2 (K2P) [64] and uncorrected distance (p-distance) parameters in MEGA11 [65]. These distance data were exported as a MEGA file to be used in Automatic Barcode Gap Discovery (ABGD) analysis, one of the species delimitation tests [36]. The best-fit model of nucleotide substitution was determined using jModelTest 2.1.7 [66] and fasta file were created using only 1st and 2nd codon positions by MEGA11 due to the substitution saturation in 3rd codon positions [65]. The dataset was used both for the construction of Maximum Likelihood (ML) tree using Randomized Axelerated Maximum Likelihood-High Performance Computing (RAxML-HPC) v.8 with 1000 bootstrap replications in CIPRES portal [67] and construction of Neighbor-Joining (NJ) tree with 1000 bootstrap replications in MEGA11. ML and NJ tree files in newick format were visualized using FigureTree (v 1.4.4) [68].

2.3. Species delimitation analyzes

Five different approaches were preferred for species delimitation analyzes: The General Mixed Yule Coalescent (GMYC) model [69] with a single threshold, (ABGD) [36], the Poisson Tree Processes (PTP) (<https://species.h-its.org/>) [70], Assemble Species by Automatic Partitioning (ASAP) (<https://bioinfo.mnhn.fr/abi/public/asap/>) [71] and TaxonDNA [72]. However, it was performed two different analyzes using p-distance and K2P distance parameters in ABGD approach. So, this study was planned a total of six analyzes based on five different approaches. For ABGD analysis, which is a distance-based method, the model setting was set as follows: TS/TV (ratio of transition to transversion) is 0.967, variability (P) is between 0.001 (P-min) and 0.132 (P-max), K2P and P distance, minimum gap width (×) of 0.1-1.5. To apply the GMYC delimitation method, an ultrametric tree was constructed with “force.ultrametric” command and was checked using “is.ultrametric” command in R [73]. The obtained ultrametric tree for GMYC was used with single threshold method using the “gmyc” function under the “SPLITS” package (R Development Core Team, www.R-project.org). For PTP, the RAxML tree was employed as input file and analyzed via the PTP web server (<https://species.h-its.org/>) with all parameters given by default, except for the number of generations, which was set to 100,000 generations. The most proper group was found by objective clustering with p-distance thresholds at 1–6% using TaxonDNA 1.8. Best Close Match (BCM) test in TaxonDNA/Species Identifier 1.8 was used to select the best threshold value and to evaluate the potential of the *COI* dataset for species identification. ASAP approach [71] is distance-based method like ABGD, and this analysis has performed in web interface. In this method, p distance parameter was preferred simple distance (p- distances).

3. Results and Discussion

Six analyses with five different methodologies (tree-based and distance- based) were conducted in this study. The compared methods used in this study all rely on a single locus for identifying species boundaries. Information

of thirty-six *Dolerus* samples identified according to these analyzes is given in Table 1. These species are *Dolerus triplicatus* (Klug, 1818), *Dolerus germanicus* (Fabricius, 1775), *Dolerus puncticollis* Thomson, 1871, *Dolerus nigratus* (Müller, 1776), *Dolerus* sp1, *Dolerus* sp2, *Dolerus* sp3 and *Dolerus* sp4. They were determined that the putative *Dolerus* sp3 and *Dolerus* sp4 species were separated from *D. puncticollis* and *D. nigratus* species, respectively.

Table 1. Information of *Dolerus* samples.

Specimens	Localities of specimens	Identification according to ITS2 (Gülmez et al, 2022)	ABGD-p, ABGD-K2P, ASAP, DNA Taxon	PTP, GMYC
spcmn1	Erzurum-Tortum	<i>D. triplicatus</i>	<i>D. triplicatus</i>	<i>D. triplicatus</i>
spcmn2	Erzurum-Tortum	<i>D. triplicatus</i>	<i>D. triplicatus</i>	<i>D. triplicatus</i>
spcmn3	Erzurum-Tortum	<i>D. triplicatus</i>	<i>D. triplicatus</i>	<i>D. triplicatus</i>
spcmn4	Erzincan-Refahiye	<i>D. triplicatus</i>	<i>D. triplicatus</i>	<i>D. triplicatus</i>
spcmn5	Erzincan-Refahiye	<i>D. triplicatus</i>	<i>D. triplicatus</i>	<i>D. triplicatus</i>
spcmn6	Erzincan-Refahiye	<i>D. triplicatus</i>	<i>D. triplicatus</i>	<i>D. triplicatus</i>
spcmn7	Kütahya-Altıntaş	<i>D. germanicus</i>	<i>D. germanicus</i>	<i>D. germanicus</i>
spcmn8	Kütahya-Altıntaş	<i>D. germanicus</i>	<i>D. germanicus</i>	<i>D. germanicus</i>
spcmn9	Uşak-Banaz	<i>D. germanicus</i>	<i>D. germanicus</i>	<i>D. germanicus</i>
spcmn10	Ankara-Bala	<i>D. germanicus</i>	<i>D. germanicus</i>	<i>D. germanicus</i>
spcmn11	Erzincan-Refahiye	<i>D. germanicus</i>	<i>D. germanicus</i>	<i>D. germanicus</i>
spcmn12	Erzincan-Refahiye	<i>D. germanicus</i>	<i>D. germanicus</i>	<i>D. germanicus</i>
spcmn13	Erzincan-Refahiye	<i>D. germanicus</i>	<i>D. germanicus</i>	<i>D. germanicus</i>
spcmn14	Erzincan-Refahiye	<i>D. germanicus</i>	<i>D. germanicus</i>	<i>D. germanicus</i>
spcmn15	Erzincan-Refahiye	<i>D. germanicus</i>	<i>D. germanicus</i>	<i>D. germanicus</i>
spcmn16	Erzincan-Refahiye	<i>D. germanicus</i>	<i>D. germanicus</i>	<i>D. germanicus</i>
spcmn17	Erzurum-Tortum	<i>D. puncticollis</i>	<i>D. puncticollis</i>	<i>D. puncticollis</i>
spcmn18	Erzurum-Tortum	<i>D. puncticollis</i>	<i>D. puncticollis</i>	<i>D. puncticollis</i>
spcmn19	Nevşehir-Ürgüp	<i>D. puncticollis</i>	<i>D. puncticollis</i>	<i>D. puncticollis</i>
spcmn20	Nevşehir-Ürgüp	<i>D. puncticollis</i>	<i>D. puncticollis</i>	<i>Dolerus</i> sp3*
spcmn21	Nevşehir-Ürgüp	<i>D. puncticollis</i>	<i>D. puncticollis</i>	<i>Dolerus</i> sp3*
spcmn22	Ankara-Beyşehir	<i>D. puncticollis</i>	<i>D. puncticollis</i>	<i>D. puncticollis</i>
spcmn23	Sivas-Gürün	<i>Dolerus</i> sp1	<i>Dolerus</i> sp1	<i>Dolerus</i> sp1
spcmn24	Ankara-Beyşehir	<i>D. puncticollis</i>	<i>D. puncticollis</i>	<i>D. puncticollis</i>
spcmn25	Niğde-Çamardı	<i>D. puncticollis</i>	<i>D. puncticollis</i>	<i>D. puncticollis</i>
spcmn26	Niğde-Çamardı	<i>D. puncticollis</i>	<i>D. puncticollis</i>	<i>D. puncticollis</i>
spcmn27	Kastamonu-Tosya	<i>D. nigratus</i>	<i>D. nigratus</i>	<i>D. nigratus</i>
spcmn28	Kastamonu-Tosya	<i>D. nigratus</i>	<i>D. nigratus</i>	<i>Dolerus</i> sp4*
spcmn29	Kastamonu-Tosya	<i>D. nigratus</i>	<i>D. nigratus</i>	<i>D. nigratus</i>
spcmn30	Erzincan-Refahiye	<i>D. nigratus</i>	<i>D. nigratus</i>	<i>D. nigratus</i>
spcmn31	Erzurum-Oltu	<i>D. nigratus</i>	<i>D. nigratus</i>	<i>D. nigratus</i>
spcmn32	Erzincan-Refahiye	<i>D. nigratus</i>	<i>D. nigratus</i>	<i>D. nigratus</i>
spcmn33	Erzurum-Oltu	<i>D. nigratus</i>	<i>D. nigratus</i>	<i>D. nigratus</i>
spcmn34	Kütahya-Altıntaş	<i>Dolerus</i> sp2	<i>Dolerus</i> sp2	<i>Dolerus</i> sp2
spcmn35	Kütahya-Altıntaş	<i>Dolerus</i> sp2	<i>Dolerus</i> sp2	<i>Dolerus</i> sp2

spcmn36	Kütahya-Altıntaş	<i>Dolerus</i> sp2	<i>Dolerus</i> sp2	<i>Dolerus</i> sp2
---------	------------------	--------------------	--------------------	--------------------

*: As a result of the species delimitation analysis, it was determined that the putative species.

The percentage of the average nucleotide composition of *COI* sequences of each species is given in Table 2. Ratio of nucleotide compositions of the *COI* sequences of each species are variable. AT contents of the examined sequences ranged between 71.80% (*Dolerus* sp1) and 73.10% (*D. nigratus*) (Table 2). The average AT content of the *COI* region mentioned in the study of Hebert (2003) [74], which is considered as the DNA barcode region and used in the analyzes, showed an AT content, like other Hymenoptera members that have been reported [75-80]. Moreover, additional proof that the sequences are *COI* comes from the fact that the “Blastn algorithm” [62] produced Per-Identities scores for the genus *Dolerus* ranging from 93 to 98%.

Table 2. Average nucleotide content of *COI* gene belongs to each species.

Specimens	Species Name	T%	C%	A%	G%	AT%
spcmn1-6	<i>D. triplicatus</i>	39.15	14.02	33.23	13.58	72.38
spcmn7-16	<i>D. germanicus</i>	38.85	14.55	32.96	13.65	71.81
spcmn17,18,19,22,24,25,26	<i>D. puncticollis</i>	38.64	13.93	33.64	13.79	72.29
spcmn20,21	<i>Dolerus</i> sp3	38.2	13.95	33.9	14	72.10
spcmn23	<i>Dolerus</i> sp1	38.6	14.9	33.2	13.3	71.80
spcmn27,29,30,31,32,33	<i>D. nigratus</i>	39.10	13.67	34.00	13.25	73.10
spcmn28	<i>Dolerus</i> sp4	38.9	13.7	33.9	13.4	72.80
spcmn34-36	<i>Dolerus</i> sp2	39.73	13.40	32.97	13.87	72.70

A=Adenine T=Thymine, C=Cytosine, G=Guanine, AT= Adenine – Thymine content

As a result of genetic distance, the interspecies distance in eight species was designated as a maximum of 9.7% (*D. nigratus-Dolerus* sp1 vs *D. germanicus*) and a minimum of 1.6% (*D. puncticollis* vs *Dolerus* sp1) (Table 3). In the intra-species genetic distance results, *D. puncticollis* samples have the maximum distance (0.872%) (Table 4). Since *Dolerus* sp1 and *Dolerus* sp4 are represented by one sample each, their interspecies genetic distances could not be calculated. According to Hebert et al (2004) [81], a 10-fold difference between mean intraspecific and interspecific differences is specified as the standard *COI* threshold for identifying animal species. This Figure. is over the designated threshold value, as evidenced by the fact that it was 13 times in the study (the difference between the average interspecies (4%) and intraspecific divergence (0.30%)). Comparison of average intraspecific and interspecific genetic distances is widely used in species delimitation, as well as in barcoding studies. Maximum distances between *Dolerus* species reflect the pattern seen in species delimitation analyses, where well-supported clusters (clades) consist of more than one species.

Table 3. Interspecific genetic distance.

No	Species	Genetic Distance							
		1	2	3	4	5	6	7	8
1	<i>D. triplicatus</i>								
2	<i>D. germanicus</i>	7.5%							
3	<i>D. puncticollis</i>	6.8%	7.8%						
4	<i>Dolerus</i> sp1	6.9%	7.6%	1.6%					
5	<i>Dolerus</i> sp3	6.4%	8.9%	3.5%	4%				
6	<i>Dolerus</i> sp4	6.6%	8%	4.5%	4.4%	4.9%			
7	<i>D. nigratus</i>	7.7%	9.7%	7.1%	6.9%	6.5%	6.6%		

8 *Dolerus* sp2 8.3% 9.7% 6.4% 6.6% 6% 4.3% 2.9%

Table 4. Intraspecific genetic distance of *Dolerus* species.

Species	d	SE
<i>D. triplicatus</i>	0	0
<i>D. germanicus</i>	0.2%	0.122%
<i>D. puncticollis</i>	0.9%	0.251%
<i>Dolerus</i> sp3	0.3%	0.232%
<i>Dolerus</i> sp1*	n/c	n/c
<i>Dolerus</i> sp4*	n/c	n/c
<i>D. nigratus</i>	0.06%	0.055%
<i>Dolerus</i> sp2	0.04%	0.228%

* *D. sp1* and *D. sp4* are represented by one sample each.

To compare the species delimitation analyses of the *Dolerus* genus, a total of six analyses based on five different approaches were conducted. In addition, we employed comparison analyses to utilize the ITS2 results (phylogenetic tree and CBCs) from our previous research [25]. Comparison analyses summarizing the results of the six different species delimitation analyses and the results of Gülmez et al. (2022) [25] (ITS2) are shown on a RaxML tree (Fig. 2). These analyses led to the identification of eight groups from tree-based analyses (PTP and GMYC) and six groups from distance-based ones (ABGD-p, ABGD-K2P, ASAP, and DNA Taxon) (Figs 2). The reason for the variability in the number of species is the use of approaches with different algorithms. The recursive partitioning of data using ABGD and ASAP techniques, which are computationally and time-efficient, involves comparing sequence differences to identify a "barcode gap" that may indicate the boundaries of different species [40]. Tree-based methods identify species boundaries by calculating branch variation using a phylogenetic tree.

Two different inputs, P-dist and K2P distance, were used in four distance-based analyzes. In the consequence of ABGD-P-dist analysis and ASAP analysis, it was observed that there were respectively 0.036% and 0.045% barcode gaps between the maximum intraspecific distance and minimum interspecific distance values in the COI data set of *Dolerus* species (Fig. 1). Despite using the same distance data, the barcode gaps were different. However, both analyses grouped the same number of species. Similarly, DNA-Taxon analysis which a species delimitation tool that clusters using intraspecific genetic distances [72], also found that same number groups as other distance-based analyses. The six groups identified by ASAP, ABGD-p, ABGD-K2P, and Taxon DNA analyses yielded identical species groups to those reported in our earlier study [25]. Moreover, for detailed comparison of intraspecific relationships, a distance-based NJ tree was also examined. The NJ tree exhibited the same topology with RAxML.

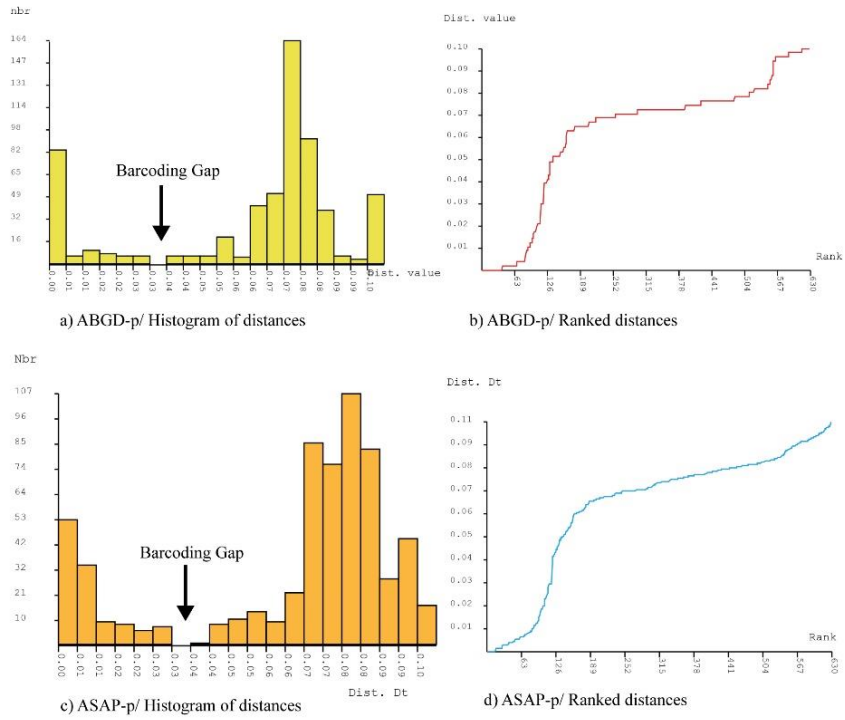


Fig. 1. a) ABGD-p/ Histogram of distances. b) ABGD-p/ Ranked distances. c) ASAP/ Histogram of distances. d) ASAP/ Ranked distances

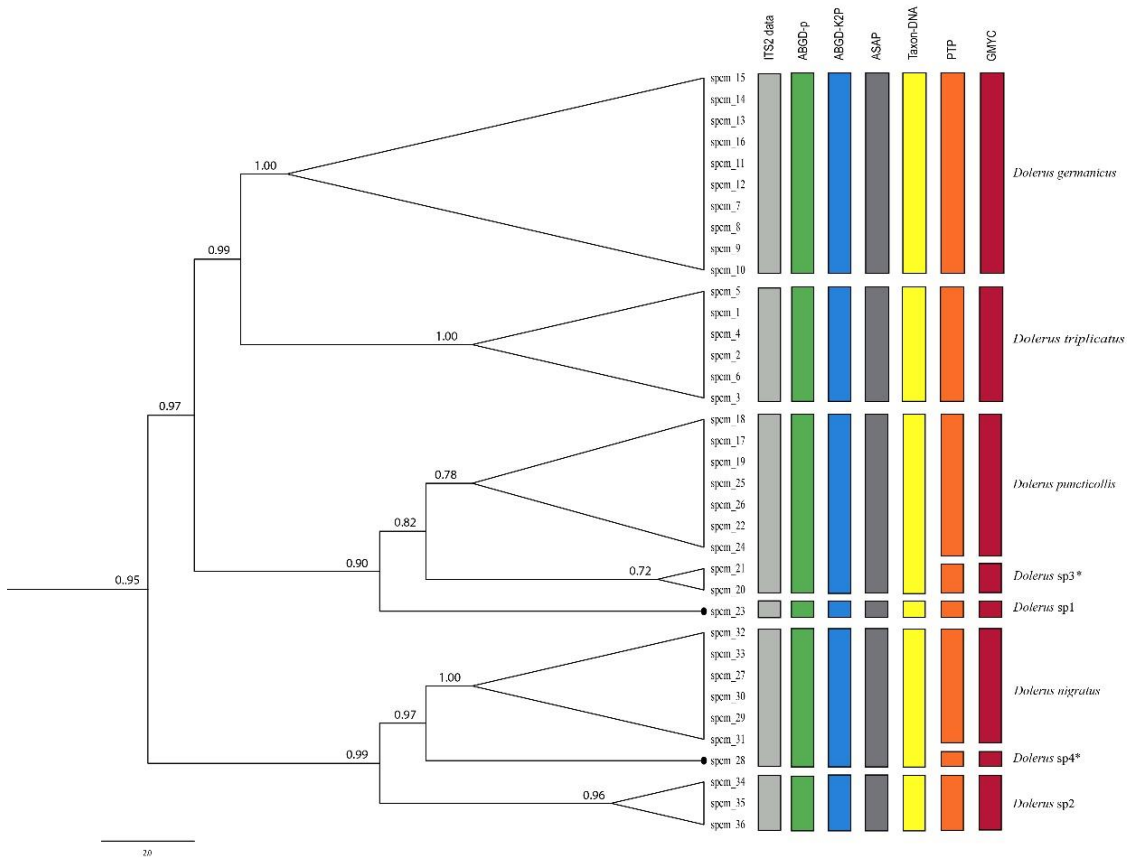


Fig. 2. Comparative all species delimitation analyzes on consensus tree of *COI* region conducted by RaxML.

Samples placed between spcm1 and spcm6 are grouped together in both distance and tree-based analyses. Since these samples represent the group defined as *D. triplicatus* according to ITS2 results (Grey color in Fig. 2), they gave similar results in both studies. Similarly, in all analyses, samples between spcm7 and spcm16 were assigned to a single group and were compatible with the *D. germanicus* species represented in ITS2 results (Grey color in Fig. 2). The spcm23 sample, which was named *Dolerus* sp1 in the previous study (Fig. 2), was in a different group in all species delimitation analyses. Its appearance in the different group supported the previous study. Consistent with the prior study's designation of these specimens as *Dolerus* sp2, all species delimitation analyses included spcm34, spcm35 and spcm36 in same group (Fig. 2). These results support comparison of all species delimitation analyzes and the ITS2 results (Grey color in Fig. 2). Distance-based analyzes (ABGD, ASAP and Taxon DNA) have given spcm20 and spcm21 with *D. puncticollis* in the same group. The distance-based analyzes of COI data and the tree topology of ITS2 results support each other. However, GMYC and PTP analyzes grouped these two samples separately from the *D. puncticollis* group. As seen in Fig. 2, spcm20 and spcm21 samples were separated from the *D. puncticollis* species group and formed a different clad (*Dolerus* sp3). When the results of the previous study are examined, it is seen that spcm20 and spcm21 samples are separated from other samples by two CBCs. The existence of these CBCs are supported by this study [25]. For this study, it is thought that the ITS2 phylogeny and CBC presence together with species delimitation analyzes will provide more informative species-level identifications. However, since both ITS2 results and the groups given by tree-based approaches do not support

each other, it was named as the putative *Dolerus* sp3.

The spcm28 sample was found in the same group with *D. nigratus* in distance-based approaches, which shows its similarity with the previous study. Tree-based analyses, however, revealed that this species belonged to a distinct single group. Although, there is no CBC presence between them, the spcm28 sample showed separate branching from the *D. nigratus* clade in the ML tree. The ML tree and species delimitation analyses supported each other, and therefore it was named as the putative species *Dolerus* sp4. GMYC approach is a coalescent-based phylogenetic method that sets thresholds between coalescent and species-level processes to species boundaries. PTP approach models speciation events by the number of substitutions in each branch, which equates to a higher expected number of substitutions between species than within species. In this context, the tree based GMYC model is an analytical approach that an ultrametric phylogenetic tree as the most likely point of transition from merging to speciation branching models [69, 80]. These models continue to be used successfully in recent times to delimit species in a wide variety of little-known insect taxa. [82-85].

GMYC and PTP analyzes generally produce similar estimates of species boundaries [86-89]. Same species groups were determined from PTP and GMYC analyzes, using the RAxML and ultrametric tree as input. Branching points or nodes in a tree are considered to indicate speciation. In monophyletic trees, each node represents the last common ancestor of two lineages that diverged from that node [85]. Therefore, the fact that each of the two main clades containing *D. puncticollis* and *D. nigratus* species have three nodes, as well as the presences of CBCs shown on the tree topology in the previous study, supported the existence of putative species groups emerged in these analyses (Fig. 2). GMYC and PTP also offer some distinct advantages over the other four types of delimitation analysis. The main benefit of these approaches is that they are far less reliant on the threshold value and integrate evolutionary theory [28]. The CBCs identified in the previous investigation support the suggested species boundaries in this analysis. Although analysis of single-locus mtDNA data and decisions based on small sample sizes pose interpretation risks, processing the data with species delimitation analyzes can provide accurate estimates of the number of species [90].

4. Conclusion

The species delimitation methods correctly group known species into clusters in most cases. The grouping of the ASAP analysis [71], which is based on the best scoring algorithm, is supported by other distance-based analyzes. In addition, PTP and GMYC analyzes are internally consistent. The main reason for this difference is the use of the ultrametric tree in the PTP and GMYC analyzes. On this tree, rates of branching events are estimated to reveal patterns of speciation (interspecific relation) and coalescence (intraspecific relation) [69, 71]. Therefore, tree-based analyzes take longer to complete than distance-based analyzes in terms of time. Although this is stated to be a disadvantage by some researchers, these analyzes among the most popular approaches to provide reliable results.

There was no consensus on the number of common species in both distance and tree-based analyzes. However, the reliability of tree-based analyzes interpreted using additional data such as ITS2 and CBC, is one step forward. Puillandre et al., (2020) [71], reported that the performance of ABGD was similar to that of ASAP, and although PTP did not perform very well, GMYC performed very well as long as the number of species was not too high. Since GMYC and PTP analyzes are based on evolutionary relationships, we named the groups separated from *D. puncticollis* and *D. nigratus* as *Dolerus* sp3 and *Dolerus* sp4, respectively. As this is the first study with this taxon group, testing species delimitation analyzes will serve as a resource for future studies for this important family.

Acknowledgments

This study was financially supported by the TÜBİTAK (The Scientific and Technological Research Council of Turkey) via a research project with grant number 113Z753. The members of Cumhuriyet University Evolutionary Bioinformatics Research Group (EBRG) are thanked for their contributions in bioinformatic work. Mehmet GÜLMEZ who first author is supported by the Council of Higher Education (CoHE) of Turkey with 100/2000 PhD Scholarship

References

- [1] E. M. Viitasari, "Sawflies I. A review of the suborder, the Western Palaearctic taxa of Xyeloidea and Pamphilioidea", Tremex Press, Helsinki, pp. 516, 2002.
- [2] A. P. Aguiar et al., "Order Hymenoptera. In: Animal Biodiversity: An Outline of Higher-level Classification and Survey of Taxonomic Richness", *Zootaxa*, vol. 3703, no. 1, pp. 51-65, 2013, doi: <https://doi.org/10.11646/zootaxa.3703.1.12>.
- [3] G. Niu et al., "Mitochondrial Phylogenomics of Tenthredinidae (Hymenoptera: Tenthredinoidea) Supports the Monophyly of Megabelesinae as a Subfamily", *Insects*, vol. 12, pp. 495, 2021, doi: <https://doi.org/10.3390/insects12060495>.
- [4] I. D. Gauld and B. Bolton, "The Hymenoptera", Oxford University Press, Oxford, 1988.
- [5] D. A. Grimaldi and M. S. Engel, "Evolution of the Insects", Cambridge University Press, New York, 2005.
- [6] S. Schmidt et al., "Identification of sawflies and horntails (Hymenoptera, Symphyta) through DNA barcodes: Successes and Caveats", *Molecular Ecology Resources*, vol. 17, no. 4, pp. 670-685, 2017, doi: <https://doi.org/10.1111/1755-0998.12614>.
- [7] A. Taeger et al., "ECatSym. Electronic World Catalog of Symphyta (Insecta, Hymenoptera)", Program version 5.0 (19 Dec 2018), data version 40 (23 Sep 2018). Senckenberg Deutsches Entomologisches Institut (SDEI), Müncheberg. (Web page: <https://sdei.de/ecatsym/>) (Date accessed: 21 May 2019)
- [8] A. M. Barker, "The identification of larvae of eight graminivorous species of the sawfly genus *Dolerus* Panzer 1801 (Hymenoptera: Tenthredinidae) regularly found in grass and cereal fields in southern England", *Journal of Natural History*, vol. 32, no. 8, pp. 1181-1215, 1998.
- [9] J. Borowski, "Materials to the knowledge of Polish sawflies. The genus *Dolerus* Panzer, 1801 (Hymenoptera, Symphyta, Tenthredinidae, Selandriinae). Part XVIII– *Dolerus* (*Achaetoprion*) *pachycerus* Hartig, 1837 with observations on its biology and a key for identification of larvae of subgenus *Achaetoprion* Goulet, 1986", *Polish Journal of Entomology*, vol. 92, pp. 1-6, 2023, doi: 10.5604/01.3001.0053.3993.
- [10] M. D. Schwarzfeld and F. A. H. Sperling, "Species delimitation using morphology, morphometrics, and molecules: definition of the *Ophion scutellaris* Thomson species group, with descriptions of six new species (Hymenoptera, Ichneumonidae)", *ZooKeys*, vol. 462, pp. 59–114, 2014, doi: 10.3897/zookeys.462.8229.
- [11] J. T. Longino, and M. G. Branstetter, "Phylogenomic species delimitation, taxonomy, and 'bird guide' identification for the Neotropical ant genus *Rasopone* (Hymenoptera: Formicidae)", *Insect Systematics and Diversity*, vol. 4, no. 2, pp. 1, 2020, doi: <https://doi.org/10.1093/isd/ixaa004>.
- [12] C. Waichert, J. S. Wilson, J. P. Pitts and C. D. V. Dohlen, "Phylogenetic species delimitation for the widespread spider wasp *Ageniella accepta* (Hymenoptera: Pompilidae), with new synonyms", *Insect Systematics and Evolution*, vol. 51, no. 3, pp. 532-549, 2020, doi: <https://doi.org/10.1163/1876312X-00002207>.
- [13] Z. Liu et al., "Tackling the Taxonomic Challenges in the Family Scoliidae (Insecta, Hymenoptera) Using an Integrative Approach: A Case Study from Southern China", *Insects*, vol. 12, no. 10, pp. 892, 2021, doi: <https://doi.org/10.3390/insects12100892>.
- [14] M. M. Prebus, "Phylogenomic species delimitation in the ants of the *Temnothorax salvini* group (Hymenoptera: Formicidae): an integrative approach", *Systematic Entomology*, vol. 46, no. 2, pp. 307-326, 2021, doi: <https://doi.org/10.1111/syen.12463>.
- [15] P. C. S. Barroso, R. S. T. Menezes, M. L. de Oliveira, and A. Somavilla, "A systematic review of the Neotropical social wasp genus *Angiopolybia* Araujo, 1946 (Hymenoptera: Vespidae): species delimitation, morphological diagnosis, and geographical distribution", *Arthropod Systematics and Phylogeny*, vol. 80, pp. 75-97, 2022, doi: 10.3897/asp.80.e71492.
- [16] A. Somavilla, M. L. D. Oliveira, R. S. Menezes and P. C. S. Barroso, "A systematic review of the Neotropical social wasp genus *Angiopolybia* Araujo, 1946 (Hymenoptera: Vespidae): species delimitation, morphological diagnosis, and geographical distribution", *Arthropod Systematics and Phylogeny*, vol. 80, pp. 75-97, 2022, doi: 10.3897/asp.80.e71492.
- [17] S. Shimizu and K. Maeto, "A New Distinctive Darwin Wasp Represents the First Record of the *Ophion minutus* Species-group (Hymenoptera: Ichneumonidae: Ophioninae) from Japan and the Far East, with an Analysis of DNA Barcode-based Species Delimitation in *Ophion*". *Zoological Studies*, vol. 62, 2023, doi: 10.6620/ZS.2023.62-27.
- [18] S. Schulmeister, "Simultaneous analysis of basal Hymenoptera (Insecta): introducing robust-choice sensitivity analysis", *Biological Journal of the Linnean Society*, vol. 79, no. 2, pp. 245-275, 2023, doi: <https://doi.org/10.1046/j.1095-8312.2003.00233.x>.
- [19] M. Prous, M. Heidema, S. Akihiko and V. Soon, "Review of the sawfly genus *Empria* (Hymenoptera, Tenthredinidae) in Japan", *ZooKeys*, vol. 150, pp. 347-380, 2011, doi: 10.3897/zookeys.150.1968.
- [20] S. A. Leppänen, E. Altenhofer, A. D. Liston and T. Nyman, "Phylogenetics and evolution of host-plant use in leaf mining sawflies (Hymenoptera: Tenthredinidae: Heterarthrinae)", *Molecular Phylogenetics and Evolution*, vol. 64, no. 2, pp. 331-341, 2012, doi: <https://doi.org/10.1016/j.ympev.2012.04.005>.
- [21] Y. Isaka and T. Sato, "Molecular phylogenetic and divergence time estimation analyzes of the sawfly subfamily Selandriinae (Hymenoptera: Tenthredinidae)", *Entomological Science*, vol. 17, no. 4, pp. 435-439, 2014, doi: <https://doi.org/10.1111/ens.12080>.
- [22] T. Malm and T. Nyman, "Phylogeny of the Symphytan grade of Hymenoptera: new pieces into the old jigsaw (fly) puzzle", *Cladistics*, vol. 31, no. 1, pp. 1-17, 2015, doi: <https://doi.org/10.1111/cla.12069>.
- [23] L. Vilhelmsen, "Morphological phylogenetics of the Tenthredinidae (Insecta: Hymenoptera)", *Invertebrate Systematics*, vol. 29, no. 2, pp. 164-190, 2015, doi: <http://dx.doi.org/10.1071/IS14056>.
- [24] M. Budak, M. Güler, E. M. Korkmaz, S. H. Örgen and H. H. Başibüyük, "The characterization and taxonomic utility of ITS2 in *Tenthredopsis* Costa, 1859 (Tenthredinidae: Hymenoptera) with some new records from Turkey", *Biochemical Systematics and Ecology*, vol. 66, pp. 76-85, 2016, doi: <https://doi.org/10.1016/j.bse.2016.03.008>.
- [25] M. Gülmez, M. Budak, E. M. Korkmaz, S. H. Örgen and H. H. Başibüyük, "Characterization and taxonomic utility of ITS2 in *Dolerus* Panzer, 1801 (Hymenoptera: Tenthredinidae)", *Turkish Journal of Entomology*, vol. 46, no. 1, pp. 13-23, 2022, doi: <http://dx.doi.org/10.16970/entoted.1018061>.
- [26] J. W. Sites Jr and J. C. Marshall, "Delimiting species: a Renaissance issue in systematic biology", *Trends in Ecology and Evolution*, vol. 18, no. 9, pp. 462-470, 2003, doi: [https://doi.org/10.1016/S0169-5347\(03\)00184-8](https://doi.org/10.1016/S0169-5347(03)00184-8).

- [27] A.D. Roe and F. A. Sperling, “Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding”, *Molecular Phylogenetics and Evolution*, vol. 44, no. 1, pp. 325-345, 2007, doi: <https://doi.org/10.1016/j.ympev.2006.12.005>.
- [28] M. D. Schwarzfeld and F. A. Sperling, “Comparison of five methods for delimitating species in *Ophion* Fabricius, a diverse genus of parasitoid wasps (Hymenoptera, Ichneumonidae)”, *Molecular Phylogenetics and Evolution*, vol. 93, pp. 234-248, 2005, doi: <https://doi.org/10.1016/j.ympev.2015.08.003>.
- [29] P. N. D. Hebert and T. R. Gregory, “The promise of DNA barcoding for taxonomy”, *Systematic Biology*, vol. 54, pp. 852–859, 2005, doi: <https://doi.org/10.1080/10635150500354886>.
- [30] A. Valentini, F. Pompanon and P. P. Taberlet, “DNA barcoding for ecologists”, *Trends in Ecology and Evolution*, vol. 24, pp. 110–117, 2009, doi: <https://doi.org/10.1016/j.tree.2008.09.011>.
- [31] P. Z. Goldstein and R. DeSalle, “Integrating DNA barcode data and taxonomic practice: determination, discovery, and description”, *Bioessays*, vol. 33, pp.135–147, 2011, doi: <https://doi.org/10.1002/bies.201000036>.
- [32] M. S. Caterino, S. Cho and F. A. H. Sperling, “E current state of insect molecular systematics: a thriving tower of babel”, *Annual Review of Entomology*, vol. 45, no. 1, pp. 1–54, 2000, doi: <https://doi.org/10.1146/annurev.ento.45.1.1>.
- [33] P. D. N. Hebert, A. Cywinska, S. L. Ball and J. R. deWaard. “Biological identifications through DNA barcodes”, *Proceedings of the Royal Society B. Biological Sciences*, vol. 270, pp. 313–321, 2003a, doi: <https://doi.org/10.1098/rspb.2002.2218>.
- [34] C. R. Bonvicino, B. Lemos and H. N. Seuánez, “Molecular phylogenetics of howler monkeys (*Alouatta*, Platyrrhini) A comparison with karyotypic data”, *Chromosoma*, vol. 110, pp. 241–246, 2001, doi: <https://doi.org/10.1007/s004120000128>.
- [35] F. F. Nascimento, C. R. Bonvicino, and H. N. Seuánez, “Population genetic studies of *Alouatta caraya* (Alouattinae, Primates): inferences on geographic distribution and ecology”, *Am J Primatol*, vol. 69, pp. 1093–1102, 2007, doi: <https://doi.org/10.1002/ajp.20423>.
- [36] N. Puillandre, A. Lambert, S. Brouillet and G. Achaz, “ABGD, Automatic Barcode Gap Discovery for primary species delimitation”, *Molecular Ecology*, vol. 21, pp. 1864-1877, 2012, doi: <https://doi.org/10.1111/j.1365-294X.2011.05239.x>.
- [37] A. Zaldívar-Riverón et al., “DNA barcoding a highly diverse group of parasitoid wasps (Braconidae: Doryctinae) from a Mexican nature reserve”, *Mitochondrial DNA*, vol. 21, no. sup1, pp. 18–23, 2007, doi: <https://doi.org/10.3109/19401736.2010.523701>.
- [38] E. P. Fagan- Jeffries, S. J. Cooper, T. Bertozzi, T. M. Bradford and A. D. Austin, “DNA barcoding of microgastrine parasitoid wasps (Hymenoptera: Braconidae) using high- throughput methods more than doubles the number of species known for Australia”, *Molecular Ecology Resources*, vol. 18, no. 5, pp. 1132-1143, 2018, doi: <https://doi.org/10.1111/1755-0998.12904>.
- [39] S. K. Oberprieler, A. N. Andersen and C. C. Moritz, “Ants in Australia’s monsoonal tropics: CO1 barcoding reveals extensive unrecognised diversity”, *Diversity*, vol. 10, no. 2, pp. 36, 2018, doi: <https://doi.org/10.3390/d10020036>.
- [40] B. A. Parslow, M. P. Schwarz and M. I. Stevens, “Review of the biology and host associations of the wasp genus *Gasteruption* (Evaniodea: Gasteruptionidae)”, *Zoological Journal of the Linnean Society*, vol. 189, no. 4, pp. 1105-1122, 2020, doi: <https://doi.org/10.1093/zoolinnean/zlaa005>.
- [41] B. A. Parslow, M. P. Schwarz, and M. I. Stevens, “Molecular diversity and species delimitation in the family Gasteruptionidae (Hymenoptera: Evaniodea)”, *Genome*, vol. 64, no. 3, pp. 253-264, 2021, doi: <https://doi.org/10.1139/gen-2019-0186>.
- [42] Y. M. Zhang et al., “Delimiting the cryptic diversity and host preferences of *Sycophila* parasitoid wasps associated with oak galls using phylogenomic data”, *Molecular Ecology*, vol. 31, no. 16, pp. 4417-4433, 2022, doi: <https://doi.org/10.1111/mec.16582>.
- [43] P. C. S. Barroso, R. S. T. Menezes, M. L. de Oliveira, and A. Somavilla, “A systematic review of the Neotropical social wasp genus *Angiopolybia* Araujo, 1946 (Hymenoptera: Vespidae): species delimitation, morphological diagnosis, and geographical distribution”, *Arthropod Systematics & Phylogeny*, vol. 80, pp. 75-97, 2022, doi: [10.3897/asp.80.e71492](https://doi.org/10.3897/asp.80.e71492).
- [44] S. Shimizu and K. Maeto, “A new distinctive *Darwin* wasp represents the first record of the *Ophion minutus* species-group (Hymenoptera: Ichneumonidae: Ophioninae) from Japan and the Far East, with an analysis of DNA barcode-based species delimitation in *Ophion*”, *Zoological Studies*, vol. 62, pp. 27, 2023, doi: [10.6620/ZS.2023.62-27](https://doi.org/10.6620/ZS.2023.62-27).
- [45] M. M. M. Alam, M. D. S. T. De Croos, S. Pálsson and S. Pálsson, “Mitochondrial DNA variation reveals distinct lineages in *Penaeus semisulcatus* (Decapoda, Penaeidae) from the Indo-West Pacific Ocean”, *Mar. Ecol.*, vol. 38, pp. e12406, 2017, doi: <https://doi.org/10.1111/maec.12406>.
- [46] C. Tavares and J. Gusmao, “Description of a new Penaeidae (Decapoda: Dendrobranchiata) species, *Farfantepenaeus isabellae* sp. Nov”, *Zootaxa*, vol. 4171, pp. 505–516, 2016, doi: [10.11646/ZOOTAXA.4171.3.6](https://doi.org/10.11646/ZOOTAXA.4171.3.6).
- [47] B. Rannala and Z. Yang, “Species Delimitation.” editors In C. Scornavacca, F. Delsuc and N. Galtier, *Phylogenetics in the Genomic Era*, chapter No. 5.5, 5.5:1–5.5:18, 2020.
- [48] B. C. Carstens, T. A. Pelletier, N. M. Reid and J. D. Satler, “How to fail at species delimitation”, *Molecular Ecology*, vol. 22, no. 17, pp. 4369-4383, 2013, doi: <https://doi.org/10.1111/mec.12413>.
- [49] M. H. Shirley, K. A. Vliet, A. N. Carr and J. D. Austin, “Rigorous approaches to species delimitation have significant implications for African crocodylian systematics and conservation”, *Proceeding of the Royal Society B: Biological Sciences*, vol. 281, pp. 20132483, 2014, doi: <https://doi.org/10.1098/rspb.2013.2483>.
- [50] N. Puillandre, S. Brouillet and G. Achaz, G, “ASAP: assemble species by automatic partitioning”, *Molecular Ecology Resources*, vol. 21, no. 2, pp. 609-620, 2021, doi: <https://doi.org/10.1111/1755-0998.13281>.
- [51] F. M. Bianchi and L. T. Gonçalves, “Borrowing the Pentatomomorpha tome from the DNA barcode library: Scanning the overall performance of *cox1* as a tool”, *J. Zool. Syst. Evol. Res.*, vol. 59, pp. 992–1012, 2021, doi: <https://doi.org/10.1111/jzs.12476>.
- [52] B. C. Carstens, T. A. Pelletier, N. M. Reid and J. D. Satler, “How to fail at species delimitation”, *Molecular Ecology*, vol. 22, no. 17, pp. 4369-4383, 2013, <https://doi.org/10.1111/mec.12413>.
- [53] M. H. Shirley, K. A. Vliet, A. N. Carr and J. D. Austin, “Rigorous approaches to species delimitation have significant implications for African crocodylian systematics and conservation”, *Proceeding of the Royal Society B: Biological Sciences*, vol. 281, pp. 2013-2483, 2014, doi: <https://doi.org/10.1098/rspb.2013.2483>.
- [54] A. Luo, C. Ling, S. Y. Ho and C. D. Zhu, “Comparison of methods for molecular species delimitation across a range of speciation scenarios”, *Systematic Biology*, vol. 67, no 5, pp. 830–846, 2018, doi: <https://doi.org/10.1093/sysbio/syy011>.

- [55] Yang, B., Cai, J., & Cheng, X. (2011). Identification of astigmatid mites using ITS2 and COI regions. *Parasitology research*, vol. 108, pp. 497-503, doi: <https://doi.org/10.1007/s00436-010-2153-y>.
- [56] D. P. Chobanov, S. Kaya, B. Grzywacz, E. Warchalowska-Śliwa and B. Çıplak, “The Anatolio-Balkan phylogeographic fault: a snapshot from the genus *Isophya* (Orthoptera, Tettigoniidae)”, *Zoologica Scripta*, vol. 46, no. 2, pp.165–179, 2016, doi: <https://doi.org/10.1111/zsc.12194>.
- [57] B. Çıplak, S. Kaya, Z. Boztepe and I. Gündüz, “Mountainous genus *Anterastes* (Orthoptera, Tettigoniidae): Autochthonous survival across several glacial ages via vertical range shifts”, *Zoologica Scripta*, vol. 44, pp. 534–549, 2015, doi: <https://doi.org/10.1111/zsc.12118>.
- [58] M. Heidema, M. Nuorteva, J. Hantula and U. Saarma, “*Dolerus asper* Zaddach, 1859 and *Dolerus brevicornis* Zaddach, 1859 (Hymenoptera: Tenthredinidae), with notes on their phylogeny”, *European Journal of Entomology*, vol. 101, no. 4, pp. 637-650, 2004.
- [59] S. M. Aljanabi and I. Martinez, “Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques”, *Nucleic acids research*, vol. 25, pp. 4692–4693, 1997.
- [60] C. Simon, F. Frati, A. Beckenbach, B. Crespi, H. Liu, P. Flook, “Evolution, weighting, and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain- reaction primers”, *Annals of the Entomological Society of America*, vol. 87, pp. 651-701, 1994.
- [61] M. Kearse et al., “Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data”, *Bioinformatics*, vol. 28, no. 12, pp. 1647-1649, 2012, doi: <https://doi.org/10.1093/bioinformatics/bts199>.
- [62] Blastn algorithms. <https://blast.ncbi.nlm.nih.gov/Blast.cgi> [accessed 2023 Dec 5].
- [63] K. Katoh and D. M. Standley, “MAFFT multiple sequence alignment software version 7: improvements in performance and usability”, *Molecular Biology and Evolution*, vol. 30, no. 4, pp. 772-780, 2013, doi: <https://doi.org/10.1093/molbev/mst010>.
- [64] M. Kimura, “A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences”, *Journal of Molecular Evolution*, vol. 16, pp. 111-120, 1980.
- [65] S. Kumar, G. Stecher and K. Tamura, “MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets”, *Molecular Biology and Evolution*, vol. 33, no. 7, pp. 1870-1874, 2016, doi: <https://doi.org/10.1093/molbev/msw054>.
- [66] D. Darriba, G. L. Taboada, R. Doallo and D. Posada, “jModelTest 2: more models, new heuristics and parallel computing”, *Nature Methods*, vol. 9, no 8, pp. 772-772, 2012, doi: 10.1038/nmeth.2109.
- [67] A. Stamatakis, “RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies”, *Bioinformatics*, vol. 30, no. 9, pp. 1312-1313, 2014, doi: <https://doi.org/10.1093/bioinformatics/btu033>.
- [68] A. Rambaut, “FigTree v.1.4.4. [accessed 2020 Oct 25]. <http://tree.bio.ed.ac.uk/software/figtree/>”, 2017.
- [69] J. Pons et al., “Sequence-based species delimitation for the DNA taxonomy of undescribed insects”, *Systematic Biology*, vol. 55, pp. 595-609, 2006, doi: <https://doi.org/10.1080/10635150600852011>.
- [70] J. Zhang, P. Kapli, P. Pavlidis and A. Stamatakis, “A general species delimitation method with applications to phylogenetic placements”, *Bioinformatics*, vol. 29, no. 22, pp. 2869-2876, 2013, doi: <https://doi.org/10.1093/bioinformatics/btt499>.
- [71] N. Puillandre, S. Brouillet and G. Achaz, “ASAP: Assemble species by automatic partitioning”, *Molecular Ecology Resources*, vol. 21, pp. 609–620, 2021, doi: <https://doi.org/10.1111/1755-0998.13281>.
- [72] R. Meier, K. Shiyang, G. Vaidya and P. K. Ng, “DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success”, *Systematic Biology*, vol. 55, no 5, pp 715-728, 2006, doi: <https://doi.org/10.1080/10635150600969864>.
- [73] R Core Team, “R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria”. ISBN 3-900051-07-0, URL (Web page: <http://www.R-project.org/>)(Date accessed: July 2021), 2014.
- [74] P. D. Hebert, S. Ratnasingham and J. R. De Waard, “Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species”, *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 270, no: suppl_1, pp. S96-S99, 2003, doi: <https://doi.org/10.1098/rsbl.2003.0025>.
- [75] R. Leys S. J. B. Cooper and P. Schwarz “Molecular phylogeny of the carpenter bees, genus *Xylocopa* (Hymenoptera: Apidae), based on mitochondrial DNA sequences”, *Mol Phylogenet Evol.*, vol. 17, pp. 407-418, 2000, doi: <https://doi.org/10.1006/mpev.2000.0851>.
- [76] B. N. Danforth, L. Conway and S. Ji, “Phylogeny of eusocial *Lasioglossum* reveals multiple losses of eusociality within a primitively eusocial clade of bees (Hymenoptera: Halictidae)”, *Syst Biol*, vol. 52, pp. 23-36, 2003, doi: <https://doi.org/10.1080/10635150390132687>.
- [77] M. Budak, E. M. Korkmaz and H. H. Başibüyük, “A molecular phylogeny of the Cephinae (Hymenoptera, Cephidae) based on mtDNA COI gene: a test of traditional classification”, *ZooKeys*, vol. 130, pp. 363, 2011, doi: 10.3897/zookeys.130.1466.
- [78] E. M. Korkmaz, M. Budak and H. H. Başibüyük, “Utilization of *cytochrome oxidase I* in *Cephus pygmeus* (L.) (Hymenoptera: Cephidae)”, *Turkish Journal of Biology*, vol. 35, no. 6, pp. 713-726, 2011, doi: 10.3906/biy-1003-65.
- [79] P. Y. Chen, B. Y. Zheng, J. X. Liu and S. J. Wei, “Next-generation sequencing of two mitochondrial genomes from family Pompilidae (Hymenoptera: Vespoidea) reveal novel patterns of gene arrangement”, *International Journal of Molecular Sciences*, vol. 17, no. 10, pp. 1641, 2016, doi: <https://doi.org/10.3390/ijms17101641>.
- [80] D. Fontaneto, C. Boschetti and C. Ricci, “Cryptic diversification in ancient asexuals: evidence from the bdelloid rotifer *Philodina flaviceps*”, *Journal of Evolutionary Biology*, vol. 21, pp. 580–587, 2008, doi: <https://doi.org/10.1111/j.1420-9101.2007.01472.x>.
- [81] P. D. N. Hebert, M. Y. Stoeckle, T. S. Zemplak and C. M. Francis, “Identification of birds through DNA barcodes”, *Plos Biology*, vol. 2, pp. 1657-1720, doi: <https://doi.org/10.1371/journal.pbio.0020312>.
- [82] Monaghan, et. al., “Accelerated species inventory on Madagascar using coalescent-based models of species delineation”, *Systematic Biology*, vol. 58, no. 3, pp. 298-311, 2009, doi: <https://doi.org/10.1093/sysbio/syp027>.
- [83] F. S. Ceccarelli, M. J. Sharkey and A. Zaldívar-Riverón, “Species identification in the taxonomically neglected, highly diverse, neotropical parasitoid wasp genus *Notiospathius* (Braconidae: Doryctinae) based on an integrative molecular and morphological approach”, *Molecular Phylogenetics and Evolution*, vol. 62, pp. 485–495, 2019, doi: <https://doi.org/10.1016/j.ympev.2011.10.018>.
- [84] S. Fernández- Flores, J. L. Fernández- Triana, J. J. Martínez and A. Zaldívar- Riverón, “DNA barcoding species inventory of Microgasterinae wasps (Hymenoptera, Braconidae) from a Mexican tropical dry forest”, *Molecular Ecology Resources*, vol. 13, no. 6, pp. 1146-1150, 2013, doi: <https://doi.org/10.1111/1755-0998.12102>.

- [85] D. Baum, "Reading a phylogenetic tree: the meaning of monophyletic groups", *Nature Education*, vol. 1, no. 1, pp. 190, 2008.
- [86] A. S. Lang, G. Bocksberger and M. Stech, "Phylogeny and species delimitations in European *Dicranum* (Dicranaceae, Bryophyta) inferred from nuclear and plastid DNA", *Molecular Phylogenetics and Evolution*, vol. 92, pp. 217-225, doi: <https://doi.org/10.1016/j.ympev.2015.06.019>.
- [87] R. Arrigoni et al., "Species delimitation in the reef coral genera *Echinophyllia* and *Oxypora* (Scleractinia, Lobophylliidae) with a description of two new species", *Molecular Phylogenetics and Evolution*, vol. 105, pp. 146-159, 2016, doi: <https://doi.org/10.1016/j.ympev.2016.08.023>.
- [88] C. Wang, S. Agrawal, J. Laudien, V. Häussermann, V. and C. Held, "Discrete phenotypes are not underpinned by genome-wide genetic differentiation in the squat lobster *Munida gregaria* (Crustacea: Decapoda: Munididae): a multi-marker study covering the Patagonian shelf", *BMC Evolutionary Biology*, vol. 16, no. 1, pp. 1-16, 2016, doi: [10.1186/s12862-016-0836-4](https://doi.org/10.1186/s12862-016-0836-4).
- [89] A. Luo, C. Ling, S. Y. Ho and C. D. Zhu, "Comparison of methods for molecular species delimitation across a range of speciation scenarios", *Systematic Biology*, vol. 67, no. 5, pp. 830-846, 2018, doi: <https://doi.org/10.1093/sysbio/syy011>.
- [90] M. Kekkonen and P. D. Hebert, "DNA barcode- based delineation of putative species: efficient start for taxonomic workflows", *Molecular Ecology Resources*, vol. 14, no. 4, pp. 706-715, 2014, doi: <https://doi.org/10.1111/1755-0998.12233>.



RESEARCH ARTICLE

Receive Date: 07.07.2023

Accepted Date: 07.11.2023

Hyperspectral anomaly detection with an improved approach: integration of go decomposition algorithm and laplacian matrix modifier

Fatma Küçük^{a,*}

^aAnkara Yıldırım Beyazıt University, Faculty of Engineering and Natural Sciences, Software Engineering Department, Ankara, Türkiye,
ORCID: 0000-0002-7052-362X

Abstract

In this study, a hyperspectral anomaly detection method based on Laplacian matrix (HADLAP) is proposed. This paper addresses the problem of determining covariance matrix inversion in high-dimensional data and proposes a new approach for identifying anomalies in hyperspectral images (HSIs). The study's goals are to find anomalous locations in HSIs and to deal with the problem of calculating the inversion of the covariance matrix of high dimensional data. The method is centered on two main concepts. One of them is decomposition process. The other one is detection process. First, HSI data is decomposed as a low rank and sparse matrices. Second, the sparse component of the data is used to build Mahalanobis Distance (MD). In this study, go decomposition (GoDec) algorithm is employed to decompose the data. Then, the distance is calculated by obtained matrix with aim of detection of anomalous pixels in the HSIs. The method differs from previous studies that covariance matrix in the distance is computed with Laplacian matrix and MD. Experiments conducted on three hyperspectral datasets present the superiority and effectiveness of the proposed framework in terms of detection performance with respect to state-of-the-art methods.

© 2023 DPU All rights reserved.

Keywords: Anomaly detection; low rank matrix; laplacian matrix

1. Introduction

Hyperspectral imaging captures the reflected light from hundreds of narrow bands objects throughout the earth's surface. The information of high spectral resolution of the hyperspectral image with these bands allows for the differentiation of various ground objects. The imaging technology is applied in the field of image analysis, remote sensing, classification, and target detection [1]. Hyperspectral images (HSIs) hold abundant spectral information

* Corresponding author. Tel.: 0 312 9062209
E-mail address: fatmakucuk@aybu.edu.tr

about characteristics of objects which enables to have an idea about the image scene [2]. Hyperspectral anomaly detection is an unsupervised approach that evaluates targets of interest against the background without any prior information about the target in advance [3]. By this methodology, abnormalities in HSIs are identified for usage with different purposes in applications such as camouflage detection, identification of minerals, fine agriculture, change detection etc. Hyperspectral anomaly detection has received substantial research in the literature. The most common methods for hyperspectral anomaly detection are statistical based models. Reed-Xiao (RX) detector is the most popular approach based on statistical model [4]. This approach relies on the idea of HSI the background follows a multivariate Gaussian distribution. Mahalanobis distance (MD) between the test pixels and the background is computed in order to identify the anomalous targets. Later, several expanded RX algorithms developed, including the subspace RX (SSRX) algorithm, which minimizes the effect of anomaly contamination on background estimation, and the local RX (LRX) algorithm, which models the local background using the inner and outer double windows approach [5, 6]. Support Vector Machines (SVM) and Random Forests (RF) techniques are also anomaly detection strategies based on advanced statistical model and get benefit from machine learning algorithms [7, 8].

Hyperspectral anomaly detection, known as unsupervised-based techniques, separates meaningful targets from the background without the need for prior information. Their aim is to distinguish outliers from background objects. Breaking down hyperspectral data into low rank and sparse components is a popular method that can be used to distinguish suspicious anomalies from background information. This process of decomposition promotes the identification of anomalous data by utilizing the data's inherent structure to differentiate context from potential anomalies.

A variety of low rank and sparse matrix decomposition-based techniques have been successfully used for hyperspectral anomaly detection. It is assumed that the anomalies are sparse, and background has low rank property. The advantage of this method is splitting HSIs as sparse and low rank matrices holding anomaly and background information respectively. Go Decomposition (GoDec) algorithm is one of the most used methods to decompose datasets can be either HSI or image or video [12]. The GoDec algorithm solves a convex optimization problem to separate datasets. As in the other datasets, it is used for background and foreground separations in HSIs. These separated matrices accurately detect the data's basic structure making it possible to explore anomalies. Low Rank and Sparse Matrix Decomposition (LRaSMD) model can be considered a novel strategy for finding hyperspectral anomalies as demonstrated in [13].

In addition to the previously discussed low rank and sparse matrix-based hyperspectral anomaly detection techniques, the robust subspace recovery algorithm via bi-sparsity pursuit is a further alternative decomposition strategy to GoDec. In [15], a robust subspace recovery algorithm is employed to isolate the data. MD is composed of the image's sparse components. Later, a different method based on LRaSMD was proposed in [17] that uses a Laplacian matrix to reconstruct MD. During the alteration of the distance function used to calculate the distance is employed as a weight function. In the context of this research, low rank and sparse matrices from the HSI dataset have been created using the GoDec algorithm. A map that detects anomalous behavior is created by executing MD on the sparse matrix. Notably, the Laplacian matrix is employed to invert the covariance matrix in MD, which increases the accuracy and effectiveness of anomalous behavior detection. Therefore, a new anomaly detection method for HSIs, HADLAP, is constructed and proposed.

This study is organized into various sections to present proposed method for hyperspectral anomaly detection. Section 2 provides information about the proposed method in detail. The datasets utilized for assessment are expressed in Section 3. Empirical findings are demonstrated in Section 4 where evaluations for proposed method with state-of-the-art methods is given. This section emphasizes how well the suggested strategy performs and how successful it is in finding anomalies in hyperspectral data. conclusion of the study is drawn in Section 5. Finally, Section 6 includes a discussion that highlights the importance of the sparse and low rank matrix decomposition based technique for hyperspectral anomaly identification.

2. Experimental design

In this section, the hyperspectral image is first separated aside GoDec method, which yields the low rank and sparse matrices. The sparse component is next substituted to MD. Cauchy function is used in place of explicitly inverting covariance matrix, allowing accurate estimates for anomalous pixels with comparable characteristics to be captured. The Laplacian matrix is then used to derive background statistics. This strategy has a number of beneficial aspects. First of all, only relevant bits of the data are inverted, which considerably lowers the cost of calculation. Additionally, it prevents the use of inaccurate statistics. As a result, the technique achieves high computing efficiency. By using the Laplacian matrix in MD after these stages, as opposed to computing the inverse of the complete dataset, the final anomaly map is created. This hybrid method, which combines matrix decomposition methods with modified MD computations, works well for locating hyperspectral anomalies.

Hyperspectral data in three dimensions is initially converted to a 2D matrix. Thus, the data $\mathbf{X} \in \mathbb{R}^{h \times w \times b}$ may be expressed by $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_b]$, $\mathbf{X} \in \mathbb{R}^{N \times b}$ in which N is the total number of pixels and b represents the quantity of bands in a spectrum. The data matrix \mathbf{X} is written as in Eq. 1

$$\mathbf{X} = \mathbf{L} + \mathbf{E} \quad (1)$$

where $\mathbf{L} \in \mathbb{R}^{N \times b}$ is the background matrix and $\mathbf{E} \in \mathbb{R}^{N \times b}$ is the sparse matrix that are two dimensional matrices with the size $N \times b$. The optimization problem is then formulated as in Eq. 2.

$$\min_{\mathbf{L}, \mathbf{E}} \|\mathbf{X} - \mathbf{L} - \mathbf{E}\|_F^2, \text{rank}(\mathbf{L}) \leq r, \text{card}(\mathbf{E}) \leq kN \quad (2)$$

where r is the maximal rank of \mathbf{L} and k is the cardinality of \mathbf{E} . In order to solve the problem in Eq. 2 GoDec algorithm below is applied. After applying the algorithm, low rank \mathbf{L} and sparse \mathbf{E} matrices are extracted. By extracting \mathbf{E} , MD distance in Eq. 3 is built.

Algorithm 1. GoDec Algorithm

Inputs:

$\mathbf{X} \in \mathbb{R}^{h \times w \times b}$; data matrix, r ; rank of \mathbf{L} matrix, s ; cardinality of \mathbf{E} matrix, Iter; max iteration number

Outputs:

\mathbf{L} ; low rank matrix, \mathbf{E} ; sparse matrix

Initialize: $\mathbf{L}_0 = \mathbf{X}, \mathbf{S}_0 = \mathbf{0}, i := 0, \mathbf{A}_1 = \text{randn}(\mathbf{L}, r)$;

Repeat: Iter times

- 1) $i := i + 1$;
- 2) $\mathbf{Y}_1 = (\mathbf{X} - \mathbf{E}_{i-1})\mathbf{A}_1, \mathbf{A}_2 = \mathbf{Y}_1; \mathbf{Y}_2 = (\mathbf{X} - \mathbf{E}_{i-1})^T \mathbf{A}_2$;
- 3) If $\text{rank}(\mathbf{A}_2^T \mathbf{Y}_1) < r$ then $r := \text{rank}(\mathbf{A}_2^T \mathbf{Y}_1)$, go to step 2); end;
- 4) $\mathbf{L}_i = \mathbf{Y}_1 (\mathbf{A}_2^T \mathbf{Y}_1)^{-1} \mathbf{Y}_2^T$
- 5) $\mathbf{E}_i = \text{P}_{\Omega}(\mathbf{X} - \mathbf{L}_{i-1})$ until $\|\mathbf{X} - \mathbf{L} - \mathbf{E}\|_F^2 / \|\mathbf{X}\|_F^2 < \varepsilon$

$$D(\mathbf{E}) = (\mathbf{E} - \mu)^T \Gamma^{-1} (\mathbf{E} - \mu) \quad (3)$$

where μ is mean and Γ^{-1} is inverse of the covariance matrix of \mathbf{E} . Each pixel in the \mathbf{E} matrix is evaluated by considering itself and its neighboring pixels. In addition, Laplacian matrix and Cauchy function are implemented to calculate Γ^{-1} unlike other studies. Cauchy distance with spatial variant is applied. It returns the likelihood map of each pixel to be anomalous. Each pixel is evaluated by considering itself and its four-connected neighbors. Thus, the distance formula becomes as in Eq. 4

$$D_{HADLAP}(E) = (E - \mu)^T L (E - \mu) \quad (4)$$

where the Laplacian matrix L is used for instead of Γ^{-1} by modifying MD. L matrix is computed by the following equations:

$$L = D - W \quad (5)$$

where D and W are degree and weight matrices respectively. W in Eq. 6 is Cauchy function by which pixels representing similar features are detected,

$$W_{xy} = \frac{1}{1 + \left(\frac{\mu_x + \mu_y}{\beta}\right)^2} \quad (6)$$

where μ_x and μ_y are means for band images x and y and β is a scaling parameter. Therefore, the anomaly map is extracted by computing the distance D . The proposed method's outcomes and performance results of detectors are given in the next section. The weights are normalized since it is more beneficial and preferred. Then, using Eq. 7,

$$L = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} \quad (6)$$

the symmetric normalized Laplacian matrix is calculated. Finally, anomaly detection maps for each image are obtained.

3. Experimental data

The information about the data used for assessments is provided in this part. The following three hyperspectral datasets are employed: Airport Beach Urban (ABU) Airport_3, ABU Urban_1 and Salinas implemented 4 (Imp_4) datasets. Figure 1 shows the original band images. Their ground truth images are presented in Figure 2.

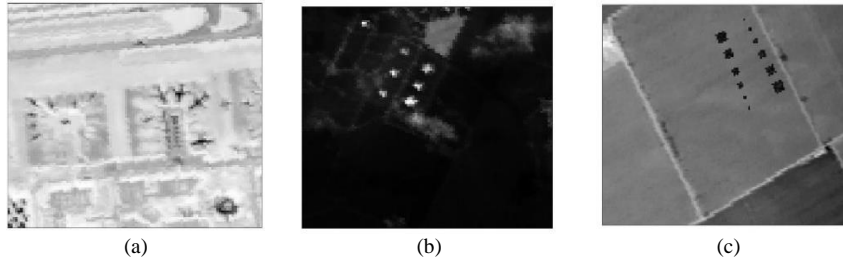


Fig. 1. Original band HSI. (a) Airport_3. (b) Urban_1. (c) Salinas Imp_4.

The first dataset in Figure 1 (a) has a resolution of 7.1m and 100 x 100 pixels with 205 spectral bands. The data is gathered by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) in Los Angeles. The second dataset has a resolution of 17.2m and 100 x 100 pixels with 204 spectral bands is presented in Figure 1 (b). The data is collected by AVIRIS in Texas coast. Salinas Hyperspectral Dataset in Figure 1 (c) is lastly used hyperspectral dataset which is a popular data set used for several remote sensing applications. It obtains 224 bands which are also captured AVIRIS to provide the dataset for this investigation over an agricultural region close to Salinas Valley in California, USA. The original collection, which has a resolution of 512x217 pixels, includes images of 16 various kinds of vegetables, bare soils, and vineyard fields. The subset is sized as 126x150 with 204 spectral bands. Figure 2 presents

ground truth images for all datasets. In Fig. 2a and 2b, planes are anomalies for ABU Airport_3 and ABU Urban_1. Anomalies are artificially implemented in Salinas. They are not real objects. The steps taken to create anomalies in the image are explained in depth in [18].

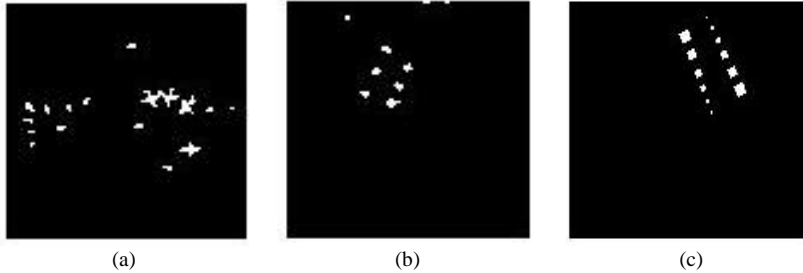
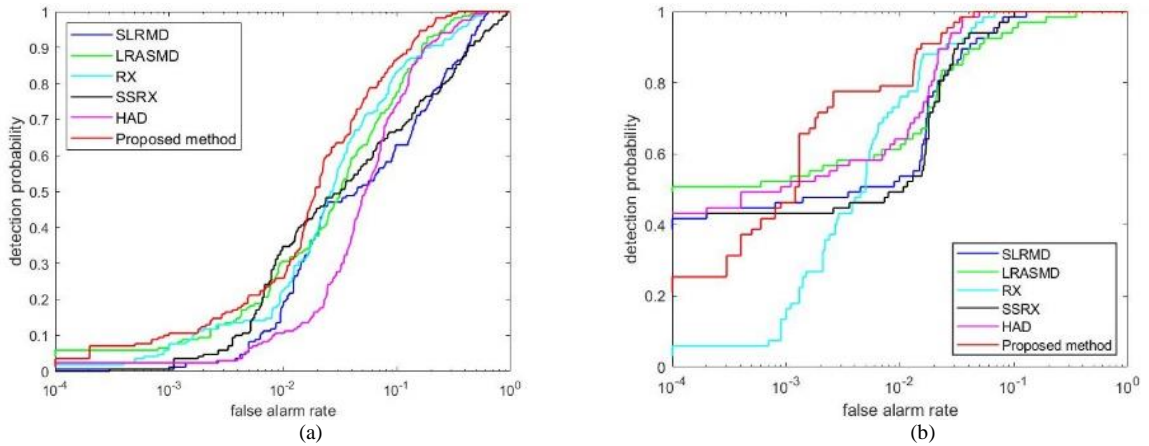


Fig. 2. Ground truth images. (a) Airport_3. (b) Urban_1. (c) Salinas Imp_4.

4. Experimental results

The experimental findings from an extensive research study were carried out to determine the effects of the proposed method compared with the state-of-the-art of the methods. Five anomaly detection techniques are compared with HADLAP: HADM, SLRMD, LRASMD, RX, and SSRX. In order to examine the accuracy of various approaches, Receiver Operating Characteristic (ROC) curves and Area Under the ROC curve (AUC) measures are applied in this study. Figure 3 shows ROC curves of each datasets which includes six hyperspectral anomaly detection results false alarm rate (far) versus detection probability (pd).



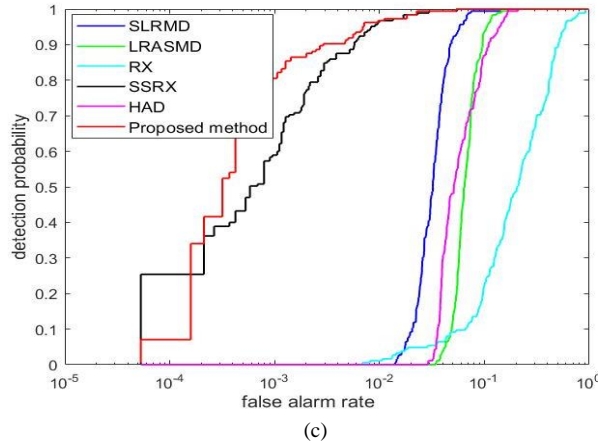
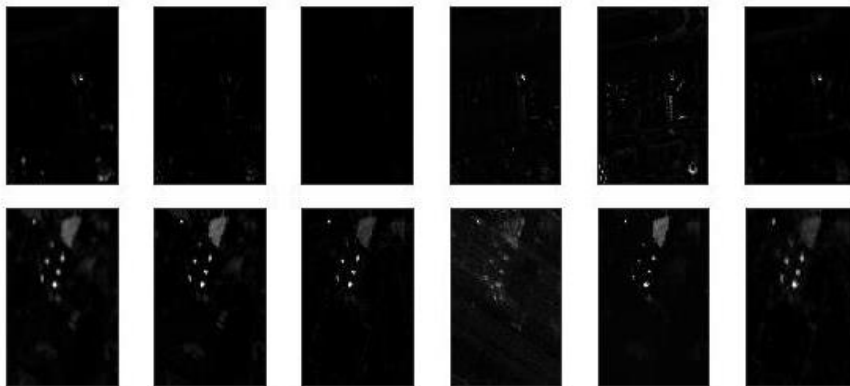


Fig. 3. ROC curves of (a) Airport_3. (b) Urban_1. (c) Salinas Imp_4.

The graph in Figure 3 (a) demonstrates the variety of methods used, from conventional methods to novel approaches. The suggested technique distinguishes out from the other methods since it has less far with higher pd. As may be observed from the figure that suggested technique is always above the other methods especially after 0.01far. LRASMD and RX come after suggested approach. LRASMD has higher pd than RX while its line approaching 100% pd. HADM and SLRMD is after them. It can be concluded that GoDec algorithm provides better detection rate than RoSuRe decomposition algorithm. While distance calculation is compared, modified MD gives better detection rate than original MD.

In the following Figure 3 (b) is ROC curve for ABU urban_1 datasets. The suggested technique performs better than the alternatives, producing better outcomes in terms of far and pd. It stands out for showing a distinct detection rate after a FAR of 0.001, while keeping a 100% PD rate from 50%. The suggested strategy is successful, as evidenced by the decline in the false alarm rate as the detection rate rises. In terms of performance, the HAD approach is quite similar to the suggested methodology.

In Figure 3 (c), the results for the Salinas imp_4 image are presented. The proposed method demonstrates a higher pd compared with the other methods. The SSRX approach, the second-most competitive technology, comes closely behind. The SLRMD, HADM, LRASMD, and RX techniques show increasingly greater detection rates in descending order. The figure highlights HADM superiority regarding Salinas imp_4 dataset detection performance.



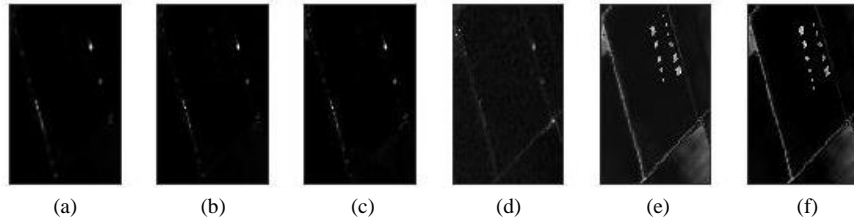


Fig. 4: 2D results for (a) HADM. (b) SLRMD. (c) LRaSMD. (d) RX. (e) SSRX. (f) HADLAP.

Figure 4 shows two dimensional results for each image obtained by the evaluated methods. The last column in Figure 4 (f) is the result for the proposed method named HADLAP. Analyzing the figure reveals that the proposed method regularly results in images with greater clarity, finer details, and overall visual representation, making it stand out as a top candidate among the five approaches examined. In addition, the suggested method's higher visual quality and distinctiveness are highlighted noticeably in the figure, underlining its promise as a sophisticated and cutting-edge way of hyperspectral image processing. By comparing the proposed approach, HADLAP, to existing modern techniques, the assessment of ROC findings and AUC values in Table 1 reveals that HADLAP performs better overall.

Table 1 presents the AUC findings for five different approaches that were used on three different HSIs. One of these techniques stands out as the suggested strategy. The accuracy and discrimination performance of each approach are shown by the AUC metric, which also acts as an assessment metric.

Table 1. AUC values for the methods.

HSI data	AUC VALUES					
	Proposed	HADM	SLRMD	LRaSMD	RX	SSRX
Airport_3	0.9561	0.9157	0.8712	<u>0.9346</u>	0.9300	0.8623
Urban_1	0.9950	<u>0.9914</u>	0.9849	0.9796	0.9907	0.9850
Salinas	0.9984	0.9358	0.9660	0.9313	0.7463	<u>0.9978</u>

We may evaluate the effectiveness of the suggested strategy in comparison to the alternatives by comparing the AUC scores. By comparing the proposed approach's AUC measures to the results attained using the other methodologies, the table highlights the worth of the suggested approach. The proposed method has the highest AUC values which are 0.956, 0.995, and 0.998 respectively. It is followed by RX for Airport_3 data. For Urban_1 image, HADM has the second highest AUC value. Finally, the second-highest AUC value is held by SSRX in Salinas dataset.

5. Conclusion

In this study, low rank and sparse matrix decomposition-based method is proposed. GoDec algorithm is applied for low-rank in order to get background and foreground information. Modified MD is then adopted to get anomaly pixels. Different than previous studies MD is rebuilt by using Laplacian matrix. This is resulted with higher detection performance. There hyperspectral datasets are used for evaluations. Anomalies in Salinas data are artificially implemented. Five state-of-the-art approaches are compared with the suggested method. Two of these methods are the most known traditional anomaly detection methods; RX and SSRX. The others are LSDM based methods similar approaches with the proposed methods. ROC curves and AUC are used for evaluations. According to these performance analysis, proposed method outperforms the other methods with higher detection rate overall 90%.

6. Discussions

Here This section discusses five alternative approaches for finding hyperspectral anomalies using three different hyperspectral images. The five methods that are taken into account in this study are HADLAP, HADM, LRaSMD, RX, and SSRX. To find anomalies in the hyperspectral images, each strategy makes use of a different collection of algorithms, methods, or statistical techniques. Various approaches in an effort are compared to find the best successful strategy for hyperspectral anomaly detection. RX and SSRX are traditional anomaly detection methods. The others, however, employ strategy of low rank and sparse matrix decomposition. HADLAP and HADM use RoSuRe algorithm whereas LRaSMD uses GoDec algorithm to partition the data. When all findings considered decomposition based methods outperforms traditional ones in terms of anomaly detection. If the suggested approach and LRaSMD which both employ the GoDec algorithm for decomposition but employ various distances are compared, it may be concluded that the modified MD by the laplacian matrix performs better than MD. Overall, this thorough analysis of five approaches for detecting hyperspectral anomalies using three hyperspectral data gives researchers insightful information that will help them choose the best strategy for precisely identifying anomalies in subsequent applications.

Acknowledgements

This research received no specific grants from any funding agency in public, commercial or non-profit sectors.

References

- [1] C. I. Chang, *Hyperspectral data processing: algorithm design and analysis*. Maryland, USA: Wiley, 2013.
- [2] C. I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Maryland, USA: Springer, 2003.
- [3] W. Li, G. Wu, and Q. Du, "Transferred deep learning for anomaly detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 597-601, May 2017, doi: 10.1109/LGRS.2017.2657818.
- [4] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 10, pp. 1760-1770, 1990, doi: 10.1109/29.60107.
- [5] D. W. J. Stein et al., "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58-69, Jan. 2002, doi: 10.1109/79.974730.
- [6] A. P. Schaum, "Hyperspectral anomaly detection beyond RX," *Proc. SPIE*, vol. 6565, pp. 13-25, May 2007, doi: 10.1117/12.718789.
- [7] T. C. M. Rao, G. J. Sankar, and T. R. Kumar, "A hierarchical hybrid SVM method for classification of remotely sensed data," *J Indian Soc Remote Sens.*, vol. 40, pp. 191-200, June 2012, doi: 10.1007/s12524-011-0149-4.
- [8] L. Wan et al., "Collaborative active and semisupervised learning for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2384-2396, Nov. 2014, doi: 10.1109/TGRS.2014.2359933.
- [9] H. Su et al., "Hyperspectral anomaly detection: A survey," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, pp. 64-90, Sept. 2021, doi: 10.1109/MGRS.2021.3105440.
- [10] Y. Xu et al., "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54(4), pp. 1990-2000, Nov. 2015, doi: 10.1109/TGRS.2015.2493201
- [11] W. Xie et al., "Spectral adversarial feature learning for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 58(4), pp. 2352-2365, Nov. 2019, doi: 10.1109/TGRS.2019.2948177.
- [12] T. Zhou and D. Tao, "Godec: Randomized low-rank & sparse matrix decomposition in noisy case," *Proc. 28th Int. Conf. Machine Learn. (ICML)*, 2011, pp. 33-40.
- [13] W. Sun et al., "Low-rank and sparse matrix decomposition-based anomaly detection for hyperspectral imagery," *J. Appl. Remote Sens.*, vol. 8, no. 1, pp. 083641, 2014, doi: 10.1117/1.JRS.8.083641
- [14] G. Lerman and T. Maunu, "An overview of robust subspace recovery," *Proc. of IEEE*, vol. 106(8), pp. 1380-1410, Aug. 2018, doi: 10.1109/JPROC.2018.2853141
- [15] F. Küçük, B. Töreyn and F. V. Çelebi, "Sparse and low-rank matrix decomposition-based method for hyperspectral anomaly detection," *J. Appl. Remote Sens.*, vol. 13, no. 1, pp. 014519, Feb. 2019, doi: 10.1117/1.JRS.13.014519.
- [16] F. Küçük, B. U. Töreyn, F. V. Çelebi, "Anomaly detection in hyperspectral data with matrix decomposition," in: 26th Signal Processing and Communications Applications Conference, SIU, 2018, pp. 1-4, doi: 10.1109/SIU.2018.8404658.

- [17] F. Küçük, "Hybrid anomaly detection method for hyperspectral images," *Signal Image Video Process.*, vol. 17, pp. 2755-2761, Jan. 2023, doi: 10.1007/s11760-023-02492-4.
- [18] F. Verdoja and M. Grangetto, "Graph Laplacian for image anomaly detection," *Mach. Vis. Appl.*, vol. 31, no. 1–2, Jan. 2020, doi: 10.1007/s00138-020-01059-4.
- [19] F. Zhang, F. He, and H. Hu, "Laplacian matrix graph for anomaly target detection in hyperspectral images," *Elec. Lett.*, vol. 58(8), pp. 312-314, Feb. 2022, doi: 10.1049/ell2.12449.
- [20] L. J. Grady and J. R. Polimeni, *Discrete Calculus: Applied Analysis on Graphs for Computational Science*. London, UK, Springer, 2010.



RESEARCH ARTICLE

Receive Date: 19.02.2024

Accepted Date: 18.03.2024

Evaluation of overall equipment effectiveness (OEE) for mining equipment (shovel-truck): A case study Manisa-Soma (Turkey) open pit mine

Sedat Toraman^{a*}

^aGeneral Directorate of Turkish Coal Enterprises, Ankara, 06560, Türkiye, ORCID: 0000-0003-2503-7320

Abstract

Mining is a costly activity that requires huge investments and large equipment. Therefore, it is crucial to use mining equipment efficiently to maximize productivity. This requires the continuous analysis of equipment efficiency and taking action to reduce negative impacts. Accurately estimating equipment efficiency is essential to increase its effectiveness. One globally recognized and accepted measurement is Overall Equipment Effectiveness (OEE). In this study was conducted to calculate the OEE values of shovels and trucks, which are the primary equipment used in open pit mining. The study presented different loss times and quality values while calculating the OEE values of mining equipment. The OEE values of electric trucks and shovels were found to be consistent with actual work quantities in the Manisa-Soma Lignite open pit operation. This study is the first applied study of OEE analysis in Turkey.

© 2023 DPU All rights reserved.

Keywords: OEE; Equipment Utilization; Mining; Time Losses; open pit equipment

1. Introduction

The speed at which modernization and Industry 4.0 are spreading around the globe has raised the capacity of mining machines and equipment. This capacity growth has brought attention to the necessity of using the gear and equipment even more effectively, nevertheless. Analysing the overall productivity of the equipment utilized in mining operations is therefore considerably more important. As in other sectors of business, mining businesses now need to make sure that their machinery and equipment are used as efficiently as possible. Utilizing equipment at its optimal efficiency level is crucial and is contingent upon the operational parameters of the business. In the event that the operational efficiency falls to a level below that which is required, it is very important that quickly to one responds improving it. By checking efficiency of equipment and action just in time in order to correct, businesses can run their level of equipment with that which is required, lowers business production cost, and in investment cost at the same time reduces.

* Corresponding author. Tel.: 0 312 5401000, E-mail address: sedat.toraman@gmail.com

Overall Equipment Efficiency, OEE, Analysis provides performance management of the equipment required for the mining operations. Since the method provides fast and accurate answers, many different businesses use this analysis method.

S. Nakajima has defined total Productive Maintenance (TPM) as having its positives and negatives. According to the author's comparison of TPM to OEE, the author states a point of simplicity and being user-friendly to the user [1]. Sharma et al. found that a relevant study of a whole range of manufacturing process equipment efficiency reflects competitiveness, hence would be ideal for measuring total equipment effectiveness and efficiency [2]. K. Yagi. To calculate their OEE in the industry, Jeong et al. used a different technique for loss prediction over the time production lines [3]. M. Braglia et al calculated the OEE of all equipment in the production line using a different method [4]. R. Oechsner et al proposed new approaches and calculation methods to determine overall plant efficiency instead of overall equipment efficiency [5]. J.A. Garza-Reyes et al investigated the correlation between system capacity and overall equipment efficiency (OEE) [6].

S. Elevli and his team applied theoretical OEE calculation methods and formulas to mining equipment in their study [7]. N.R. Sharma conducted a study on the relationship of OEE with the implementation of Preventative Maintenance systems for the reduction of equipment downtime, especially at the availability front of the equipment [8]. J.M. Akande et al. did a study on equipment optimization conducted on loaders and rigid frame trucks in a mining operation [9]. Conveyer and bucket-based backhoes efficiency study was done by M. Mohammadi et al. [10].

The idea of the study was to determine how efficient the electric trucks and shovels operating at the Manisa-Soma coal works. Data on shovels and trucks in Manisa-Soma (Turkey) coal mine for the years 2021 and 2022 were collected and analysed. Among the findings of this analysis was the determination of the state of equipment efficiency as well as suggestions on how overall equipment efficiency figures can be raised.

2. Methodology

OEE is an analytics tool that enables enterprises to make decisions with concerns to the already available equipment in the enterprise. After the 2000s, that kind of analysis was much favoured because it's logically simple and not too time-consuming. Following completion of the OEE assessment process, if the percentage OEE is found to be less than the targeted level, then some of the factors that need to be analysed include operator quality, down time of equipment, repair and installation times, lost time due to minor amount of equipment malfunction and problems which are caused due to the speed with which the equipment are operated. A business has the opportunity to become more efficient if it is able to identify and correct these problems.

OEE is one of the important techniques of progress control over and improvement in efficiency of equipment used in the mining operation. This method accounts for the losses incurred during the work process and is used to calculate real efficiency. Table 1 displays the six most prevalent major loss times. [11].

Table 1. Six Big Losses [11].

Factor	Six Big Loss Category	OEE Loss Category	OEE Factor
Mechanical	Equipment Failure	DowntimeLosses	Availability (A)
Human	Setup and Adjustment		
Operational	Idling and Minor Stoppages	Speed Losses	Performance (P)
Mechanical or Operational	Reduced Speed		

Mechanical	Reduced Yield	Defect Losses	Quality(Q)
Operational	Quality Defects		

Production line equipment availability is affected by any losses that occur during use, such as downtime and repair times. The availability can be calculated using the formula 1.

$$\text{Availability (\%)} = \frac{\text{Net Available Hours} - \text{Downtime Losses}}{\text{Net Available Hours}} * 100 \quad (1)$$

$$\text{Net available hours} = \text{Total time-planned downtime}$$

Equipment performance is affected by speed losses resulting from user error and operating conditions. Formula 2 can be used to calculate the speed loss.

$$\text{Performance (\%)} = \frac{\text{Operating Times} - \text{Speed Losses}}{\text{Operating Times}} * 100 \quad (2)$$

$$\text{Operation Times: Net available hours-Downtime losses}$$

"Quality" refers to a product's standard and includes the consideration of "product loss". The formula 3 is used to determine the amount of product loss.

$$\text{Quality (\%)} = \frac{\text{Net Operating Times} - \text{Defect Losses}}{\text{Net Operating Times}} * 100 \quad (3)$$

$$\text{Net operation time} = \text{Operation time} - \text{Speed losses, Defect Losses: Operator defects}$$

Once the Overall Equipment Effectiveness (OEE) has been analysed, it needs to be checked for compliance with industry standards. The acceptable average efficiency in the industry is typically around 85-90%. If the calculated total equipment efficiency value falls below the specified operating efficiency value, corrective measures are taken to increase system efficiency. It is important to base the calculated OEE value on accurate and realistic data, as using unrealistic data can lead to incorrect equipment efficiency readings. To calculate OEE, the actual total time, planned downtime, and unplanned downtime within this total time should be accurately calculated.

2.1. Calculating the Overall Equipment Effectiveness (OEE) for Mining Equipment

Mining operations are significantly distinct and more intricate than the manufacturing sector. As a result, both the expected and unexpected losses for mining equipment must be analysed based on their operating systems. However, collecting data on the planned time and losses for mining equipment is a more complex process. Table 2 illustrates the procedure for determining a truck's Overall Equipment Effectiveness (OEE). Due to differences in mining equipment, the collection of data can vary. However, the process is often challenging due to the following reasons:

- Mining activities involve several operations, such as drilling, blasting, excavation, loading and unloading. The efficiency of the equipment used in these operations is highly dependent on the efficiency of the previous operation. Therefore, to analyse the overall efficiency of the entire system, it is crucial to consider the efficiency of each operation.
- The productivity of mining operations is greatly impacted by the high equipment capacities used.
- The severe working conditions found in mining operations can have a negative effect on OEE efficiency, which lowers standard productivity figures.
- Elements in the working environment, like dust and lighting, have an impact on the efficiency of mining equipment. [12].
- The matching factor between trucks and shovels is a crucial consideration in open pit mining OEE (Overall

Equipment Effectiveness) calculations. For the truck and the bucket, the matching factor should ideally be 1. The truck will wait for the bucket to load it, increasing the loading time, if the matching factor is less than 1. Productivity will suffer if the matching factor is less than 1.

- In addition to the truck and shovel filling factor, the quality value calculation should consider the performance of operators operating the equipment.

Table 2. The methodology for determining a truck's Overall Equipment Effectiveness (OEE).

Loss Classification	Description
Non-Scheduled Time	1. Unplanned time for operation of the equipment. 2. Planned time for periodic maintenance of the truck.
Scheduled Maintenance Time	Time spent on breakdown.
Unscheduled Maintenance Time	Equipment preparation and setup time
Setup And Adjustment Time	Equipment operational but downtime due to other factors
Idle Time Without Operator	Time duration for which truck waits to get position to be loaded
Loading Time Loss	The time when the truck waits to be loaded.
Loss of working Conditions	Time loss due to management, supervision, climate, and job conditions
Speed Loss	Time loss due to the equipment that is operating under the standard speed
Quality Loss	1. Fill factor of loader 2. Loader operator efficiency

3. Overall Equipment Effectiveness Value Calculation

3.1. Overall Equipment Effectiveness values for trucks



Fig. 1. 630ES Electric Trucks.

The operation has ten Komatsu-630ES Electric trucks, each with a capacity of 170 short tons (Fig. 1). These trucks are primarily used for transporting excavated materials and were added to the system back in 1999. The distance between the open pit mine and the dump site is 4.7 km, and it takes approximately 19 minutes for the trucks to complete the tour. Tables 3 and 4 provide statistical data on truck usage in the business for the years 2021 and 2022. In this study, the OEE values for electric trucks in the field were determined for 2021 and 2022 to observe any changes over the years.

Table 3. Statistical Data for Trucks in 2021.

Model	Truck Number	Programme (Hour)	Actual amount of work (Ton)	Failure total (Hour)	Loss of working conditions (weather opposition, etc.) (Hour)
HAULPAK-KO/630 ES	548	5.535	345.982	2.364	862
HAULPAK-KO/630 ES	549	6.128	601.300	1.275	837
HAULPAK-KO/630 ES	550	5.633	381.500	2.488	551
HAULPAK-KO/630 ES	551	6.188	674.541	792	818
HAULPAK-KO/630 ES	552	6.435	808.192	404	929
HAULPAK-KO/630 ES	553	6.255	714.896	686	983

HAULPAK-KO/630 ES	554	6.293	638.491	558	970
HAULPAK-KO/630 ES	555	6.233	718.921	549	889
HAULPAK-KO/630 ES	556	6.285	728.721	800	712
HAULPAK-KO/630 ES	557	6.225	675.955	370	987

Table 4. Statistical data for Trucks in 2022.

<i>Model</i>	<i>Truck Number</i>	<i>Programme (Hour)</i>	<i>Actual amount of work (Ton)</i>	<i>Failure total (Hour)</i>	<i>Loss of working conditions (weather opposition, etc.) (Hour)</i>
HAULPAK-KO/630 ES	548	6.383	638.678	512	628
HAULPAK-KO/630 ES	549	6.720	775.580	236	728
HAULPAK-KO/630 ES	550	5.288	62.720	3.642	137
HAULPAK-KO/630 ES	551	6.383	627.313	572	714
HAULPAK-KO/630 ES	552	6.548	611.974	820	690
HAULPAK-KO/630 ES	553	5.970	388.972	2.483	167
HAULPAK-KO/630 ES	554	6.660	705.234	265	703
HAULPAK-KO/630 ES	555	6.503	641.916	552	676
HAULPAK-KO/630 ES	556	6.345	633.101	378	672
HAULPAK-KO/630 ES	557	6.473	702.972	639	511

Table 5 displays the time durations of the different parts of Truck No. 548. In Real-time OEE calculations, the total time determined by field authorities under actual working conditions was taken into account. The amount of work carried out by the truck was compared to the amount of work it was supposed to do under those working conditions. In calculating the work done, real data such as a truck capacity of 170 short tons (154 tons) and a truck tour time of 19 minutes were used.

Table 5. Time Lengths of the Elements of Truck No. 548/2021.

Item	Description	Time (Hour)
Total Duration	24 hours/day*30 days/month*12 months/year	8640
<i>Scheduled Time (1): (Administrative Leave)</i>	2 days/month*24 hours/day*12 months/year	576
<i>Scheduled Time (2): (Meal and refreshment break)</i>	1 hour/shift*3 Shifts/day*30 days/month* 12 months/year	1080
<i>Scheduled Time (3): Time not intended for operation</i>	60 days/year*24 hours/day	1440
Total scheduled time=1+2+3		3096
Planned Maintenance	0.1 day/month* 24 hours/day* 12 months	28,8
Unplanned Failure Stops		2364
Installation and Setup	0.4 hours/shift* 3 shifts/day* 24 hours/day* 12 Months	432
Idle Time	0.6 hours/shift* 3 shifts/day* 24 hours/day* 12 Months	648
Loading Waiting Time	0.1 hour/shift* 3 shifts/day* 24 hours/day* 12 Months	108
Loss of loading time	0.3 hours/shift* 3 shifts/day* 24 hours/day* 12 Months	324
Loss of working Conditions	Losses due to weather, etc.	862
Speed Loss	0.4 hours/shift* 3 shifts/day* 24 hours/day* 12 Months	432
Quality Loss	Filling Factor (87%)*Loader Operator Factor (95%) = 82.65	

Truck 548's Availability, Performance, Quality, and OEE values were calculated using the data in Table 6.

$$\begin{aligned}
 \text{Net available hours} &= 8640 - (3096+28,8) = 5515,20 \\
 \text{Downtime Loses} &= 2364+432+648+108 = 3552 \\
 \text{Operation Times} &= 5515,20 - (324+862+432) = 3897,20 \\
 \text{Speed Losses} &= 324+862+432 = 1618
 \end{aligned}
 \quad \longrightarrow \quad
 \begin{aligned}
 \text{Availability} &= (5515,20-3552)/5515,20 = 0,3559 \\
 \text{Performance} &= (3897,20-1618)/3897,20 = 0,5848 \\
 \text{Quality} &= 0,87*0,95 = 0,8265
 \end{aligned}$$

Table 6. The OEE Calculations of Truck No:548-2021.

<i>Calculation Based on Application Time</i>	
<i>Net available hours</i>	5515,20
<i>Availability</i>	0,3559
<i>Performance</i>	0,5848
<i>Quality</i>	0,8265
OEE	Availability*Performance*Quality
	0,1721

Table 7. The calculation is based on the application time of 2021.

<i>Truck Number</i>	<i>Availability</i>	<i>Performance</i>	<i>Quality</i>	<i>OEE (%)</i>	<i>Actual Work Quantity (Ton)</i>	<i>Production Quantity (Tons) (OEE at 50% level)</i>
548	35,61	58,48	82,65	17,21	345.982	: (%50*345.982)/%17,21 = 1.005.164
549	59,76	64,81	82,65	32,01	601.300	939.152
550	34,82	69,84	82,65	20,10	381.500	949.028
551	67,65	65,38	82,65	36,55	674.541	922.714
552	75,44	64,86	82,65	40,44	808.192	999.239
553	69,98	61,38	82,65	35,50	714.896	1.006.926
554	72,13	61,97	82,65	36,94	638.491	864.149
555	72,27	64,39	82,65	38,46	718.921	934.666
556	68,26	69,39	82,65	39,15	728.721	930.676
557	75,13	61,45	82,65	38,15	675.955	885.823
				Total	6.288.499	9.437.537

Table 8. The calculation is based on the application time of 2022.

<i>Truck Number</i>	<i>Availability</i>	<i>Performance</i>	<i>Quality</i>	<i>OEE (%)</i>	<i>Actual Work Quantity (Ton)</i>	<i>Production Quantity (Tons) (OEE at 50% level)</i>
548	73,25	59,99	82,65	36,32	638.678	879.224
549	78,81	60,13	82,65	39,17	775.580	990.092
550	8,52	66,51	82,65	4,69	62.720	669.322
551	72,33	57,35	82,65	34,29	627.313	914.790
552	69,36	59,82	82,65	34,29	611.974	892.276
553	38,57	70,68	82,65	22,53	388.972	863.187
554	78,19	60,36	82,65	39,01	705.234	903.934
555	73,25	59,84	82,65	36,23	641.916	885.844
556	75,33	58,57	82,65	36,47	633.101	868.037
557	71,81	64,48	82,65	38,27	702.972	918.441
				Total	5.788.460	8.785.148

After analyzing the OEE values and actual work quantities of electric trucks operating in the field between 2021 and 2022, as shown in Table 7-8, we found that truck 548 had the lowest OEE value of 17.21% in 2021, and truck 550 had an OEE value of 4.69% in 2022. We observed that trucks with the lowest OEE value also recorded the lowest work quantities. For example, in 2021, truck 548 spent 2,364 hours in malfunction out of a planned work time of 5,515 hours. Similarly, in 2022, truck 550 had a planned work time of 5,280 hours, but its unplanned downtime was 3,642 hours.

It was observed that the trucks which had the highest OEE values also had the highest actual work quantities. In 2021, truck 552 recorded the highest OEE value of 40.44%, and in 2022, truck 549 recorded 39.17%. Additionally, it was noted that the trucks with the highest OEE value also produced the highest output when considering the actual work quantities. If the OEE value hits 50%, the total production of 6,288,499 tons in 2021 could potentially increase to 9,437,537 tons, resulting in a 66% increase.

Table 9. Illustrates the correlation between unplanned downtime and Overall Equipment Effectiveness (OEE).

2021					2022				
Truck Number	Total Duration Time (Hour)	Unplanned Failure Stops (hour)	(Unplanned Failure Stops (hour))/(Net available hours)*100	OEE %	Truck Number	Total Duration (Hour)	Unplanned Failure Stops (hour)	(Unplanned Failure Stops (hour))/(Net available hours)*100	OEE %
548	5515,2	2364	42,85	17,21	550	5280	3642	68,98	4,69
550	5640	2488	44,11	20,1	553	5976	2483	41,55	22,53
549	6120	1275	20,83	32,01	551	6360	572	8,99	34,29
553	6240	686	10,99	35,5	552	6552	820	12,51	34,29
551	6120	792	12,94	36,55	555	6504	552	8,48	36,23
554	6264	558	8,90	36,94	548	6355,2	512	8,06	36,32
557	6264	370	5,91	38,15	556	6348	378	5,95	36,47
555	6264	549	8,76	38,46	557	6480	639	9,85	38,27
556	6264	800	12,77	39,15	554	6660	265	3,97	39,01
552	6480	404	6,23	40,44	549	6720	236	3,51	39,17

Table 9 demonstrates that the Overall Equipment Effectiveness (OEE) values and unplanned downtime have an inverse relationship. By analyzing both the OEE and RAM (Reliability, Availability, and Maintainability) metrics together, the company can increase the OEE percentage. In order to improve the OEE values of the enterprise, it is necessary to carry out preventive maintenance and repair work on the trucks. It is also essential to minimize repair losses by conducting a maintainability analysis for each truck and following the types of failures and intervention times associated with those failures.

3.2. Overall Equipment Effectiveness Values of Shovel Working in Manisa-Soma Coal Mine



Fig. 2. Shovel.

Five shovels (Fig. 2.) actively used in the enterprise, which are quite old, were evaluated for their overall equipment effectiveness values using data from 2021 and 2022. Tables 10 and 11 provide statistical data on shovel usage in the business for the years 2021 and 2022.

The truck-shovel matching point value is much lower than 1, causing longer truck waiting times for shovels due to insufficient trucks.

Table 10. Statistical Data for Shovel-2021.

<i>Model</i>	<i>Shovel Number</i>	<i>Program (Hour)</i>	<i>Actual amount of work (Ton)</i>	<i>Failure total (Hour)</i>	<i>Loss of working conditions (weather opposition, etc.) (Hour)</i>
MARION/191M11	2	5910,0	1.776.600	793,50	980,50
MARION/191M-II	4	5407,5	954.800	2478,00	847,00
MARION/191M-II	5	5797,5	1.598.100	1008,50	1063,00
MARION/191M-	9	5527,5	544.600	767,50	740,50
MARION/191M11	16	5512,5	865.900	1563,50	719,00

Table 11. Statistical Data for Shovel-2022.

<i>Model</i>	<i>Shovel Number</i>	<i>Program (Hour)</i>	<i>Actual amount of work (Ton)</i>	<i>Failure total (Hour)</i>	<i>Loss of working conditions (weather opposition, etc.) (Hour)</i>
MARION/191M11	2	5400,00	532.000	2580,50	312,50
MARION/191M-II	4	6150,00	1.995.000	965,00	701,00
MARION/191M-II	5	6030,00	1.619.800	631,50	744,00
MARION/191M	9	5362,50	203.700	1604,50	609,50
MARION/191M11	16	5452,50	456.400	2365,50	378,50

Table 12 displays the time lengths for Shovel No. 2. In real-time OEE calculations, the total times determined by the field authorities under actual working conditions were taken into account. The amount of work that the shovel was expected to do according to all these working conditions was compared to the amount of work it actually did. Real data was used in the calculation of the amount of work done, which included the shovel's capacity of 19 Yd3 (14.50 m3) and 1.5 buckets per minute.

Table 12. Time Lengths of the Elements of Shovel No. 2 -2021.

Item	Description	Time (Hour)
Total Duration	24 hours/day*30 days/month*12 months/year	8640
<i>Scheduled Time (1): (Administrative Leave)</i>	2 days/month*24 hours/day*12 months/year	576
<i>Scheduled Time (2): (Meal and refreshment break)</i>	1 hour/shift*3 Shifts/day*30 days/month* 12 Months/year	1080
<i>Scheduled Time (3): Time not intended for operation</i>	37.5 days/year*24 hours/day	900
Total scheduled time=1+2+3		2556
Planned Maintenance	1 day/month* 24 hours/day* 12 months	288
Unplanned Failure Stops		794
Installation and Setup	0.5 hours/shift* 3 shifts/day* 24 hours/day* 12 Months	540
Idle Time	0.6 hours/shift* 3 shifts/day* 24 hours/day* 12 Months	648
Truck Waiting Time	0.1 hour/shift* 3 shifts/day* 24 hours/day* 12 Months	108
Loss of working Conditions	Losses due to weather, etc.	981
Speed Loss	0.4 hours/shift* 3 shifts/day* 24 hours/day* 12 Months	864
Moving Time	4 transports/month * 2hours/transport*12 Months	96
Quality Loss	Filling Factor (87%)*Loader Operator Factor (95%) = 82.65	

The Availability, Performance, Quality and OEE values of Shovel-2 are calculated as given in Table 13.

$$Net\ available\ hours = 8640 - (2556 + 288) = 5796$$

$$Downtime\ Loses = 794 + 540 + 648 + 108 = 2090$$

$$Operation\ Times = 5796 - (981 + 864 + 96) = 3855$$

$$Speed\ Losses = 981 + 864 + 96 = 1941$$

$$Availability = (5796 - 2090) / 5796 = 0,6394$$

$$Performance = (3855 - 1941) / 3855 = 0,4964$$

$$Quality = 0,87 * 0,95 = 0,8265$$

Table 13. OEE Calculations of Shovel No:2-2021.

<i>Calculation based on Application Time</i>	
<i>Net available hours</i>	5796,00

Availability	0,6394
Performance	0,4964
Quality	0,8265
OEE	0,2623

Table 14. Calculation based on Application Time-2021.

Shovel Number	Availability	Performance	Quality	OEE %	Actual Work Quantity (Ton)	Production Quantity (Tons) (OEE at 50% level)
2	63,95	49,64	82,65	26,23	1.776.600	3.384.000
4	31,33	51,02	82,65	13,21	954.800	3.613.929
5	59,31	44,44	82,65	21,78	1.598.100	3.668.733
9	61,79	54,03	82,65	27,59	544.600	986.952
16	47,05	54,88	82,65	21,34	865.900	2.028.819
Total					5.740.000	13.682.433

Table 15. Calculation based on Application Time-2022.

Shovel Number	Availability	Performance	Quality	OEE %	Actual Work Quantity (Ton)	Production Quantity (Tons) (OEE at 50% level)
2	26,58	68,25	82,65	14,99	532.000	1.774.516
4	62,54	62,03	82,65	32,07	1.995.000	3.110.384
5	67,42	59,54	82,65	33,18	1.619.800	2.440.928
9	44,69	57,29	82,65	21,16	203.700	481.333
16	31,28	66,45	82,65	17,88	456.400	1.276.286
Total					4.806.900	9.083.447

Upon analyzing Tables 14 and 15, it can be observed that shovel number 4 had the lowest OEE value of 13.21% in 2021, whereas shovel number 2 had an OEE value of 14.99% in 2022. It's worth noting that despite having the worst OEE value in 2021, shovel number 4 had the second-highest OEE value in 2022. The reason behind this is that shovel number 4 underwent an engine overhaul in 2021, which resulted in its low OEE value for that year. However, after the overhaul, shovel number 4's OEE value improved significantly, becoming the second-best in 2022. Therefore, it is recommended to conduct preventive-predictive maintenance operations for old-model shovels, especially after an engine overhaul.

Table 16. Relationship between unplanned downtime and OEE.

Shovel Number	2021				Shovel Number	2022			
	Total Duration Time (Hour)	Unplanned Failure Stops (hour)	(Unplanned Failure Stops (hour))/(Net available hours)*100	OEE %		Total Duration (Hour)	Unplanned Failure Stops (hour)	(Unplanned Failure Stops (hour))/(Net available hours)*100	OEE %
4	5496	2478	45,09	13,21	2	5280	2581	48,87	14,99
16	5400	1564	28,95	21,34	16	5328	2366	44,40	17,88
5	5664	1009	17,81	21,78	9	5244	1605	30,60	21,16
2	5796	794	13,69	26,23	4	6036	965	15,99	32,07
9	5400	768	14,21	27,59	5	5916	632	10,67	33,18

Table 16 shows the relationship between Shovel OEE analysis and unplanned downtime. Similar to the truck analysis, there is an inverse proportion between the OEE value and unplanned downtime. To ensure maximum efficiency, the company should evaluate repairability analysis and OEE analysis together and take necessary

measures. Furthermore, it's important to evaluate the OEE of shovels with the truck-shovel matching factor. If the matching factor is below 1, the shovel may be waiting for truck loading due to the lack of trucks or the long distance to the dump site. In this case, the company must dispatch enough trucks or find solutions to reduce the distance to the dump site in order to maintain optimal performance.

$$\text{Match factor} = \frac{\text{Number of truck} * \text{Total period duration of the shovel}}{\text{Number of Shovel} * \text{Truck tour time}} = \frac{3 * 180 \text{ second}}{1 * 19 \text{ minute} * 60 \text{ second}} = 0,47$$

When analyzing the Shovel OEE, it is crucial to consider the truck-shovel compatibility factor within the enterprise. This factor is particularly important when analyzing the work quantities performed by shovels in 2021. It was observed that even though the OEE value was high, the work quantities were low. Upon analyzing the shovel-truck compatibility factor in the field, a compatibility factor of 0.47 was calculated. This was due to the fact that only 3 trucks were available for shovels, and the shovel tour time was 30 seconds while the truck tour time was 19 minutes. Additionally, each truck was filled with 6 shovels. A compatibility factor below 1 indicates that the shovel has to wait for the truck to load. Therefore, establishing a good truck dispatch system on-site and increasing the number of trucks can result in a proportional change in the actual volumes of work to the OEE values. In 2022, there is better alignment between actual throughput and OEE values.

4. Results

- Increasing the OEE values of the trucks will lead to more work performed due to the observed parallelism between the two factors.
- Due to the high total working time of trucks, the frequency of breakdowns increases. To prevent this, it is important to perform a reliability analysis of the trucks and apply preventive and predictive maintenance systems.
- When calculating OEE, shovel operator productivity is crucial for quality. Therefore, increasing loader personnel productivity increases OEE.
- The actual amount of work is not compatible with the OEE values of Shovels due to several reasons.
 1. The shovels have been used for a long time, resulting in a high total working time, which, in turn, increases their failure frequency. Therefore, it is important to perform a maintainability analysis along with the OEE (Overall Equipment Effectiveness) analysis. This helps to review the types and frequencies of failures, as well as the time taken for repairs.
 2. The shovel-truck compatibility factor is an important metric to consider. If the factor is less than 1, it means that the shovel is waiting for the truck to arrive. In such a scenario, the company should either increase the number of trucks assigned to the shovel or reduce the distance to the dump site to decrease the truck tour time. This will help the operations to run smoothly and efficiently.
- There is an inverse relationship between truck and shovel downtime and OEE values. Therefore, minimizing the downtime will increase the OEE value.
- Truck failure times must be reduced if an OEE of at least 50% is to be attained. Reducing the distance to the waste area will likewise raise the OEE value concurrently. The ideal shovel-truck matching for shovels is given by a matching factor of 1, and maintaining this value will raise the OEE. When arranging shovel-truck arrangements, it is crucial to have an adequate number of trucks scheduled for this reason. To achieve this, it is recommended to buy new trucks, minimize the period when trucks and shovels break down, and shorten the duration of truck tours. Production will rise if the OEE value is raised to 50%.

Acknowledgements

This study did not receive any specific funding or financial assistance from governmental, commercial, or non-profit organizations.

References

- [1] S. Nakajima, "Introduction to Total Productive Maintenance", *Productivity Press*, Cambridge, MA, 1988
- [2] R. K. Sharma, D. Kumar & P. Kumar, "Manufacturing excellence through TPM implementation: a practical analysis", *Industrial Management & Data systems*, 106, 2006, pp. 256-280
- [3] Ki-Young Jeong, Don T. Phillips, "Operational efficiency and effectiveness measurement", *International Journal of Production and Management*, 22-11, 2001, pp.1404-1416
- [4] M. Braglia, M. Frosolini & F. Zammori, "Overall equipment effectiveness of a manufacturing line (OEEML)", *Journal of Manufacturing Technology Management*, 20-1, 2009, pp. 8-29
- [5] R. Oechsner, M. Pfeffer, L Pfitzner, H. Binder, E. Muller, "From overall equipment efficiency (OEE) to overall Fab effectiveness (OFE)", *Materials Science in Semiconductor Processing*, 5, 2003, 333–339
- [6] J. A. Garza-Reyes, S. Eldridge, K. D. Barber & H. Soriano-Meier, "Overall equipment effectiveness (OEE) and process capability (PC) measures", *International Journal of Quality & Reliability Management*, 27-1, 2010, pp. 48-62
- [7] S. Elevli And B. Elevli, "Performance measurement of mining equipments by utilizing OEE," *Acta Montanistica Slovaca*, vol.15, no.2, 2010, pp.95-101,
- [8] Niraj Ranjan Sharma, Arvind Kumar Mishra & Sandeep Jain "OEE improvement of mining shovels by survival analysis and linear optimization as per sustainable development goals", *International Journal of Mining, Reclamation, and Environment*, 36:5, 2022, 323-355, DOI: 10.1080/17480930.2022.2044138
- [9] J. M. Akande, A. I. Lawal, A. E. Aladejare, "Optimization of the Overall Equipment Efficiency (OEE) of Loaders and Rigid Frame Trucks in NAMDEB Southern Coastal Mine Stripping Fleet, Namibia." *Earth Science*. Vol. 2, No. 6, 2013, pp. 158-166. doi: 10.11648/j.earth.20130206.17
- [10] M. Mohammadi, Mousa, P. Rai, S. Gupta, Performance Evaluation of Bucket Based Excavating, Loading and Transport (BELT) Equipment – an OEE Approach." *Archives of Mining Sciences*. 62., 2017, 105-120. 10.1515/amsc-2017-0008.
- [11] R. P. Choudhary, "Optimization of Load–Haul–Dump Mining System by OEE and Match Factor for Surface Mining", *International Journal of Applied Engineering and Technology* ISSN: 2277-212X (Online) Vol. 5 (2) April-June 2015, pp. 96-102/Choudhary
- [12] M. J. Alam, B. K. Mahanta, N. Nawghade, "Comparative performance study of mine trucks by Overall Equipment Effectiveness (OEE)." *International Research Journal of Engineering and Technology (IRJET)*, 5(11), 2018, 448-453.



E-ISSN: 2687-6167

Number 56, March 2024

RESEARCH ARTICLE

Receive Date: 19.02.2024

Accepted Date: 18.03.2024

Data correlation matrix-based spam URL detection using machine learning algorithms

Funda Akar^{a,*}

^aDepartment of Computer Engineering, Erzincan Binali Yıldırım University, Erzincan 24002, Turkey,
ORCID: 0000-0001-9376-8710

Abstract

In recent years, the widespread availability of internet access has brought both advantages and disadvantages. Users now enjoy numerous benefits, including unlimited access to vast amounts of information and seamless communication with others. However, this accessibility also exposes users to various threats, including malicious software and deceptive practices, leading to victimization of many individuals. Common issues encountered include spam emails, fake websites, and phishing attempts. Given the essential nature of internet usage in contemporary society, the development of systems to protect users from such malicious activities has become imperative. Accordingly, this study utilized eight prominent machine learning algorithms to identify spam URLs using a large dataset. Since the dataset only contained URL information and spam classification, additional feature extractions such as URL length and the number of digits were necessary. The inclusion of such features enhances decision-making processes within the framework of machine learning, resulting in more efficient detection. As the effectiveness of feature extraction significantly impacts the results of the methods, the study initially conducted feature extraction and trained models based on the weight of features. This paper proposes a data correlated matrix approach for spam URL detection using machine learning algorithms. The distinctive aspect of this study lies in the feature extraction process applied to the dataset, aimed at discerning the most impactful features, and subsequently training models while considering the weighting of these features. The entire dataset was used without any reduction in data. Experimental findings indicate that tree-based machine learning algorithms yield superior results. Among all applied methods, the Random Forest approach achieved the highest success rate, with a detection rate of 96.33% for the non-spam class. Additionally, a combined and weighted calculation method yielded an accuracy of 94.16% for both spam and non-spam data.

© 2023 DPU All rights reserved.

Keywords: Classification; Machine Learning; Spam Detection; Tree-based Algorithms.

1. Introduction

The proliferation of the internet since the early 2000s has led to its widespread adoption among nearly all individuals in contemporary society. Through internet usage, users have access to vast knowledge and opportunities, with many relying on digital platforms for their daily activities, thus contributing to the increased utilization of the internet. However, this widespread adoption has also exposed users to various risks and harms, particularly through unconscious internet usage and the prevalence of fraudulent schemes. Many people have fallen victim to things like social engineering, fake websites, and email phishing. To reduce these risks, it is crucial to take preventive measures against threats such as spam emails and malicious URLs. Additionally, the increasing dependence on online transactions and the prevalence of technology, especially IoT technologies, in daily life increasingly raises issues related to web security and data protection. The use of malicious URL links causes a serious risk as it is one of the most common types of cyber-attacks and can allow unauthorized access to personal data [1].

Hence, this study was guided to define and solve the spam URL problem by using machine learning algorithms to analyze and classify internet content. In order to increase the efficacy of spam URL detection, various methods have been investigated to obtain the best results using different machine learning algorithms. The most important difference that distinguishes this study from others is the extraction of features from the data set and the training of models by taking into account the weight levels of these features.

Various studies have been conducted employing diverse methodologies targeting spam URLs, spam emails, and spam bots [2]–[8]. Numerous approaches have been employed for URL classification, culminating in the development of decision support systems and relevant applications. It is imperative to employ rigorous filtering mechanisms for URL classification, encompassing various web page advertisements and activities deemed hazardous [9]. Chen et al. utilize extensive URL databases or blacklists to detect malicious or phishing websites [10]. Frequently, attacks generate spam URLs resembling legitimate corporate websites through tactics such as social engineering. Malware is designed to disseminate through URLs, infecting computers and propagating the authors' software [11]. Frequently, attacks generate spam URLs resembling legitimate corporate websites through tactics such as social engineering. Malware is designed to disseminate through URLs, infecting computers and propagating the authors' software.

A relevant study analyzed the differences between benign and phishing URLs utilizing attributes such as URL and domain length. Through dissecting the structure of phishing URLs, domain registration, and the hosting machinery, the authors illustrated that domain names employed for phishing endeavors exhibit disparate lengths and locations [12]. Building upon domain behavior on phishing platforms, Ma et al. devised a model capable of identifying suspicious URLs with 99% accuracy, leveraging verbal and blacklisted host-based features [13].

Kwon et al. proposed a different method for detecting spam URLs. Authors focused to domain independence, competitive robustness, and semi-supervised perception. They applied machine learning techniques in their investigation, which produced 96% recall and 70% accuracy results and obtained these results using Decision Tree, Logistic Regression and Support Vector Machine [14]. Takata et al. attempted to identify a direct download attack type by analyzing it as having the relevant code in the redirection [15]. Research was carried out by Almeida and Westphall to identify harmful URLs. The URLs constructed with identity theft in mind were found in the study. Generally, they stated that they were detected with success rates between 73-97% on URLs created for phishing purposes [16]. Manyumwa et al. made multi-class classification on malicious URL detection. DMOZ, PhishTank, URLhaus, WEBSHAM datasets and XGBoost, Adaboost, LightGBM, CatBoost were used as machine learning methods in their study [11]. Rao and Pais proposed a different method for the detection of URLs developed for identity theft and tried to find the best result with many machine learning algorithms. In their study, they found the highest result with 99.31% accuracy [17]. Raj and Kang used many machine learning methods for spam URL detection in their study and the highest test accuracies were made with XGBoost and 87% success was found [18]. Another proposed system offers a dual-layered detection mechanism. Initially, URLs are categorized as either benign or malicious using a binary classifier. Subsequently, URL classes are further classified into five categories based on their features: benign, spam, phishing, malware, and defacement. In particular, we present findings on four ensemble learning methodologies, namely the ensemble of bagging trees (En_Bag) approach, the ensemble of k-nearest neighbor (En_kNN) approach, the ensemble of boosted decision trees (En_Bos) approach, and the ensemble of subspace discriminator (En_Dsc) approach. They also compare their En_Bag model with state-of-the-art solutions, demonstrating its superiority in both binary classification and multi-classification tasks, achieving accuracy rates of 99.3% and 97.92%, respectively [19].

Many studies have been done on URL spam detection before. This paper proposes a data correlated matrix approach for spam URL detection using machine learning algorithms. For this purpose, effective operations were carried out on the dataset to increase the success. First, a dataset containing large-scale data was provided in order to carry out the study. The dataset contains about 150,000 spam or non-spam URLs. Using feature engineering, the best way to approach the decision-making process was established. The features were systematically eliminated, considering their potential significance in spam detection within URLs. A correlation matrix was employed to identify the most influential features for learning. Utilizing this insight, an extensive classification process was undertaken with numerous machine learning algorithms to determine the optimal outcome. Finally, the findings were thoroughly discussed, providing insights into the efficacy of features in identifying spam URLs, thus culminating the study.

In this study, no new dataset was generated. Instead, models were trained by assigning weight values solely based on the significance of features, without any reduction or modification to the dataset; the entire dataset was utilized in its original form.

2. Material and methods

In order to detect spam URLs, first of all, it is necessary to know what the URL is and what parts it consists of (Fig. 1). URLs usually consist of domain and subdomain. Rules defined by a protocol are used to transfer data to the other party. URLs contain many numbers and letters and there can be redirects within the website in a structure similar to a folder within path.

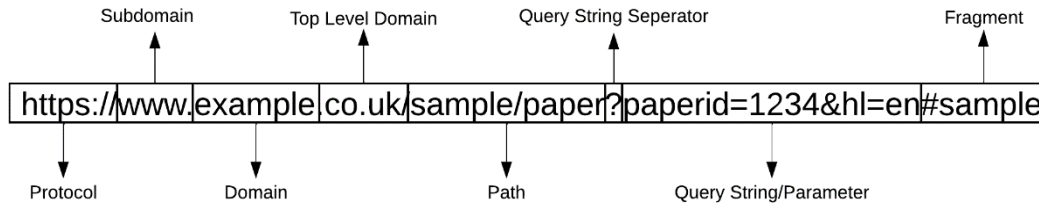


Fig 1. An example showing URL parts.

Data set from Kaggle was used in the study which contains a very large amount of data [20]. In this way, it will be better to test the transactions made. It is prepared to be given to machine learning algorithms in the most appropriate way by feature engineering with the acquired data. Data are based on general study opinion. Based on a general justification utilized in the literature search, these discrimination rates are split into 70% train and 30% test groups. Later, machine learning methods were made ready for implementation. The application is terminated by making spam URL classification. The flow chart of the processes performed in the study is shown in Fig. 2.

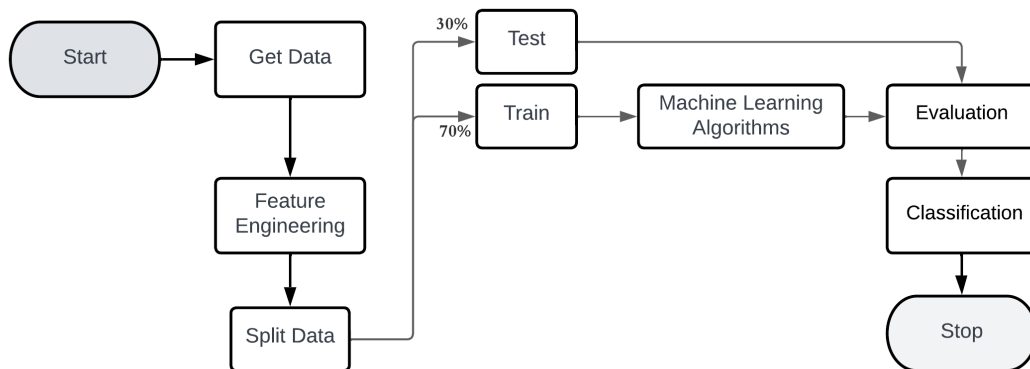


Fig. 2. Flowchart of the work.

2.1. Structure

The dataset, which has already labeled data, contains approximately 150.000 URLs [20]. Fig. 3 shows the spam and non-spam distribution of the URLs. Only URLs and their tags are included in the dataset. To determine whether it is spam, some pre-processing is required for this reason. 70% of the dataset is used for training and 30% for testing.

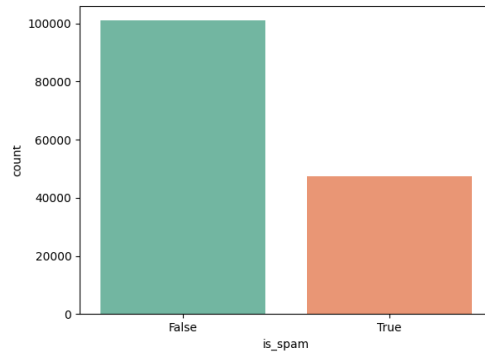


Fig. 3. Spam distribution of the dataset.

There are nearly 100.000 non-spam and 50.000 spam URLs in the dataset. This will result in varying weights for the values provided. The study utilized labeled values and did not use any missing data. Furthermore, the impact of weights on the outcomes remained unchanged. The outcomes of machine learning were included into the outcomes in the manner in which they were created.

2.2. Feature engineering

Since there is no information other than URL and label in the dataset, an extra feature must be added. Therefore, it is necessary to add data over numeric, special characters and letters in the dataset. Better results are expected after this procedure. Satisfactory results were found in the study, especially thanks to feature engineering part. Since the number of letters is very important in the transactions, first the number of letters and the length of the URL have been added. Special characters are also added one by one because there are many pieces. Since numerical data are also seen to be important, they are also indicated in the Table I.

Table 1. All features and descriptions in the dataset.

Features	Feature Description
length_url	Specifies the length of the URL.
num_digits	Shows the total number in the URL
num_letters	Total number of letters in URL
num_words	Total number of words in URL
with_https	Number of "https" in the URL
num_hashtag	Number of "#" in the URL
num_?	Number of "?" in the URL
num_/_	Number of "/" in the URL
num_!	Number of "!" in the URL
num_-	Number of "-" in the URL
num_.	Number of "." in the URL
num_*	Number of "*" in the URL
num_	Number of "_" in the URL
num_%	Number of "%" in the URL
num_&	Number of "&" in the URL
inc_www	Whether there is "www" in the URL or not
inc_subscribe	Whether there is "subscribe" in the URL or not
inc_com	Whether there is "com" in the URL or not
inc_net	Whether there is "net" in the URL or not
inc_edu	Whether there is "edu" in the URL or not
inc_org	Whether there is "org" in the URL or not

Since machine learning algorithms were used many times while determining the features, all kinds of examinations were made. Every possible feature has been tried to be added until satisfactory results are obtained. It has been observed that the procedures performed directly affect the results. In this way, it has been tried to underline these important features while performing the classification processes. The degree of effect of the results was determined using the correlation matrix. The correlation matrix depicts the intra-grid structure of variables based on the correlation effect coefficient, which was a value between [-1, 1]; whichever number was closer to 1 resulted in more correlations between data components [21]. In this way, the effect of the added features can be seen clearly. Although some added features remained ineffective, they were still not removed. Because it has been seen that it has a small effect on performance. That's why every added feature is given to machine learning algorithms.

In Fig. 4 all features are given according to the feature importance and correlation matrix is given in Fig. 5. The length of the URL, the number of letters, the number of digits and the number of "-" are the four features having the most impact, as shown in figures.

Except for feature weighting, no data manipulation has occurred. No operations, such as addition or removal of data, have been conducted that could impact the outcomes. Any modifications to the dataset would inevitably influence the results. Machine learning methods were applied while strictly adhering to the original dataset. The considerable size disparity between non-spam and spam data underscores the significance of accurately detecting non-spam instances, given its twice as large representation, which substantially affects the success rates. Additionally, the augmentation of features has introduced supplementary inputs, crucial for providing a more comprehensive benchmark for machine learning evaluation. The procedures were designed with this consideration in mind, aiming to compare the effects of features on the results, thereby discerning the efficacy of the added features for achieving superior outcomes.

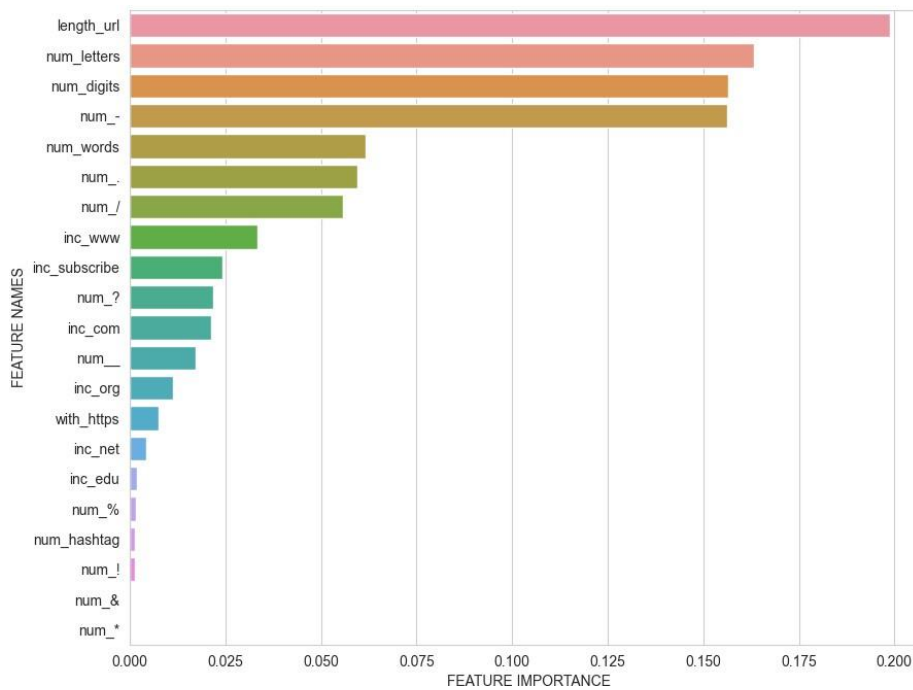


Fig. 4. Feature importance of dataset.

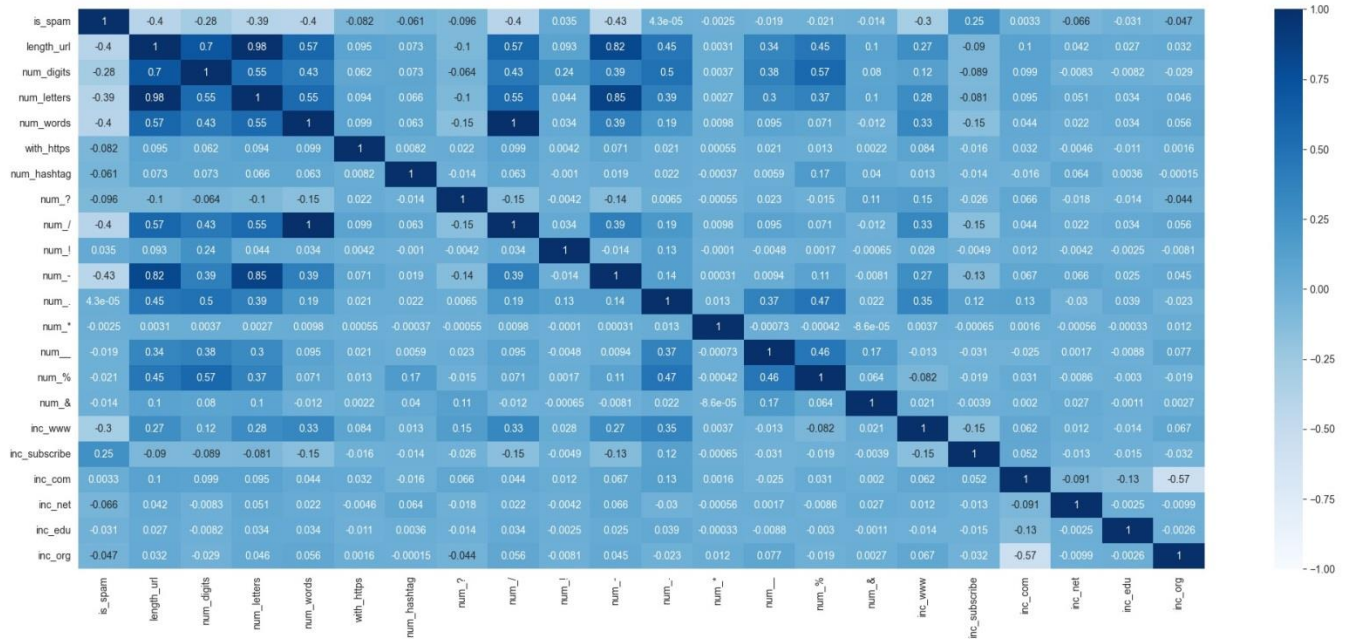


Fig. 5. Correlation matrix.

Correlation matrix was created when only features were detected, and additions were made. This clarifies which feature can match which feature. In addition to "https" and "http" has also been added in the study. However, "http" was excluded because it was not correlated with any value. If it is not removed, it will cause inaccuracies in the results. It is possible to make such a comparison for each added feature. It appears to be comparable to how significant features are in the correlation matrix. It appears that the correlational structure, or likeness to one another, increases with a higher dark blue degree. The conclusion to be made here is that this structure determines the quality of the added features. So, it is clearly seen which additional features will affect the success. Furthermore, features such as exclamation points and hashtags included in the URL have little effect. Nevertheless, the features in this structure were not removed to avoid difficulties in decision making. Because it was seen that when it was eliminated, the success rate of the experimental findings reduced. Therefore, it is important to add features by considering their advantages over each other.

2.3. Machine learning algorithms

The scope of machine learning algorithms has been expanded to allow for a more thorough evaluation in this study. While evaluating machine learning in the study, both class-based achievements and weighted average outcomes were provided. Machine learning algorithms were completed with default values. If the algorithms are executed using parameters other than the default ones, better results are likely to be attained. The key purpose here is to determine which machine learning approach is the most likely to solve this problem. Because several machine learning algorithms are employed in the study, a comprehensive explanation of each approach is not provided here. The study's major goal is not to explain the methods, but to compare the success of the methods with each other. Relevant machine learning approaches have been investigated, and detailed descriptions of the methods may be found in the publications listed below. Machine learning algorithms used in the study:

- Logistic Regression (LR) [22]–[24]: Commonly used for classification problems, especially effective in binary classification tasks.
- Decision Tree Classifier (DTC) [25]–[27]: Widely applicable in various domains, used for both classification and regression tasks. Decision trees find applications in data mining, medicine, finance, and marketing.
- Random Forest Classifier (RFC) [28]–[30]: Used in a variety of application areas. Particularly effective for classification and regression problems with large datasets.
- Naive Bayes (NB) [31]–[33]: Especially popular in text classification tasks such as spam filtering and sentiment analysis. Also applicable to multi-class classification problems.

- K- Nearest Neighbor (KNN) [34], [35]: Used for classification and regression tasks based on spatial similarities. Applications include medical diagnosis in healthcare, customer segmentation in marketing, and more.
- XGBoost (XGB) [36], [37]: Widely applicable and particularly preferred for classification and regression tasks with high-dimensional datasets. Applications span across finance, medicine, natural language processing (NLP), and more.
- AdaBoost Classifier (ABC) [38]–[40]: Constructs a strong classifier by combining weak learners. Often used in areas like face recognition and speech classification.
- Gradient Boosting Classifier (GBC) [41], [42]: Based on gradient boosting principles and commonly used for classification and regression problems. Has a broad range of applications, including web page ranking and medical diagnosis.

Each algorithm has its strengths and weaknesses, so choosing the most suitable one depends on the context of the problem and characteristics of the dataset. Since many articles on how the machine learning methods employed work provide formulas and usage reasoning, only the scanned publications are shown in this study by citing the investigated materials. Because all of these strategies are included in the study and a multi-class dataset is used, both class-based and weighted overall performance results are provided.

2.4. Evaluation metrics

We require certain performance indicators to assess the efficacy of the machine learning techniques used in the study. It is possible to compare which approach performs better than which method in this way. The technique to be used may not always be the most accurate. There are several problem and outcome-oriented algorithms available. The following metrics were utilized in the study:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

Accuracy (1) is defined as the quotient of correctly predicted results divided by total results, signifying overall performance. The F1-score represents the harmonic mean of precision and recall levels (2). Precision (3) represents the proportion of true positive results to other positive outcomes. The recall (4) value is defined as the ratio of true positive values to true negative and true positive values [41]. A Confusion Matrix must be created in order to calculate all of these parameters.

3. Results and discussion

As previously indicated, eight distinct machine learning methodologies were employed in the study. These methods yield a two-class result due to two labels in the dataset as spam or not spam. The outcomes are presented alongside the confusion matrix and ROC curves for comprehensive evaluation. The ROC curve can give us information about how accurate the operation is and how far we are from a successful method. Accuracy, F1-Score, Recall and Precision values are also given as in Section 2.4 for a better evaluation of the study. This will make it easier to compare the study to others and allow for a better evaluation of its effectiveness.

The results of Logistic Regression are presented in Fig. 6 and Table 2. It seems clear that non-spam data is better detected. According to the results obtained, it seems that Logistic Regression is not suitable for solving this problem. Although the ROC curve is not even close to 1, this machine learning method can be useful considering its speed.

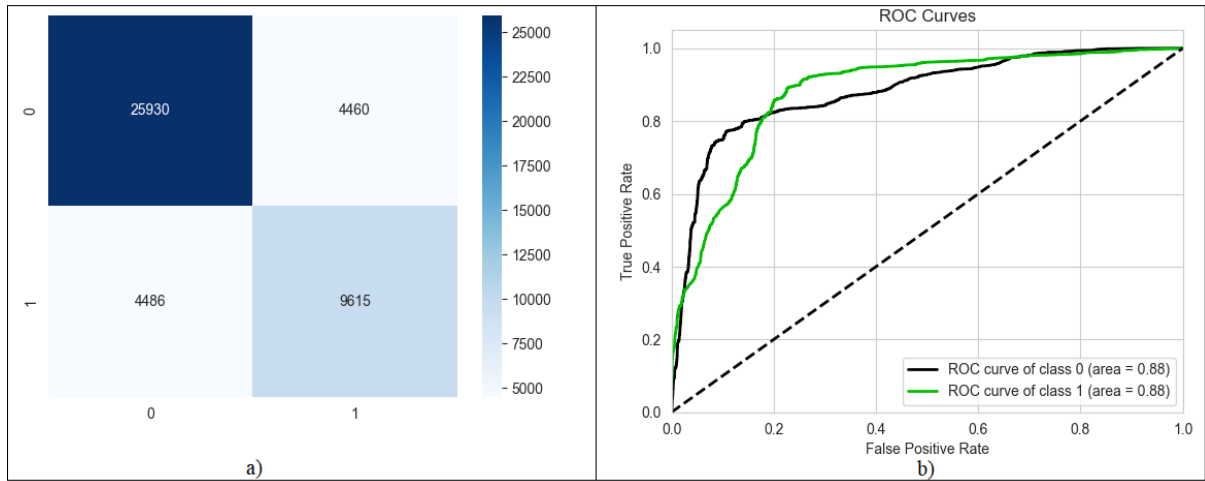


Fig. 6. (a) Confusion matrix; (b) ROC curve values of logistic regression.

Table 2. Evaluation metrics values of logistic regression.

Classes	Precision (%)	Recall (%)	F1-Score (%)	Support
Not spam (class 0)	85.25	85.32	85.29	30390
Spam (class 1)	68.31	68.19	68.25	14101

The results of the Decision Tree are presented in Fig. 7 and Table 3. It appears that non-spam data is more effectively detected. It is seen that Decision Tree generates a good result among the results obtained. ROC curve is very close to 1. As can be seen, the operations performed are very close to the truth. This method is very convenient to use because it is fast.

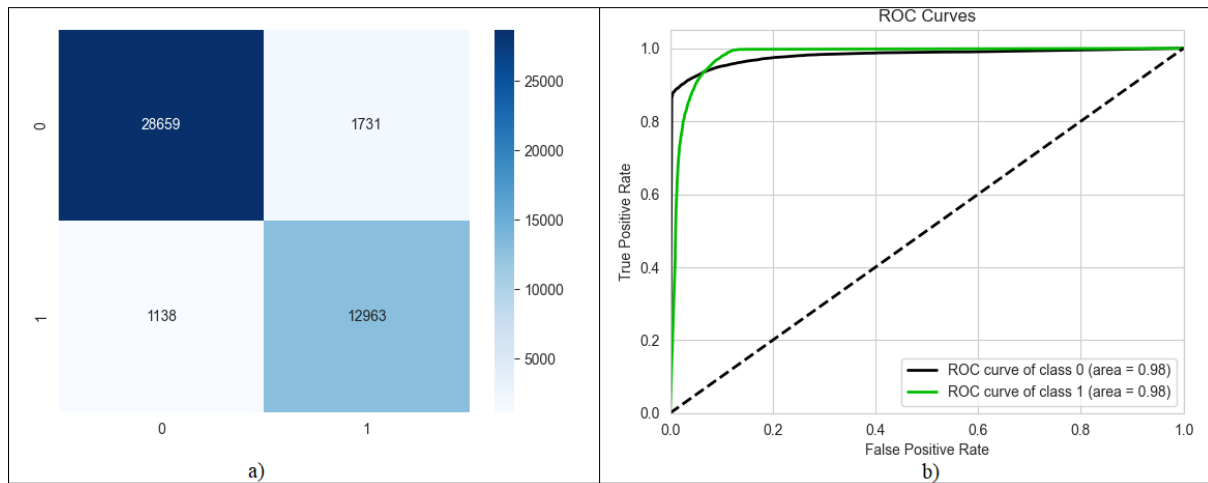


Fig 7. (a) Confusion matrix; (b) ROC curve values of decision tree.

Table 3. Evaluation metrics values of decision tree.

Classes	Precision (%)	Recall (%)	F1-Score (%)	Support
Not spam (class 0)	96.18	94.30	95.23	30390
Spam (class 1)	88.22	91.93	90.04	14101

Fig. 8 and Table 4 show the Random Forest's outcomes. Non-spam data appears to be better detected. Random Forest is the algorithm that produces the best outcomes out of all the results obtained. Because it is relatively slow, there may be a temporal contraction in larger data entries. As can be seen, ROC curve is very close to 1, the operations performed are very close to the truth indicating that the procedures were very accurate. This method is highly convenient to use because it is

fast. In addition to its high results, the difference between the two classes was found smallest in this method among all methods. By cross validating the parameters of this method, which is often employed in studies, good results can be obtained. A sufficient number of parameters have been evaluated in the study and the best results have been tried to be obtained. Moreover, it is a frequently preferred algorithm due to the widespread usage of this method and the convenience of transportation.

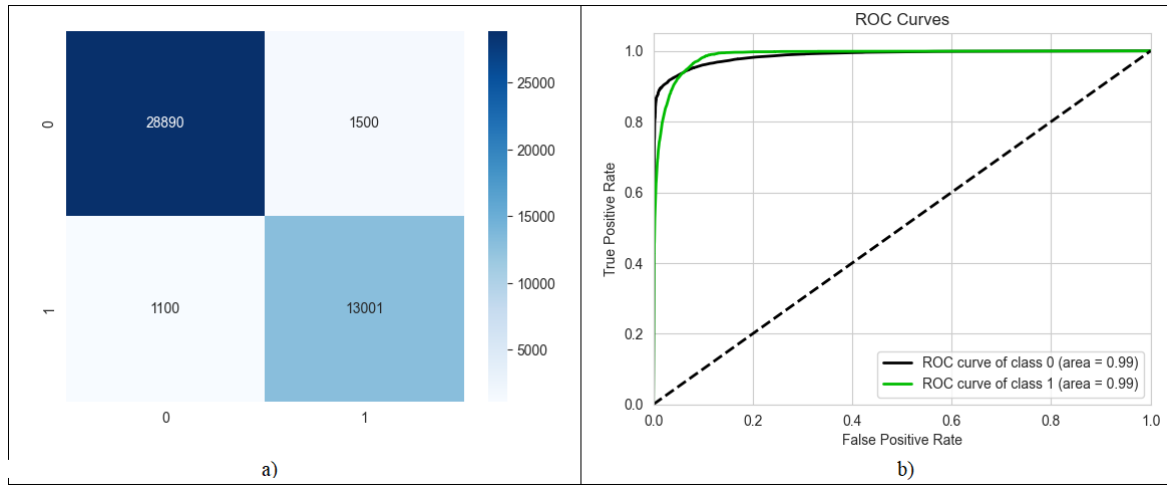


Fig. 8. (a) Confusion matrix; (b) ROC curve values of random forest.

Table 4. Evaluation metrics values of random forest.

Classes	Precision (%)	Recall (%)	F1-Score (%)	Support
Not spam (class 0)	96.33	95.06	95.69	30390
Spam (class 1)	89.66	92.20	90.91	14101

Fig. 9 and Table 5 present the findings of the Naive Bayes model. It is obvious that non-spam data is better detected but the results show that Naive Bayes is not an appropriate solution for this situation. As can be seen, ROC curve is not even close to 1. Nevertheless, this machine learning method can be useful considering that it is fast. Naive Bayes machine learning gets better results on mostly statistical data. Although the values used in this dataset are not statistical, they contain too many variables.

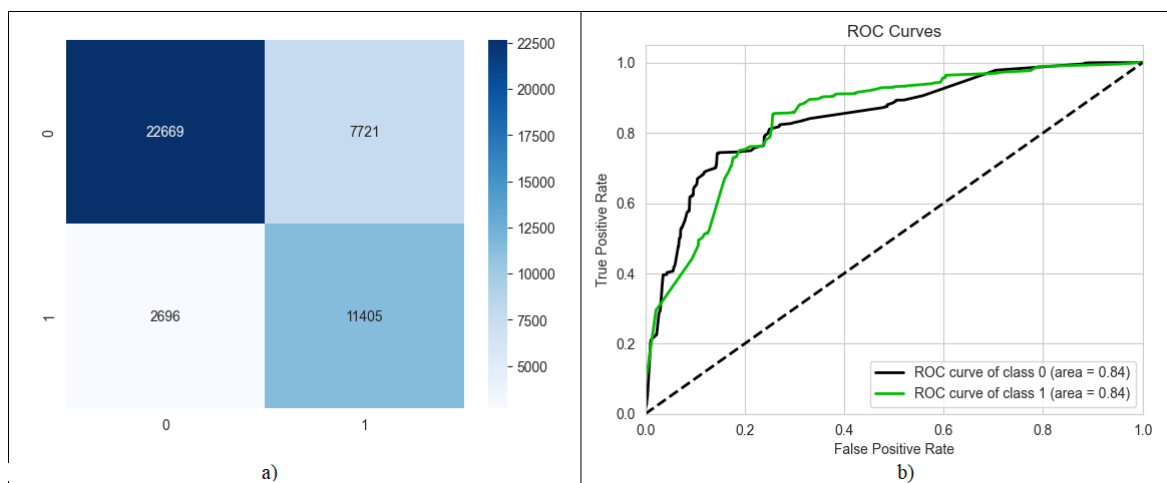


Fig. 9. (a) Confusion matrix; (b) ROC curve values of naïve bayes.

Table 5. Evaluation metrics values of naïve bayes.

Classes	Precision (%)	Recall (%)	F1-Score (%)	Support
Not spam (class 0)	89.37	74.59	81.32	30390
Spam (class 1)	59.63	80.88	68.65	14101

The K-Nearest Neighbor results are displayed in Fig. 10 and Table 6. Non-spam data is clearly detected better. The findings indicate that K-Nearest Neighbors is on the verge of solving this challenge. ROC curve has a value that close to 1. However, given how rapid this machine learning process is, it may be useful. K-Nearest Neighbor excel at classification via building neighborhood relations. Since several processes were performed by establishing correlation in this dataset it yielded good results.

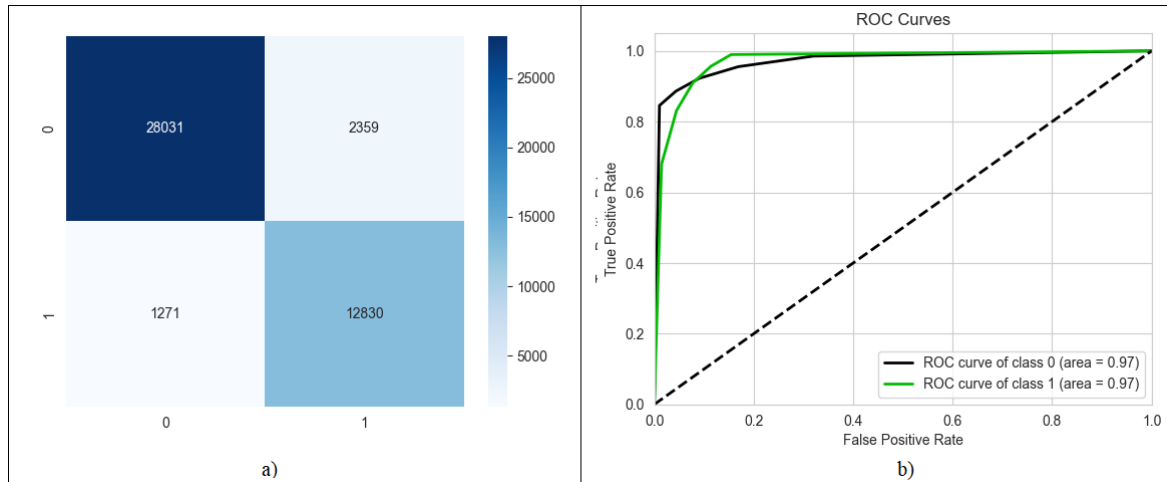


Fig. 10. (a) Confusion matrix; (b) ROC curve values of k-nearest neighbor.

Table 6. Evaluation metrics values of k-nearest neighbor.

Classes	Precision (%)	Recall (%)	F1-Score (%)	Support
Not spam (class 0)	95.66	92.24	93.92	30390
Spam (class 1)	84.47	90.99	87.61	14101

The XGBoost findings are displayed in Fig. 11 and Table 7. It is evident that non-spam data is better identified. According to the results, XGBoost is on the verge of resolving this issue. ROC curve value is close to 1. It might be beneficial, though, considering how fast this machine learning process works. XGBoost is frequently favoured in competitions and might be beneficial, though, considering how fast this machine learning process works. Aside from that, scholars prefer it because of the high worth of the results. It has a graph that is a little far from satisfactory.

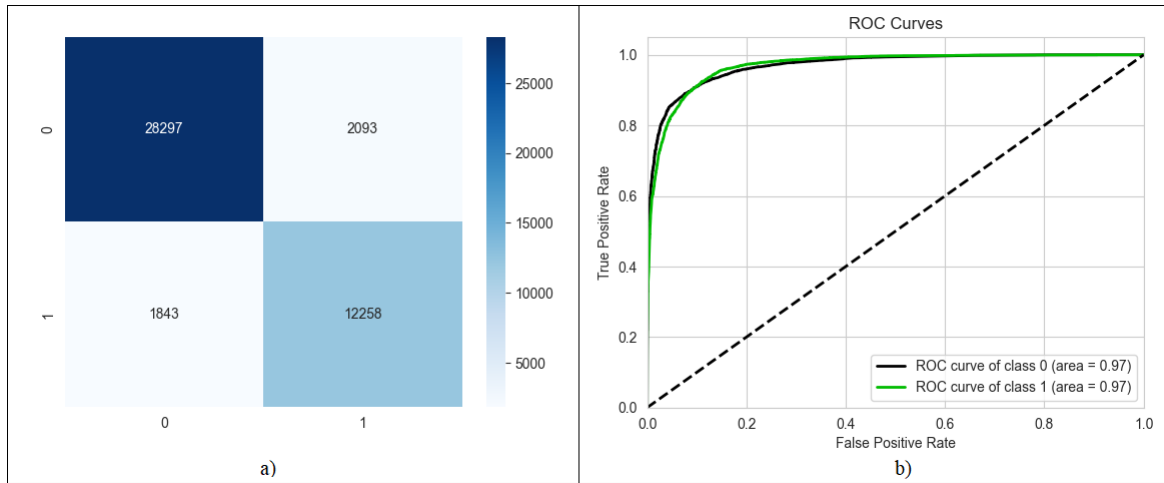


Fig. 11. (a) Confusion matrix; (b) ROC curve values of XGBoost.

Table 7. Evaluation metrics values of XGBoost.

Classes	Precision (%)	Recall (%)	F1-Score (%)	Support
Not spam (class 0)	93.89	93.11	93.50	30390
Spam (class 1)	85.42	86.93	86.17	14101

The results of AdaBoost are shown in Fig. 12 and Table 8. As seen, non-spam data is clearly detected better. The outcomes suggest that AdaBoost is close to resolving this issue. ROC curve value is slightly close to 1. However, given how rapid this machine learning process is, it may be useful. AdaBoost is frequently favored in contests and also popular among academics due to its high-value outcomes. It did not, however, provide very good remedies to this challenge. Although it is commonly used in stepped constructions, it did not produce the expected results when tackling this challenge. It is also thought that it can produce superior outcomes with an expanded parameter network.

Gradient Boosting results are given in Fig. 13 and Table 9. It is clear that non-spam data is better detected. According to the results obtained, it is seen that Gradient Boosting is comes very near to resolving this problem. ROC curve has a value slightly close to 1. However, considering that this machine learning method is fast, it can be useful. AdaBoost is often chosen in contests and also frequently preferred by researchers due to its high-value results. However, it did not provide very good solutions to this problem. Although it is mostly preferred in stepped structures, it could not give the expected results in solving this problem. It is also thought that it can give better results with an expanded parameter network.

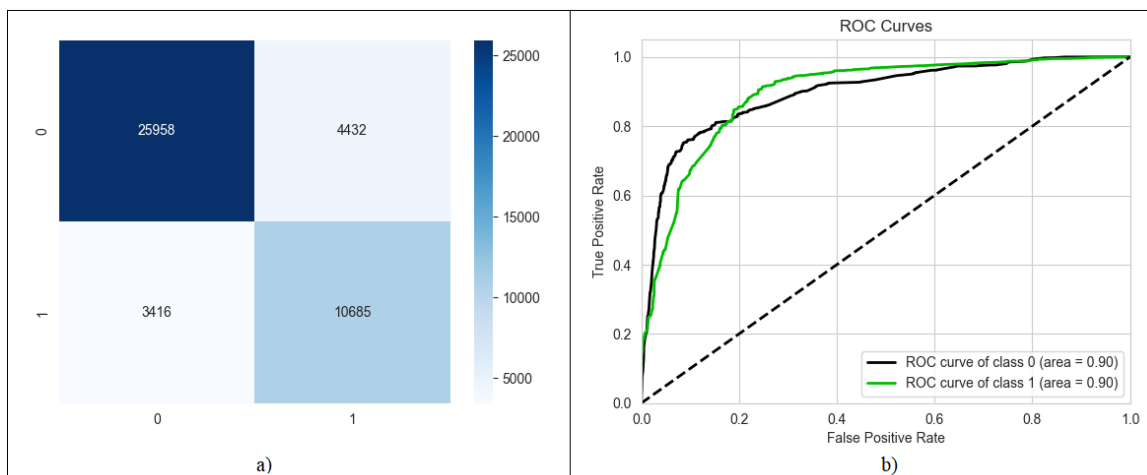


Fig. 12. (a) Confusion matrix; (b) ROC curve values of AdaBoost.

Table 8. Evaluation metrics values of AdaBoost.

Classes	Precision (%)	Recall (%)	F1-Score (%)	Support
Not spam (class 0)	88.37	85.42	86.87	30390
Spam (class 1)	70.68	75.77	73.14	14101

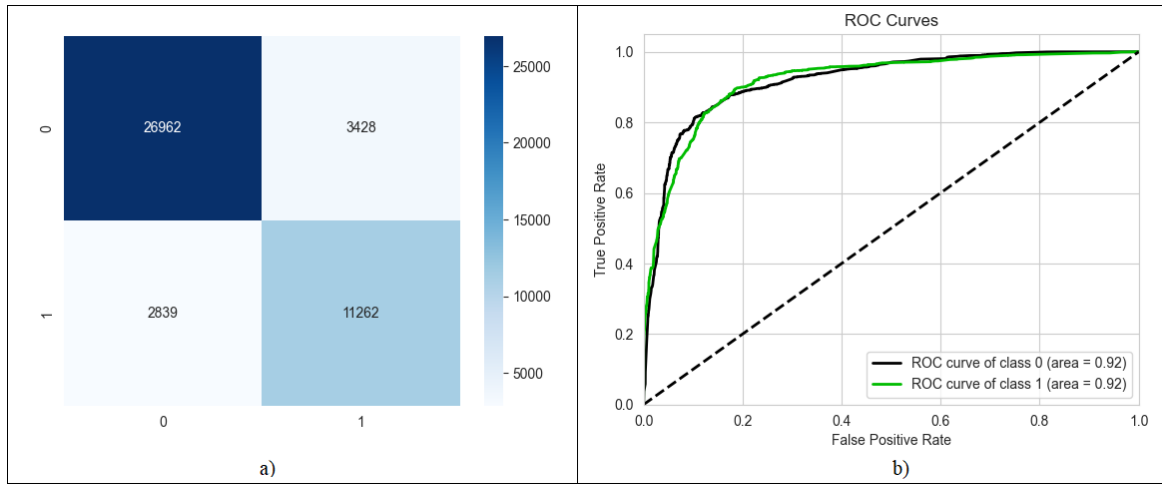


Fig. 13. (a) Confusion matrix; (b)ROC curve values of gradient boosting.

Table 9. Evaluation metrics values of gradient boosting.

Classes	Precision (%)	Recall (%)	F1-Score (%)	Support
Not spam (class 0)	90.47	88.72	89.59	30390
Spam (class 1)	76.66	79.87	78.23	14101

The weighted results of all the mentioned methods are shown in Table 10. Random Forest Classifier produces the best results with a success rate of 94.16% was achieved in all results. The most substantial aspect that distinguishes this study from other studies in the literature is that the most preferred and most successful machine learning methods are preferred and compared. Another prominent property of this study is that it worked with a large data set. Since it contains very high data, the results obtained are at a very satisfactory level. The main purpose of the study is to get impeccable results even with such large data. By using numerous machine learning methods, it has been determined which method can achieve better result. Random Forest method gave the best results in solving this problem, and the Decision Tree method gave the closest result. Additionally, the outcomes will be superior if a tree-based machine learning approach is used to solve such problems.

Table 10. Weighted metrics of the machine learning methods

ML Algorithms	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression (LR)	79.89	79.88	79.89	79.89
Decision Tree (DT)	93.55	93.66	93.55	93.59
Random Forest (RF)	94.16	94.22	94.16	94.18
Naïve Bayes (NB)	76.59	79.95	76.59	77.30
K-Nearest Neighbor (KNN)	91.84	92.11	91.84	91.92
XGBoost (XGB)	91.15	91.20	91.15	91.17
AdaBoost (ABC)	82.36	82.76	82.36	82.52
Gradient Boosting (GBC)	85.91	86.10	85.91	85.99

4. Conclusion

Working with large datasets often yields more consistent results. For this purpose, it is essential to have a large-scale

dataset. Accurate results are often difficult to find on small-scale datasets. Because if the weights of the data in the datasets are not determined very convenient, it will not be possible to get an efficiency from the operations performed. On the contrary, datasets created with very well selected data can also give very effective results with machine learning. Therefore, a large-scale dataset was selected in the study and extra features have been added to the data to be taught in machine learning. In this manner, machine learning would be able to make decisions while considering additional characteristics. The study demonstrates that machine learning algorithms of the tree-based typically produce favourable outcomes. This highlights how the decision mechanism provides appropriate responses for tree architectures. The Random Forest approach found a detection success of 96.33% for the highest non-spam class. Random Forest Classifier produced the best results in the study with 94.16%, 94.22%, 94.16% and 94.18% success was achieved in Accuracy, Recall, Precision and F1-Score values respectively for both spam and non-spam URL detection using combined and weighted results.

When the application is compared with other studies, the most striking difference is that a very large-scale dataset was selected. While working with big data, machine learning algorithms created can prevent high successes that may occur by chance. In this way, it was preferred to choose a large-scale and high-data dataset while selecting the dataset. In addition, most popular eight machine learning methods were used for comparison. As a result of the research such a comprehensive study was not found by the authors. The study provides convenience when determining which approach will be better when many algorithms are used. One of the conclusions that can be drawn from this study is that it has been clearly shown that tree-based classifiers give better results in multi-criteria decision making. There is a potential that the study will provide information for others to use in their future research.

Acknowledgements

There is no conflict of interest with any person or institution in the prepared article.

References

- [1] R. S. Arslan, "Kötüçül Web Sayfalarının Tespitinde Doc2Vec Modeli ve Makine Öğrenmesi Yaklaşımı," *European Journal of Science and Technology*, no. 27, pp. 792–801, 2021, doi: 10.31590/ejosat.981450.
- [2] D. Sahoo, C. Liu, and S. C. H. Hoi, "Malicious URL Detection using Machine Learning: A Survey," *ArXiv*, vol. abs/1701.0, 2017.
- [3] P. Kolar, A. Java, T. Finin, T. Oates, and A. Joshi, "Detecting spam blogs: A machine learning approach," *Proceedings of the National Conference on Artificial Intelligence*, vol. 2, pp. 1351–1356, 2006.
- [4] F. O. Catak, K. Sahinbas, and V. Dörtkarde\cs, "Malicious URL detection using machine learning," *Artificial intelligence paradigms for smart cyber-physical systems*, IGI Global, pp. 160–180, 2021.
- [5] A. Begum and S. Badugu, "A study of malicious url detection using machine learning and heuristic approaches," *Advances in Decision Sciences, Image Processing, Security and Computer Vision*, Springer, pp. 587–597, 2020.
- [6] S. Kumar, X. Gao, I. Welch, and M. Mansoori, "A machine learning based web spam filtering approach," *2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*, 2016, pp. 973–980.
- [7] P. Parekh, K. Parmar, and P. Aware, "Spam URL detection and image spam filtering using machine learning," *Computer Engineering*, 2018.
- [8] M. Aljabri *et al.*, "Detecting Malicious URLs Using Machine Learning Techniques: Review and Research Directions," *IEEE Access*, vol. 10, no. October, pp. 121395–121417, 2022, doi: 10.1109/ACCESS.2022.3222307.
- [9] I. Hernández, C. R. Rivero, D. Ruiz, and R. Corchuelo, "CALA: CLAssifying Links Automatically based on their URL," *Journal of Systems and Software*, vol. 115, pp. 130–143, 2016.
- [10] C.-M. Chen, J.-J. Huang, and Y.-H. Ou, "Efficient suspicious URL filtering based on reputation," *Journal of Information Security and Applications*, vol. 20, pp. 26–36, 2015.
- [11] T. Manyumwa, P. F. Chapita, H. Wu, and S. Ji, "Towards Fighting Cybercrime: Malicious URL Attack Type Detection using Multiclass Classification," *2020 IEEE International Conference on Big Data (Big Data)*, pp. 1813–1822, 2020.
- [12] D. K. McGrath and M. Gupta, "Behind Phishing: An Examination of Phisher Modi Operandi," *LEET*, vol. 8, p. 4, 2008.
- [13] J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Identifying suspicious URLs: an application of large-scale online learning," *Proceedings of the 26th annual international conference on machine learning*, pp. 681–688, 2009.
- [14] H. Kwon, M. B. Baig, and L. Akoglu, "A domain-agnostic approach to spam-url detection via redirects," *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 220–232, 2017.
- [15] Y. Takata, M. Akiyama, T. Yagi, T. Hariu, and S. Goto, "Minespider: Extracting urls from environment-dependent drive-by download attacks," *2015 IEEE 39th Annual Computer Software and Applications Conference*, vol. 2, pp. 444–449, 2015.
- [16] R. Almeida and C. Westphall, "Heuristic phishing detection and URL checking methodology based on scraping and web crawling," *2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pp. 1–6, 2020.
- [17] R. S. Rao and A. R. Pais, "Detection of phishing websites using an efficient feature-based machine learning framework," *Neural Computing and Applications*, vol. 31, no. 8, pp. 3851–3873, 2019.
- [18] R. Raj and S. S. Kang, "Spam and Non-Spam URL Detection using Machine Learning Approach," *2022 3rd International Conference for Emerging Technology (INCET)*, pp. 1–6, 2022.
- [19] Q. Abu Al-Haija and M. Al-Fayoumi, "An intelligent identification and classification system for malicious uniform resource locators (URLs)," *Neural Computing and Applications*, vol. 35, no. 23, pp. 16995–17011, 2023, doi: 10.1007/s00521-023-08592-z.
- [20] Kaggle, "Spam URLs Classification Dataset." <https://www.kaggle.com/datasets/shivamb/spam-url-prediction>.
- [21] A. Hmimou and others, "On the computation of the correlation matrix implied by a recursive path model," *2020 IEEE 6th International Conference on Optimization and Applications (ICOA)*, pp. 1–5, 2020.
- [22] S. Sperandei, "Understanding logistic regression analysis," *Biochimica medica*, vol. 24, no. 1, pp. 12–18, 2014.

- [23] D. Maulud and A. M. Abdulazeez, "A review on linear regression comprehensive in machine learning," *Journal of Applied Science and Technology Trends*, vol. 1, no. 4, pp. 140–147, 2020.
- [24] J. Chen *et al.*, "A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide," *Environment international*, vol. 130, p. 104934, 2019.
- [25] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.
- [26] Y. K. Qawqzeh, M. M. Ootom, and F. Al-Fayez, "A Proposed Decision Tree Classifier for Atherosclerosis Prediction and Classification," *International Journal of Computer Science and Network Security (IJCSNS)*, vol. 19, no. 12, pp. 197–202, 2019.
- [27] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, 2021.
- [28] L. Breiman, "Random forests; uc berkeley tr567," *University of California: Berkeley, CA, USA*, 1999.
- [29] J. R. Quinlan, *C4. 5: programs for machine learning*, Elsevier, 2014.
- [30] L. Breiman, J. Friedman, C. Stone, and R. Olshen, "Classification and regression trees (crc, boca raton, fl)," 1984.
- [31] I. Rish and others, "An empirical study of the naive Bayes classifier," *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, no. 22, pp. 41–46, 2001.
- [32] E. Frank and R. R. Bouckaert, "Naive bayes for text classification with unbalanced classes," *Knowledge Discovery in Databases: PKDD 2006: 10th European Conference on Principles and Practice of Knowledge Discovery in Databases Berlin, Germany, September 18-22, 2006 Proceedings 10*, pp. 503–510, 2006.
- [33] Ö. Şahinaslan, H. Dalyan, and E. Şahinaslan, "Naive bayes sınıflandırıcısı kullanılarak youtube verileri üzerinden çok dilli duygu analizi," *Bilişim Teknolojileri Dergisi*, vol. 15, no. 2, pp. 221–229, 2022.
- [34] Y. Wu, K. Ianakiev, and V. Govindaraju, "Improved k-nearest neighbor classification," *Pattern recognition*, vol. 35, no. 10, pp. 2311–2318, 2002.
- [35] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN Model-Based Approach in Classification," *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, pp. 986–996, 2003.
- [36] T. Chen *et al.*, "Xgboost: extreme gradient boosting," *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.
- [37] A. Asselman, M. Khaldi, and S. Aammou, "Enhancing the prediction of student performance based on the machine learning XGBoost algorithm," *Interactive Learning Environments*, pp. 1–20, 2021.
- [38] T.-K. An and M.-H. Kim, "A new diverse AdaBoost classifier," *2010 International conference on artificial intelligence and computational intelligence*, vol. 1, pp. 359–363, 2010.
- [39] X. Li, L. Wang, and E. Sung, "AdaBoost with SVM-based component classifiers," *Engineering Applications of Artificial Intelligence*, vol. 21, no. 5, pp. 785–795, 2008.
- [40] A. Vezhnevets and V. Vezhnevets, "Modest AdaBoost-teaching AdaBoost to generalize better," *Graphicon*, vol. 12, no. 5, pp. 987–997, 2005.
- [41] J. Son, I. Jung, K. Park, and B. Han, "Tracking-by-segmentation with online gradient boosting decision tree," *Proceedings of the IEEE international conference on computer vision*, pp. 3056–3064, 2015.
- [42] S. Peter, F. Diego, F. A. Hamprecht, and B. Nadler, "Cost efficient gradient boosting," *Advances in neural information processing systems*, vol. 30, 2017.



RESEARCH ARTICLE

Receive Date: 28.11.2023

Accepted Date: 01.02.2024

Drug repurposing analysis with co-expressed genes identifies novel drugs and small molecules for bladder cancer

Esra Gov^{a,*}, Gökçe Kaynak Bayrak^b

^aDepartment of Bioengineering, Faculty of Engineering, Adana Alparslan Türkeş Science and Technology University, Adana 01250, Türkiye.
ORCID: 0000-0001-7948-5087

^bDepartment of Biomedical Engineering, Faculty of Engineering and Architecture, İzmir Bakırçay University, İzmir 35665, Türkiye.,
ORCID: 0000-0002-5256-4778

Abstract

Bladder cancer (BC) is the fifth most common malignancy in humans and has poor survival rates. Although there is extensive research on the diagnosis and treatment of BC, novel molecular therapies are essential due to tumor recurrence. In this study, we aim to identify repurposed drugs or small molecules of BC with multi-omics systems biology perspective. Gene expression datasets were statistically analyzed by comparing bladder tumor and normal bladder tissues and differentially expressed genes (DEGs) were determined. Co-expression network of common DEGs for BC was constructed and co-expressed module was found by using tumors and control bladder tissues. Using independent data, we demonstrated the high prognostic capacity of the module genes. Moreover, repurposed drugs or small molecules were predicted by using L1000CDS2 gene expression based-search engine tool. We found numerous drug candidates as 480743.cdx, MK-2206, Geldanamycin, PIK-90, BRD-K50387473 (XMD8-92), BRD-K96144918 (mead acid), Vorinostat, PLX-4720, Entinostat, BIX-01294, PD-0325901 and Selumetinib, that may be used in BC therapy. We report 480743.cdx, BRD-K50387473 (XMD8-92) and mead acid as novel drugs or small molecules that offer crucial step in translational cancer research of BC.

© 2023 DPU All rights reserved.

Keywords: multi-omics data, gene expression, drug repurposing, bladder cancer.

1. Introduction

Bladder cancer (BC) is the fifth most common malignancy in humans and approximately 550,000 new cases occur per year with 200,000 of them resulting in death [1,2]. Approximately, 75% of the tumors do not invade to muscularis propria and these are classified as non-muscle invasive bladder cancer (NMIBC)[3]. Muscle invasive bladder cancer (MIBC) is associated with most bladder cancer morbidity and mortality. Early-stage cancers are mostly treated with tumor resection, but the disease recurrence rate is high (50-80%) and can progress to an invasive type depending on the stage. Therefore, patients usually undergo lifelong surveillance through cystoscopy [4,5].

Chemotherapy and radical cystectomy are the routine treatment methods and immunotherapies are in the clinical trial phase. Also, bladder cancer is a curable disease, therapy is a difficult process and MIBC has a poor survival rate

* Corresponding author. Esra GOV Tel.: 0-322-455-0000; fax: 0-322-455-0032.

E-mail address: egov@atu.edu.tr

of 5-years <50%, but if the tumor spreads to lymph nodes or organs, the survival rate decreases to 35% [3,6]. So, early diagnosis is of vital importance.

Diagnosis of bladder cancer is practiced by standard tests like urine cytology, but the sensitivity is as low as 40% [7]. Multiplex tests including different markers will probably provide higher precision. In recent years, high-throughput molecular profiling studies identified novel biomarkers and therapeutic targets for different cancers [8]. Bladder cancers can be diagnosed without muscle invasion through molecular screening by the use of tumor biomarkers. Thus, metastase- induced morbidity can be prevented and life expectancy may be improved.

The advent of molecular biology methods positively affects the diagnosis and the prediction of outcomes of several cancers. Detailed multi-omic studies are carried out for the discovery of carcinogenesis and progression. In bladder tumors MIBC and NMIBC show different molecular characteristics and their clinical behaviors are distinct [9]. Due to the complexity and heterogeneous structure of bladder cancer, multiple biomarkers are investigated concurrently and some of them accurately predict the prognosis [10]. Multi-omic studies are crucial at this point considering genomic aberrations in tumor transformations. There are few studies focused on bladder cancer, and NMIBC- related research is more limited. Goel et al. used exome and transcriptome sequencing to characterise all grades of NMIBCs to determine prognostic genes and indicated that multi-omic data may help to better identify treatment in high-risk patients [11].

Although there is extensive research on diagnosis and treatment of BC, new molecular therapies are required due to tumor recurrence. Drug repurposing is aimed at approved or failed/abandoned compounds to find new indications for use in a different disease or condition. Drug repurposing studies offer an alternative to conventional drug inventions with their cost effective, cheaper and time saving aspects. For the development and release to the market of a new drug molecule, an average of 12-13 years and an estimated 2-3 billion USD investment are required. Also, the proposed drug is safe as it has been approved by a health regulatory authority. For cancer treatment, there are three repurposed drugs [12,13]. Feng et al. investigated metformin for BC therapy. Metformin is a frequently used hypoglycemic drug and it has been reported in the study that metformin has an anti-proliferative effect on BC stem cells and support the chemotherapy drugs on BC cells [14].

Developing potential marker gene lists of bladder cancer must be the starting point. Lindsprog et al. described transcriptomic and genomic markers of NMIBC and presented an online classification tool [4]. In a whole exome sequencing study, driver mutations in FGFR3, KDMTA, and KDMT2C were found and also DNA methylation and hydroxymethylation were investigated as promising biomarker [5]. Also, it was indicated in the literature that CCNB1, FOXM1, GSN, LAMC2 genes are prognostic expression markers for non-invasive BC [7]. Besides, there are some studies about identifying key genes and pathways in bladder cancer. Gao et al. showed in their GO analysis that mitotic nuclear division, the spindle and protein binding related genes upregulated while cell adhesion, extracellular exosomes and calcium ion binding related genes downregulated [15]. In another research, differentially expressed genes in bladder cancer tissues were identified as mitotic and chromosome assembly, including nucleosome assembly, spindle checkpoint and DNA replication [16]. Also, Tang et al. reported that upregulated DEGs were associated with cell division, nucleoplasm and protein binding, while the downregulated DEGs were associated with 'extracellular matrix organization', 'proteinaceous extracellular matrix' and 'heparin binding' [17].

In the present study, differentially expressed genes (DEGs) were identified by using gene expression datasets including bladder tumor and normal bladder tissues obtained from two different studies. BC specific co-expression network of common DEGs was reconstructed. A co-expressed module was found by using cancerous and normal bladder tissues. The prognostic capability of the module was evaluated. Moreover, potential therapeutic targets and reverse the expression of co-expressed module genes were investigated through L1000CDS2 tool. We report novel

drugs or small molecules that offer crucial prospects for prognosis, treatment and translational cancer research of bladder cancer.

2. Materials and methods

2.1. Transcriptome datasets

The data of transcriptome datasets of BC including GSE7476 [18] and GSE24152 [19] were taken from Gene Expression Omnibus [20]. It was analyzed to identify differentially expressed genes (DEGs) of BC. Both datasets including the arrays of the Affymetrix platform were selected for analysis. A total of 27 samples were selected, including 17 BC and 10 normal bladder tissue samples. BLCA-TCGA-Bladder Urothelial Carcinoma obtained from the Cancer Genome Atlas (TCGA) database as an independent dataset including 390 patients was used in prognosis analysis.

2.2. Identification of differentially expressed genes

Robust Multi Array (RMA) techniques [21] were used for normalization of datasets. Linear models for microarray (LIMMA) package [22] in R language were performed for both dataset to identify DEGs in patients with BC compared to healthy individuals. The obtained DEGs were determined based on the p-values ($p < .05$) and the direction of differentiation was identified using gene expression fold changes (FC). Up regulated and down-regulated genes were identified considering the $FC > 2$ and $FC < 0.5$, respectively.

Gene enrichment analyses of DEGs were performed via the Metascape [23]. The significant terms were determined by using $p < 0.05$ which is the cut-off for statistical significance.

2.3. Differential co-expression analysis and identification of co-expressed modules

Gene expression data of common DEGs of two datasets were obtained from both tumor and control samples, separately. Our differential co-expression network analysis algorithm [24] was applied to both gene expression data of cancerous and normal tissues to identify a BC specific differentially co-expressed network. The mean value of gene expression data of each common DEG was calculated. Afterward, z score normalization of each common DEGs was found. Spearman correlation coefficients (SCC) of mean gene expression were calculated in BC and normal bladder tissues, separately since data are not normally distributed. The significant pair-wise gene correlations of common DEGs were determined by using an SCC cutoff ($p < 0.05$). It was constructed a BC specific differential co-expression network in cancerous samples compared to normal bladder tissues. Two parameters were described to identify significant differentially co- expression profiles between cancerous and normal tissues: (i) Gene pair that show a significant correlation score in the cancerous samples, but no significant correlation in the normal bladder samples. (ii) Although gene pair show a significant correlation in both cancerous and normal bladder samples, it was selected co-expression direction is different in cancerous and normal bladder samples (i.e: positive and negative correlation score).

The MCODE plugin [25] of the Cytoscape [26] was used to identify network modules of the differential co-expression network. For further analysis, modules with a minimum of 10 nodes (genes) and a network density of 0.50 were taken into consideration.

2.4. Prognostic capability analysis of co-expressed module genes

An independent BC dataset obtained from TCGA was used to investigate the prognostic capabilities of the co-expressed module genes. Cox proportional hazards regression analysis was executed via SurvExpress bioinformatics

validation tool [27]. In SurvExpress, each cancerous sample of the BLCA-TCGA-Bladder Urothelial Carcinoma was categorized according to their prognostic index as low- and high-risk groups. The prognostic performance of the module genes was determined through the log-rank test and Kaplan-Meier (KM) plots.

2.5. Drug repurposing analysis

L1000CDS2 [28] is a search tool that presents a listing of FDA-approved drugs or experimentally studied small molecules that are defined to reverse or mimic the down-regulated and up-regulated genes. It was executed L1000CDS2 analyses by using the differential co-expressed up and down regulated module genes, to identify drug or small molecules that reverse the regulation direction of genes in BC (<https://maayanlab.cloud/L1000CDS2/#/index>).

3. Results

3.1. Gene expression profiles of bladder cancer

Differentially expressed genes (DEGs) for BC were identified in BC compared to normal bladder samples through statistical analysis. We obtained 2490 DEGs (p value<0.05) where 727 upregulated and 1769 downregulated genes from GSE7475 datasets. Analysis of GSE24152 identified 832 DEGs (p value <0.05), 483 upregulated and 725 downregulated DEGs. (Figure 1A). 131 common DEGs between the two datasets were determined (Figure 1B). The gene enrichment analysis of common DEGs indicated that signaling by aberrant PI3K in cancer, ras signaling, cytoskeleton and proteoglycan-related biological processes were significantly enriched terms (Figure 1C).

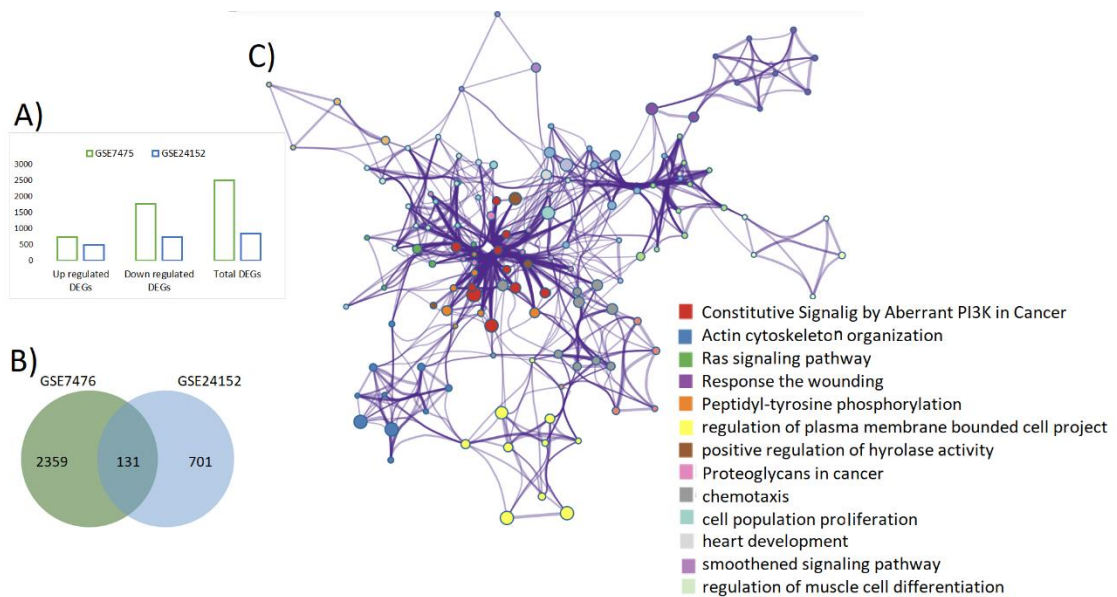


Fig. 1. Gene expression analysis results of bladder cancer. (A) The graph of up-regulated genes, down-regulated genes and total differentially expressed genes (DEGs) ($p < .05$). (B) The venn diagram represents the number of common DEGs between both datasets. (C) Biological pathway and gene ontology enrichment analysis results of common DEGs. The network was obtained from Metascape bioinformatics tool.

3.2. Differential co-expression network in bladder cancer

The differential co-expression network analyses resulted in a total of 759 significant differential correlations among 131 common DEGs in cancerous tissues compared with normal bladder tissues. Differentially co-expressed gene module which is highly-clustered co-expressed genes including the number of 22 nodes (ETV4, SLC44A5, EVPL, ARL13B, COPZ1, EPRS, ELN, EPHA3, GLTP, SSBP2, SLC39A11, MSN, SCRIB, FGFR1, MLXIP, EMP1, EPHA7, FAM83B, JAZF1, CCT5, FGFR3 and EPB41L2) and 191 edges and a network density is 83% was obtained (Figure 2). It was performed gene enrichment analysis on the module genes. The statistically significant biological process associated GO terms (p value <0.05) were obtained, Top three terms were identified as transmembrane receptor protein tyrosine kinase signaling pathway (GO:0007169), cell recognition (GO:0008037) and morphogenesis of an epithelium (GO:0002009) (Figure 2).

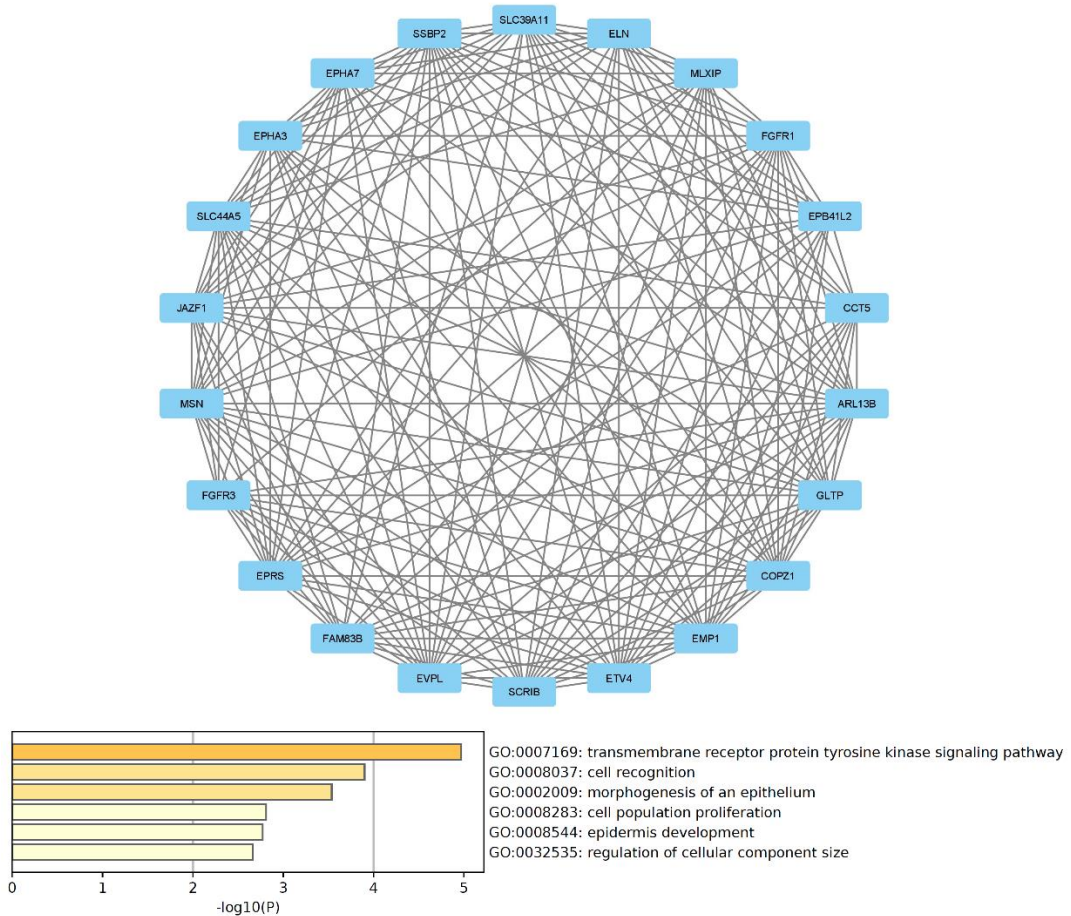


Fig. 2. Differential co-expressed module in bladder cancer. Statistically significant differentially correlated common DEGs were represented as nodes and significant Spearman correlation values between the DEGs were represented as edges.

3.3. Prognostic capability of module genes by using independent bladder cancer dataset

To determine the prognostic capability of the module genes, Cox proportional hazards regression analysis was

executed in SurvExpress validation bioinformatics tool. Cancerous tissue samples were categorized into low- and high-risk groups according to their prognostic index calculated by using survival times. For this purpose, an independent RNA-Seq dataset (n = 390), BLCA-TCGA-Bladder Urothelial Carcinoma-July 2016, was performed and prognostic capabilities of the module genes based on survival data were analyzed by using the log-rank test and Kaplan-Meier plots (p<0.01) (Figure 3).

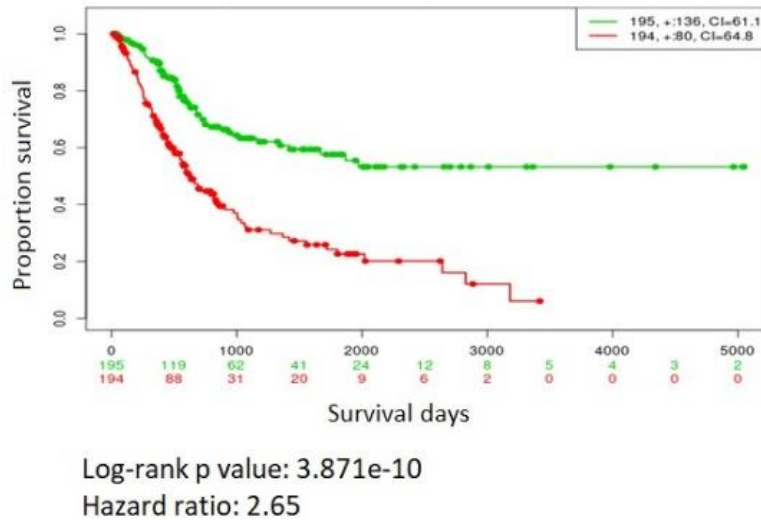


Fig. 3. The survival analysis results of the BC specific module genes were represented with Kaplan-Meier curve. Low-risk and high-risk of patients samples were represented as green and red colors, respectively.

3.4. Putative repurposed drugs and small molecules for bladder cancer

The LINCSL1000 gene expression based-search engine tool was utilized concerning to the up and down-regulated module genes as predictive molecular targets in BC. The L1000CDS2 web tool was used to enter the list of up and down regulated module genes in order to search for drugs or small compounds that could change the gene expression profiles of the relevant genes. With an overlap score of ≥ 0.1765 , the highest score to reverse expression profiles on up-regulated and/or down-regulated module genes in various cell lines, thirteen drugs or small compounds were obtained as repurposed drugs, demonstrating potential (Table 1). Repurposed drug or small molecule candidates were investigated in the literature to understand whether BC is associated with or not.

Table 1. Repurposed drugs and small molecules for bladder cancer with an overlap score of ≥ 0.1765 according to LINCS L1000 gene expression based-search engine tool analysis.

Rank	Score	Perturbation	Cell line	Dose, μm	Time, h	Reversed expression of genes
1	0.1765	480743.cdx	HT29	80.0	24.0	<i>EPB41L2, EPHA3, FGF1</i>
2	0.1765	MK-2206	LOVO	10.0	6.0	<i>EMP1, FGFR1, MSN</i>
3	0.1765	Geldanamycin	NCIH2073	10.0	6.0	<i>EMP1, ELN, EPHA3</i>
4	0.1765	PIK-90 (PI 3-K inhibitor IX)	RMGI	10.0	6.0	<i>EMP1, FGFR1, MSN</i>
5	0.1765	BRD-K50387473(XMD8-92)	HEPG2	10.0	6.0	<i>EMP1, FGFR3, GLTP</i>
6	0.1765	BRD-K96144918(Mead acid)	A549	10.0	6.0	<i>EMP1, FGFR3, SSBP2</i>
7	0.1765	Vorinostat(SAHA,suberoylanilide	MCF7	10.0	6.0	<i>EPB41L2, EPHA3, MSN</i>

		hydroxamic acid)				
8	0.1765	PLX-4720	A375	1.11	24	<i>ETV4, EPHA3, SSBP2</i>
9	0.1765	PLX-4720	A375	0.37	24	<i>ETV4, EPHA3, SSBP2</i>
10	0.1765	Entinostat (BRD-K77908580)	MCF7	3.33	24	<i>EPRS, FGFR1, MSN</i>
11	0.1765	BIX-01294	HEPG2	10	24	<i>FGFR3, FGFR1, MSN</i>
12	0.1765	PD-0325901	A375	0.04	24	<i>ETV4, EPHA3, SSBP2</i>
13	0.1765	Selumetinib	HME1	10	24	<i>EMPI, ETV4, SSBP2</i>

4. Discussion

Bladder cancer (BC) is among the top 10 most common tumors with the 6th most diagnosed cancer worldwide [14]. Pathologically, BC is categorized as NMIBC and MIBC. Due to its clinical and molecular complexity, it is not possible to forecast which stage tumor will progress to an aggressive form. In our study, we found thirteen drugs or small molecules that can reverse gene expressions and searched for BC. There are few genomic aberrations in FGFR3, PIK3CA, RAS oncogenes on bladder cancer and these genes targeted treatments are being investigated in several clinical trials.

Also, there are essential molecular pathways that act in urothelial tumorigenesis as the RAS-MAPK pathway and PI3K-Akt pathway [29]. MK2206 is an allosteric Akt inhibitor, that blocks the phosphorylation and activation of Akt, therefore prevent the proliferation of many human cancer cell lines [30]. There are several searches on the effect of MK2206 on breast, thymic, lung, colorectal, endometrial, renal cancers [31-34]. Sathe et al. examined MK2206 on 11 BC cell lines and stated a decreasing AKT phosphorylation depending on the dose. They also specified that an increase in caspase 3/7 activity of sensitive cells is related with an increase in apoptosis [31]. In another study Sun et al. showed the booster effect of MK2206 on cisplatin (CDDP)-induced cytotoxicity and bladder cancer cells apoptosis and suppressive effect on tumor growth in subcutaneous xenograft models [35]. They also have seen the same incidence of testicular cancer and elucidated the mechanism by the suppressed expression of Akt pathway [36]. Zhang et al. studied on drug sensitivity and remarked that bladder tumor cells with Retinoblastoma 1 (RB1, a cell division regulator) mutation (found in 25% of patients) are more resistant to MK-2206, Dactolisib and GNE-317 [6].

The second drug candidate is geldanamycin, a type of antibiotic from the benzoquinone ansamycins (BAs) family. It has heat shock protein 90 (Hsp90) inhibitory properties. Hsp90 upregulation in malignancies acts as protective for tumorigenesis, therefore Hsp90 inhibition is one of the trend cancer treatment modalities. Geldanamycin has been used in numerous cancer studies as gall bladder, thyroid, liver, osteosarcoma, lungs, melanoma, cervical, breast, prostate, colorectal cancers [37-40]. Germano et al. emphasized that Ron can gain tumorigenic potential by single point mutations and aberrant activation of Ron has been determined in colorectal adenocarcinomas, non-small cell lung tumors and primary breast carcinomas. Also, it is stated that Ron expression is correlated with bladder tumor phase [41]. Karkoulis et al., demonstrated anti-neoplastic properties of geldanamycin in human bladder tumor cell lines (RT4 and T24) [42]. Unfortunately, it is unstable, cardiotoxic, oculotoxic, hepatotoxic and has low aqueous dissolubility, therefore several GA analogues have been produced [39,43].

Re-regulation of the phosphoinositide 3-kinase (PI3K)/AKT/mammalian target of rapamycin (mTOR) pathway in tumors is also one of the most researched cancer therapies. PIK90 is a small molecule that acts as a PI3K inhibitor and its antitumor activity has been shown for breast, ovarian cancer cells [44] and bladder cells [45]. In metastatic bladder cancers, the PI3K pathway is around 72% overactive. Sathe et al. characterized molecular mechanisms of PI3K pathway signaling in bladder cancer cells in their study. They used different pathway inhibitors and small molecules such as PIK90 and MK-2206 and demonstrated their anti-cancer properties. Finally, they emphasized that

simultaneous targeting of the PI3K, AKT and mTORC1 pathways is required for effective tumor growth inhibition [46].

Histone deacetylases regulate the expression of numerous proteins involved in malignant tumor initiation and progression. Therefore, histone deacetylase inhibitors (HDACi) are being produced for cancer treatment. Vorinostat is one of them and approved by the U.S. Food and Drug Administration (FDA) for the cure of advanced and refractory cutaneous T-cell lymphoma. It has been also applied in numerous clinical cancer therapy trials such as head and neck squamous cell carcinomas, breast, lymphoma, non-small cell lung cancer (NSCLC), glioblastoma multiforme [47]. Besides, vorinostat reportedly has antiproliferative effects in various cancer cells such as ovarian cancer cells [48], renal cancer cells (in combination with Fluvastatin [49]), breast cancer [50]. and human bladder cancer cell lines [51]. Kaletsch et al. examined the effect of a novel HDACi 19i (LMK235) with vorinostat (SAHA) and the HDAC4-specific HDACi TMP269 on urothelial carcinoma cells, and stated the disturbed mitosis with apoptosis of cells after treatment [52]. In a clinical study, Quinn et al. observed the toxic effect of Vorinostat and proposed to use a lower dose [53]. Additionally, vorinostat has low solubility and permeability. Moreover, due to its high metabolized in the liver, its therapeutic benefits are poor when used as monotherapy. So various delivery systems are being developed to increase its clinical utility [47].

Nearly 20% of urothelial tumors were identified as RAF1 activation dependent and they use RAF/MEK/ERK signaling pathway. Therefore, these cells are sensitive to RAF inhibitors and RAF plus MEK inhibition combination. Bekele et al. specified in their investigation that bladder cancer cell lines are sensible to BRAFV600E inhibitor PLX4720 [54]. Also, Chen et al., determined Transient receptor potential (TRP) family gene expression in bladder and para-carcinoma tissues and TRP expression is associated with the sensibility to several drugs inclusive of PLX-4720 [55]. Otherwise, PLX4720 is applied for other cancer types such as melanoma [56,57] and colorectal carcinoma [58].

Entinostat is also a selective HDAC inhibitor like vorinostat, and has been investigated as either a single agent or a combination in non-small cell carcinoma, Hodgkin's lymphoma, breast, and myelodysplastic syndrome. Pili et al., used entinostat and 13-cis retinoic acid on patients having solid tumours and searched for its safety [59]. Truong et al., evaluated the anticancer activity and cell-autonomous mechanism of entinostat in bladder cancer and they proved that entinostat had substantial antitumor efficiency in immune-competent but not immune-compromised hosts. Also, they indicated that when entinostat was combined with programmed cell death protein 1 (PD-1), its antitumor responses were increased, and long-term immunologic memory was induced in host [60]. In another study, entinostat was treated with a combination of the approved drug decitabine on platinum resistant bladder tumor cells and researchers proposed its usage on cisplatin-resistant bladder cancer [61]. Additionally, macrophages play an essential role in immune response and tumor-associated macrophages partake in solid tumor development with anti-tumorigenic or pro-tumorigenic character based on their polarization. So, there are various clinical trials targeting macrophages. In bladder cancer research different agents were used as Vorinostat and Entinostat [62].

BIX-01294 is the first selective G9a, a histone methyltransferase also known as euchromatic histone-lysine N-methyltransferase 2 (EHMT2), inhibitor. G9a is highly active in some cancer types as esophageal, ovarian, and gastric cancer. Inhibition of this gene prevents tumor cell proliferation and metastases by stimulating autophagic cell death, apoptosis, and cell cycle arrest. Also, antiapoptotic proteins are decreased and proapoptotic proteins are increased with G9a depletion. BIX-01294 was proven to induce cell death in breast cancer, head and neck squamous cell carcinoma, neuroblastoma cells and bladder cancer cells [63-65]. Cui et al., showed BIX-01294 induced endoplasmic reticulum stress and apoptosis in human bladder cancer cells occurred via caspase-dependent pathway [63]. Li et al. also examined the anti-proliferative effect of BIX-01294 on bladder carcinoma cell lines T24 and UMUC-3 and proposed that G9a might be a good therapeutic target in bladder cancer [66]. In a clinical study, the role of G9a in Bacillus Calmette-Guerin (BCG)-treated NMIBC patients was investigated. BCG bladder instillation is the gold standard treatment in high-risk NMIBC patients. In the experiments BIX-01294 was used to examine the

effect on trained immunity responses in vitro and finally it was emphasized that suppression of G9a is important in the stimulation of trained immunity [67].

Cirone et al. investigated the effect of PI3K/mTOR inhibitor PF-0469502 and a MEK inhibitor PD-0325901 in in vitro and in vivo models of bladder cancer and determined the slowed tumor growth. Therefore, they suggested the therapy as a potential treatment approach for bladder cancer [68]. Also, in an in vivo study, bladder cancer invasion was prevented by TGF β receptor inhibitor LY2157299 and MEK inhibitor PD-0325901 [69]. Zhang et al., investigated the effect of PD0325901 (MEK/ERK inhibitor), CHIR99021 (GSK3 pathway inhibitor), small molecule inhibitors SB431542 (ALK inhibitor) and valproic acid (VPA; HDAC inhibitor) on uterine cervix carcinoma cells, bladder cancer cells and squamous cell carcinoma cells. The results of the analysis reveal that combined inhibition of MEK/ERK, ALK and GSK3 may be a potential cancer therapy [70].

Selumetinib is a small molecule, having a short half-life, and acts as an oral mitogen-activated ERK kinase (MEK)-inhibitor. There are lots of research about its usage in various cancers such as melanoma, colorectal, pancreatic, breast cancers, papillary thyroid carcinoma, non-small cell lung cancer, pediatric low-grade gliomas and neurofibromatosis 1 (NF1) [71]. Preclinical studies suggest that it may improve the effect of chemotherapeutic drugs. LoRusso et al. demonstrated that selumetinib was safe when combined with docetaxel and dacarbazine in advanced solid tumors [72]. Schulz et al. used gamma-secretase inhibitor (GSI) dibenzazepine and selumetinib in bladder cancer cell line and suggested inhibition of both NOTCH and MAPK signaling most strongly suppressed tumor growth [73]. Additionally, there is an ongoing clinical trial using Selumetinib in Muscle Invasive Bladder Cancer (NCT02546661)

In cancer related studies it was indicated that mead acid containing diet inhibited breast cancer by suppressing cell proliferation, also mead acid inhibited some tumorigenic features of human breast, urothelium, and colon cell lines [74-76].

BRD-K50387473 (XMD8-92) is an extracellular signal-regulated kinase 5 (ERK5), a member of the mitogen-activated protein kinase (MAPK) family, inhibitor. Kang et al., induced apoptosis of acute myeloid leukemia cell lines and they proposed that XMD8-92 may be an efficient adjuvant in AML chemotherapy [77]. Besides, XMD8-92 was utilized in hepatocellular carcinoma cells [78], lung and cervical cancers, pancreatic cancers [79]. and Yang et al., emphasized that XMD8-92 may be an effective approach for treating human cancer [80].

In summary, in our study bladder cancer associated genes were found to be associated with transmembrane receptor protein tyrosine kinase signaling pathway, cell recognition and morphogenesis of an epithelium. We found numerous drug candidates as 480743.cdx, MK-2206, Geldanamycin, PIK-90, BRD-K50387473 (XMD8-92), BRD-K96144918 (mead acid), Vorinostat, PLX-4720, Entinostat, BIX-01294, PD-0325901 and Selumetinib, that may be used in bladder cancer therapy. As discussed above, all candidates are investigated in terms of bladder cancer except from 480743.cdx, BRD-K50387473 (XMD8-92) and mead acid. These three prospective drugs may be evaluated for bladder cancer therapy after carrying out more advanced analyses and preclinical studies.

Acknowledgement

All authors declared that there are no conflicts of interest.

Appendix

Availability of data and materials

Transcriptomic data have been deposited in Gene Expression Omnibus with an accession number GSE7476 and GSE24152. For validation studies, it was used BLCA-TCGA-Bladder Urothelial Carcinoma-July 2016 data obtained from TCGA.

References

- [1] Tang, C., Yu, M., Ma, J., and Zhu, Y., “Metabolic classification of bladder cancer based on multi- omics integrated analysis to predict patient prognosis and treatment response”, *J Transl Med*, vol. 19 no. 205, 2021, doi: 10.1186/s13073-022-01056-4.
- [2] Yu, E.Y.-W., Zhang, H., Fu Y., et al., “Integrative multi-omics analysis for the determination of non-muscle invasive vs. muscle invasive bladder cancer: a pilot study”, *Curr Oncol*, vol. 29, no. 8, pp. 5442–5456, 2022, doi: 10.3390/currenocol29080430.
- [3] Mo, Q., Li, R., Adeegbe, D.O., Peng, G., and Chan, K.S., “Integrative multi-omics analysis of muscle-invasive bladder cancer identifies prognostic biomarkers for frontline chemotherapy and immunotherapy”, *Commun Biol*, Vol. 3, no. 784, 2020, doi:10.1038/s42003-020-01491-2.
- [4] Lindskog, S.V., Prip, F., Lamy, P., et al., “An integrated multi-omics analysis identifies prognostic molecular subtypes of non-muscle invasive bladder cancer.”, *Nat Commun.*, vol. 12, no. 2301, 2021. doi: 10.1038/s41467-021-22465-w.
- [5] Shi, Z.-D. Han, X.-X., Song, Z.-J., et al., “Integrative multi- omics analysis depicts the methylome and hydroxymethylome in recurrent bladder cancers and identifies biomarkers for predicting PD- L1 expression.”, *Biomark. Res.*, vol. 11, no. 47, 2023, doi: 10.1186/s40364-023-00488-3.
- [6] Zhang, X., Wang, J., Lu, J. et al., “Robust prognostic subtyping of muscle-invasive bladder cancer revealed by deep learning-based multi-omics data integration”, *Front. Oncol.*, vol. 11, no. 689626, 2021. doi.org/10.3389/fonc.2021.689626.
- [7] You, C., Piao, X.M., Kang, K., Kim, Y.J., and Kang, K., “Integrative transcriptome profiling reveals ska3 as a novel prognostic marker in non-muscle invasive bladder cancer.”, *Cancers*, vol. 13, no. 18, 4673, 2021, doi: 10.3390/cancers13184673.
- [8] Demirtas, T.Y., Rahman, R., Capkin Yurtsever, M., and Gov, E., “Forecasting gastric cancer diagnosis, prognosis, and drug repurposing with novel gene expression signatures.”, *OMICS A J Integr. Biol.*, vol. 26, no. 1, 2022, DOI:10.1089/omi.2021.0195.
- [9] Knowles, M.A., and Hurst, C.D., “Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity.”, *Nat. Rev. Cancer*, vol. 15, no. 1, pp. 25-41, 2014, doi: 10.1038/nrc3817.
- [10] Hurst, C.D., Cheng, G., and Platt, F.M., “Stage-stratified molecular profiling of non-muscle-invasive bladder cancer enhances biological, clinical, and therapeutic insight.”, *Cell Rep. Med.*, vol. 2, 100472, 2021.
- [11] Goel, A., Ward, D.G., Noyvert, B., et al., “Combined exome and transcriptome sequencing of non- muscle- invasive bladder cancer: associations between genomic changes, expression subtypes, and clinical outcomes.”, *Genome Med.*, vol. 14, no. 59, 2022, doi: 10.1186/s13073-022-01056-4.
- [12] Gonzalez- Fierro, A., Romo- Perez, A., Chavez- Blanco, A., Dominguez- Gomez, G., and Duenas- Gonzalez, A., “Does therapeutic repurposing in cancer meet the expectations of having drugs at a lower price?”, *Clin. Drug Investig.*, vol. 43, no. 4, pp. 227–239, 2023, doi: 10.1007/s40261-023-01251-0.
- [13] Malik, J.A., Ahmed, S., Momin, S.S., et al. “Drug repurposing: A new hope in drug discovery for prostate cancer”, *ACS Omega*, vol. 8, no. 1, pp. 56–73, 2023, doi: 10.1021/acsomega.2c05821
- [14] Feng, Y., Jia, B., and Shen, Z., “Metformin and bladder cancer: Drug repurposing as a potential tool for novel therapy: A review”, *Medicine*, vol. 101, no. 45, 2022, doi: 10.1097/MD.00000000000031635.
- [15] Gao, X., Chen, Y., Chen, M., Wang, S., Wen, X., Zhang, S., “Identification of key candidate genes and biological pathways in bladder cancer.” *Peer J*, vol. 6, 2018, doi: 10.7717/peerj.6036.
- [16] Wang, J.P., Leng, J.Y., Zhang, R.K., Zhang, L., Zhang, B., Jiang, W.Y., Tong, L., “Functional analysis of gene expression profiling-based prediction in bladder cancer.”, *Oncol. Lett.*, vol. 15, no. 6, pp. 8417-8423, 2018, doi: 10.3892/ol.2018.8370.
- [17] Tang, F., He, Z., Lei, H., Chen, Y., Lu, Z., Zeng, G., Wang, H., “Identification of differentially expressed genes and biological pathways in bladder cancer”, *Molecular Medicine Reports*, vol. 17, no. 5, pp 6425-6434, 2018, doi: 10.3892/mmr.2018.8711.
- [18] Mengual, L, Burset, M, Ars, E, et al., “DNA microarray expression profiling of bladder cancer allows identification of noninvasive diagnostic markers.”, *J Urol.*, vol. 182, no. 2, pp. 741-748, 2009, doi: 10.1016/j.juro.2009.03.084.
- [19] Zhang, Z., Furge, K.A., Yang, X.J., The, B.T., and Hansel, D.E., “Comparative gene expression profiling analysis of urothelial carcinoma of the renal pelvis and bladder.”, *BMC Med Genom.*, Vol. 3, no. 58, 2010, doi: 10.1186/1755-8794-3-58.
- [20] Barrett, T, Wilhite, S.E., Ledoux, P, et al., “NCBI GEO: Archive for functional genomics data sets-Update”, *Nucleic Acids Res.*, 41(D1), D991–D995, 2013, doi: 10.1093/nar/gks1193.
- [21] Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P., “A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.”, *Bioinformatics*, vol. 19, no. 2, pp. 185–193, 2013, doi: 10.1093/bioinformatics/19.2.185.
- [22] Gentleman, R., Carey, V.J., Huber, W., Irizarry, R.A., Dudoit, S., and Smyth, G.K. LIMMA: Linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions Using R and Bioconductor.*, eds. Springer: New York, New York, USA. 2005, pp. 397–420.
- [23] Zhou, Y., Zhou, B., Pache, L., et al., “Metascape provides a biologist-oriented resource for the analysis of systems-level datasets.”, *Nat Commun*, 10, 1523, 2019, https://doi.org/10.1038/s41467-019-09234-6.
- [24] Gov, E., and Arga, K.Y., “Differential co-expression analysis reveals a novel prognostic gene module in ovarian cancer.”, *Sci Rep.*, vol. 7, no. 4996, 2017, doi:10.1038/s41598-017-05298-w
- [25] Saito, R, Smoot, M.E., Ono, K., et al., “A travel guide to Cytoscape plugins.”, *Nat Methods*, vol. 9, pp. 1069-1076, 2012, doi: 10.1038/nmeth.2212.
- [26] Cline, M.S., Smoot, M., Cerami, E., et al., “Integration of biological networks and gene expression data using Cytoscape”, *Nat Protoc*, vol. 2, pp. 2366–2382, 2007, doi:10.1038/nprot.2007.324
- [27] Aguirre-Gamboa, R., Gomez-Rueda, H., Martinez-Ledesma, E., et al., “SurvExpress: An online biomarker validation tool and database for cancer gene expression data using survival analysis.”, *PLoS One*, vol. 8, e74250, 2013, doi:10.1371/journal.pone.0074250

- [28] Duan, Q., Reid, S.P., Clark, N.R., et al., "L1000CDS2: LINCSL1000 characteristic direction signatures search engine.", *NPJ Syst Biol Appl*, vol. 2, no. 16015, 2016, doi:10.1038/npsba.2016.15
- [29] Kompier, L.C., Lurkin, I., van der Aa, M.N.M., et al., "FGFR3, HRAS, KRAS, NRAS and PIK3CA mutations in bladder cancer and their potential as biomarkers for surveillance and therapy." *PLoS ONE*, 5, vol. 11, e13821, 2010, doi: 10.1371/journal.pone.0013821.
- [30] Lai, W.T., Cheng, K.L., Baruchello, R., et al., "Hemiasterlin derivative (R)(S)(S)-BF65 and Akt inhibitor MK-2206 synergistically inhibit SKOV3 ovarian cancer cell growth." *Biochem. Pharmacol.*, vol. 113, pp. 12–23, 2016, doi: 10.1016/j.bcp.2016.06.010.
- [31] Sathe, A., Guerth, F., Cronauer, M.V., et al., "Mutant PIK3CA controls DUSP1-dependent ERK 1/2 activity to confer response to AKT target therapy", *Br. J. Cancer*, vol. 111, pp. 2103–2113, 2014, doi: 10.1038/bjc.2014.534.
- [32] Jonasch, E., Hasanov, E., Corn, P.G., et al., "A randomized phase 2 study of MK-2206 versus everolimus in refractory renal cell carcinoma." *Ann Oncol.*, vol. 28, pp. 804–808, 2017, doi: 10.1093/annonc/mdw676.
- [33] Lee, E.K., Tan-Wasielewski, Z., Aghajanian, C. et al., "Results of an abbreviated phase II study of AKT inhibitor MK-2206 in the treatment of recurrent platinum-resistant high grade serous ovarian, fallopian tube, or primary peritoneal carcinoma (NCT 01283035)", *Gynecol Oncol Rep.*, vol. 32, no. 100546, 2020, doi: 10.1016/j.gore.2020.100546
- [34] Stover, E.H., Xiong, N., Myers, A.P., et al., "A phase II study of MK-2206, an AKT inhibitor, in uterine serous carcinoma." *Gynecol Onc Rep.*, vol. 40, no. 100974, 2022, doi: 10.1016/j.gore.2022.100974.
- [35] Sun, D., Sawada, A., Nakashima, M., Kobayashi, T., Ogawa, O., and Matsui, Y., "MK2206 potentiates cisplatin-induced cytotoxicity and apoptosis through an interaction of inactivated Akt signaling pathway." *Urol Oncol: Semin Orig.*, vol. 33, no. 3, e17-26, 2015, doi: 10.1016/j.urolonc.2014.10.018.
- [36] Sun, D., Wang, J., Zhang, H. et al., "MK2206 Enhances Cisplatin-Induced Cytotoxicity and Apoptosis in Testicular Cancer Through Akt Signaling Pathway Inhibition." *Transl Oncol.*, vol. 13, no. 100769, 2020, doi: 10.1016/j.tranon.2020.100769.
- [37] Wang, J., Li, Z., Lin, Z., et al., "17-DMCHAG, a new geldanamycin derivative, inhibits prostate cancer cells through Hsp90 inhibition and survivin downregulation." *Cancer Lett.*, vol. 362, pp. 83-96, 2015, doi: 10.1016/j.canlet.2015.03.025.
- [38] Zeynali-Moghaddam, S., Mohammadian, M., Kheradmand, F., et al., "A molecular basis for the synergy between 17-allylamino-17-demethoxy geldanamycin with Capecitabine and Irinotecan in human colorectal cancer cells through VEGF and MMP-9 gene expression." *Gene*, vol. 684, pp. 30–38, 2019, doi: 10.1016/j.gene.2018.10.016.
- [39] Liew, H.Y., Tan, X.Y., Chan, H.H., Khaw, K.Y., and Ong, Y.S., "Natural HSP90 inhibitors as a potential therapeutic intervention in treating cancers: A comprehensive review." *Pharmacol Res.*, vol. 181, no. 106260, 2022, doi: 10.1016/j.phrs.2022.106260.
- [40] Parma, B., Wurdak, H., and Ceppi, P., "Harnessing mitochondrial metabolism and drug resistance in non-small cell lung cancer and beyond by blocking heat-shock proteins." *Drug Resist Updat.*, vol. 65, no. 100888, 2022, doi: 10.1016/j.drug.2022.100888.
- [41] Germano, S., Barberis, D., Santoro, M.M., et al., "Geldanamycins trigger a novel ron degradative pathway, hampering oncogenic signaling." *J Biol Chem.*, vol. 281, no. 31, pp. 21710-21719, 2006, doi:10.1074/jbc.M602014200
- [42] Karkoulis, P.K., Stravopodis, D.J., Konstantakou, E.G., and Voutsinas, G.E., "Targeted inhibition of heat shock protein 90 disrupts multiple oncogenic signaling pathways, thus inducing cell cycle arrest and programmed cell death in human urinary bladder cancer cell lines." *Cancer Cell Int.*, vol. 13, no. 11, 2013, doi.org/10.1186/1475-2867-13-11
- [43] Tang, Y., Zhou, Y., Fan, S., Wen, Q., "The multiple roles and therapeutic potential of HSP60 in cancer." *Biochem Pharmacol.*, vol. 201, no. 115096, 2022, https://doi.org/10.1016/j.bcp.2022.115096
- [44] Dockx, Y., Vangestel, C., Van den Wyngaert, T., et al., "Early changes in [18F]FDG Uptake as a readout for PI3K/Akt/mTOR targeted drugs in HER-2-positive cancer xenografts." *Mol Imaging.*, pp. 1-14, 2021, doi: 10.1155/2021/5594514.
- [45] Tong, Z., Sathe, A., Ebner, B., et al., "Functional genomics identifies predictive markers and clinically actionable resistance mechanisms to CDK4/6 inhibition in bladder cancer." *J Exp Clin Cancer Res.*, vol. 38, no. 322, 2019, doi.org/10.1186/s13046-019-1322-9.
- [46] Sathe, A., Chalaud, G., Oppolzer, I., et al., "Parallel PI3K, AKT and mTOR inhibition is required to control feedback loops that limit tumor therapy." *PloS one*, 13, 1, e0190854, 2018, doi: 10.1371/journal.pone.0190854.
- [47] Le, V.K.H., Pham, T.P.D., and Truong, D.H., "Delivery systems for vorinostat in cancer treatment: An updated review." *J Drug Deliv Sci Technol.*, vol. 61, no. 102334, 2021, https://doi.org/10.1016/j.jddst.2021.102334
- [48] Ma, X., Wang, J., Liu, J., et al., "Targeting CD146 in combination with vorinostat for the treatment of ovarian cancer cells." *Oncol Lett.*, vol. 13, pp. 1681-1687, 2017, doi: 10.3892/ol.2017.5630
- [49] Okubo, K., Isono, M., Miyai, K., Asano, T., and Sato, A., "Fluvastatin potentiates anticancer activity of vorinostat in renal cancer cells." *Cancer Sci.*, vol. 111, no. 1, pp. 112-126, 2020, doi: 10.1111/cas.14225.
- [50] Wawruszak, A., Borkiewicz, L., Okon, E., Kukula-Koch, W., Afshan, S., and Halasa, M., "Vorinostat (SAHA) and breast cancer: An overview." *Cancers*, vol. 13, no. 18, 2021, doi: 10.3390/cancers13184700.
- [51] Wang, D., Ouyang, S., Tian, Y., et al., "Intravesical treatment with Vorinostat can prevent tumor progression in MNU induced bladder cancer." *J Cancer Ther.*, vol. 4, no 6, 2013, DOI: 10.4236/jct.2013.46A3001.
- [52] Kaletsch, A., Pinkerneil, M., Hoffmann, M.J., et al., "Effects of novel HDAC inhibitors on urothelial carcinoma cells." *Clin Epigenetics.*, vol. 10, no. 100, 2018, https://doi.org/10.1186/s13148-018-0531-y.
- [53] Quinn, D.I., Tsao-Wei, D.D., Twardowski, P., et al., "Phase II study of the histone deacetylase inhibitor vorinostat (Suberoylanilide Hydroxamic Acid; SAHA) in recurrent or metastatic transitional cell carcinoma of the urothelium – an NCI-CTEP sponsored: California Cancer Consortium trial, NCI 6879." *Investig New Drugs.*, no. 39, pp. 812-820, 2021, doi: 10.1007/s10637-020-01038-6.
- [54] Bekele, R.T., Samant, A.S., Nassar, A.H., "RAF1 amplification drives a subset of bladder tumors and confers sensitivity to MAPK-directed therapeutics." *J Clin Invest.*, vol. 131, no. 22, 2021, doi: 10.1172/JCI147849.

- [55] Chen, Z., Zhao, Y., Tian, Y., Cao, R., Shang, D., “Pan-cancer analysis of the TRP family, especially TRPV4 and TRPC4, and its expression correlated with prognosis, tumor microenvironment, and treatment sensitivity.” *Biomolecules*, vol. 13, no. 282, 2023, doi: 10.3390/biom13020282.
- [56] Wang, L., de Oliveira, R.L., Huijberts, et al., “An acquired vulnerability of drug-resistant melanoma with therapeutic potential.” *Cell.*, vol. 173, pp. 1413-1425, 2018, doi: 10.1016/j.cell.2018.04.012.
- [57] Capone, E., Lamolinara, A., D'Agostino, D., et al., “EV20-mediated delivery of cytotoxic auristatin MMAF exhibits potent therapeutic efficacy in cutaneous melanoma.” *J Control Release.*, vol. 277, pp. 48-56, 2018, doi: 10.1016/j.jconrel.2018.03.016.
- [58] Rohde, S., Lindner, T., Polei, S., et al., “Application of in vivo imaging techniques to monitor therapeutic efficiency of PLX4720 in an experimental model of microsatellite instable colorectal cancer.”, *Oncotarget.*, vol. 8, no. 41, pp. 69756-69767, 2017, doi: 10.18632/oncotarget.19263.
- [59] Pili, R., Salumbides, B., Zhao, M., et al., “Phase I study of the histone deacetylase inhibitor entinostat in combination with 13-cis retinoic acid in patients with solid tumours.” *Br J Cancer.*, vol. 106, no. 1, pp. 77-84, 2012, doi: 10.1038/bjc.2011.527.
- [60] Truong, A.S., Zhou, M., Krishnan, B., et al., “Entinostat induces antitumor immune responses through immune editing of tumor neoantigens.” *J Clin Invest.*, vol. 131, no. 6, 2021, doi: 10.1172/JCI138560.
- [61] Wang, C., Hamacher, A., Petzsch, P., et al., “Combination of Decitabine and Entinostat synergistically inhibits urothelial bladder cancer cells via activation of FoxO1.”, *Cancers*, vol. 12, no. 2, 2020, doi: 10.3390/cancers12020337.
- [62] Leblond, M., Zdimerova, H., Desponds, E., Verdeil, G., “Tumor-Associated Macrophages in Bladder Cancer: Biological Role, Impact on Therapeutic Response and Perspectives for Immunotherapy.”, *Cancers*, vol. 13, no. 18, 2021, doi: 10.3390/cancers13184712.
- [63] Cui, J., Sun, W., Hao, X., et al., “EHMT2 inhibitor BIX-01294 induces apoptosis through PMAIP1-USP9X-MCL1 axis in human bladder cancer cells.”, *Cancer Cell Int.*, vol. 15, no. 4, 2015, doi: 10.1186/s12935-014-0149-x.
- [64] Cao, Y., Sun, J., Li, M., et al., “Inhibition of G9a by a small molecule inhibitor, UNC0642, induces apoptosis of human bladder cancer cells.”, *Acta Pharmacol Sin.*, vol. 40, pp. 1076-1084, 2019, doi:10.1038/s41401-018-0205-5
- [65] Yang, Z., Wang, H., Zhang, N., et al., “Chaetocin Abrogates the self-renewal of bladder cancer stem cells via the suppression of the KMT1A–GATA3–STAT3 circuit.”, *Front. Cell Dev. Biol.*, vol. 8, no. 424, 2020, doi: 10.3389/fcell.2020.00424.
- [66] Li F, Zeng J, Gao Y, Guan Z, Ma Z, Shi Q, et al., “G9a Inhibition Induces Autophagic Cell Death via AMPK/mTOR Pathway in Bladder Transitional Cell Carcinoma.”, *PLoS ONE*, vol. 10, no. 9, 2015, doi:10.1371/journal.pone.0138390.
- [67] Mourits, V.P., van Puffelen, J.H., Novakovic, B., et al., “Lysine methyltransferase G9a is an important modulator of trained immunity.”, *Clin Trans Immunol.*, vol. 10, 2021, doi: 10.1002/cti2.1253.
- [68] Cirone, P., Andresen, C.J., Eswaraka, J.R., Lappin, P.B., and Bagi, C.M., “Patient-derived xenografts reveal limits to PI3K/mTOR-and MEK-mediated inhibition of bladder cancer.”, *Cancer Chemother Pharmacol.*, vol. 73, pp. 525-538, 2014, doi: 10.1007/s00280-014-2376-1.
- [69] Sim, W.J., Iyengar, P.V., Lama, D., et al., “c-Met activation leads to the establishment of a TGFβ-receptor regulatory network in bladder cancer progression.”, *Nat Commun.*, vol. 10, no. 1, 2019, doi: 10.1038/s41467-019-12241-2.
- [70] Zhang, Y., Zhang, Y., Li, M., et al., “Combination of SB431542, CHIR99021 and PD0325901 has a synergic effect on abrogating valproic acid- induced epithelial- mesenchymal transition and stemness in HeLa, 5637 and SCC- 15 cells.”, *Oncol Rep.*, vol. 41, pp. 3545-3554, 2019, doi: 10.3892/or.2019.7088.
- [71] Campagne, O., Yeo, K.K., Fangusaro, J., Stewart, C.F., “Clinical pharmacokinetics and pharmacodynamics of Selumetinib.”, *Clin pharmacokinet.*, vol. 60, no. 3, pp. 283-303, 2021, doi: 10.1007/s40262-020-00967-y.
- [72] LoRusso, P.M., Infante, J.R., Kim, KB, et al., “A phase I dose-escalation study of selumetinib in combination with docetaxel or dacarbazine in patients with advanced solid tumors.”, *BMC Cancer.*, vol. 17, no. 173, 2017, doi: 10.1186/s12885-017-3143-6.
- [73] Schulz, G.B., Elezkurtaj, S., Börding, T., et al., “Therapeutic and prognostic implications of NOTCH and MAPK signaling in bladder cancer.”, *Cancer Sci.*, vol. 112, pp. 1987-1996, 2021, doi: 10.1111/cas.14878
- [74] Kinoshita, Y., Yoshizawa, K., Hamazaki, K., et al., “Dietary effects of mead acid on N-methyl-N-nitrosourea-induced mammary cancers in female Sprague-Dawley rats.”, *Biomed Rep.*, vol. 4, pp. 33-39, 2016, doi: 10.3892/br.2015.530.
- [75] Kinoshita Y, Yoshizawa K, Hamazaki K, et al., “Mead acid inhibits the growth of KPL-1 human breast cancer cells in vitro and in vivo.”, *Oncol Rep.*, vol. 32, pp. 1385-1394, 2014, DOI: 10.3892/or.2014.3390.
- [76] Farag, M.A., and Gad, M.Z., “Omega- 9 fatty acids: potential roles in inflammation and cancer management.”, *J Genet Eng Biotechnol.*, vol. 20, no. 48, 2022, doi: 10.1186/s43141-022-00329-0.
- [77] Kang, C., Kim, J.S., Kim, C.Y., Kim, E.Y., and Chung, H.M., “The pharmacological inhibition of ERK5 enhances apoptosis in acute myeloid leukemia cells.”, *Int J Stem Cells*, vol. 11, no. 2, pp. 227-234, 2018, doi: 10.15283/ijsc18053.
- [78] Rovida, E., Di Maira, G., Tusa, I., et al., “The mitogen-activated protein kinase ERK5 regulates the development and growth of hepatocellular carcinoma.”, *Eur J Cancer.*, vol. 64, no. 9, pp. 1454-1465, 2015, doi: 10.1136/gutjnl-2014-306761.
- [79] Sureban, S.M., Maya, R., Weygant, N., et al., “XMD8-92 inhibits pancreatic tumor xenograft growth via a DCLK1-dependent mechanism.”, *Cancer Lett.*, vol. 351, pp. 151-161, 2014, doi: 10.1016/j.canlet.2014.05.011.
- [80] Yang, Q., Deng, X., Lu, B., et al., “Pharmacological inhibition of BMK1 suppresses tumor growth through PML.” *Cancer Cell.*, vol. 18, no. 3, pp. 258-267, 2010, doi: 10.1016/j.ccr.2010.08.008.



RESEARCH ARTICLE

Receive Date: 15.01.2024

Accepted Date: 27.02.2024

A comparative study on the estimation of ultimate bearing capacity of rock masses using finite element and limit equilibrium methods

Serdar Koltuk*

Technical University of Berlin, Ernst-Reuter-Platz 1, Berlin and 10587, Germany, ORCID:0000-0002-6214-0848

Abstract

Most rock masses are excellent foundation materials due to their bearing capacities of MPa. However, the ultimate bearing capacity of rock masses should be accurately estimated in the design of structures with high foundation loads. In this study, the ultimate bearing capacities of a strip footing built on rock masses with different geotechnical properties are determined using the finite element method (FEM) and the failure criterion of Hoek & Brown. The results of FE-analyses are compared to those obtained from the limit equilibrium methods (LEM) in the literature. It has been shown that the FEM with associated flow rule and Terzaghi's limit equilibrium method give similar failure surfaces for most cases, and the ratio of ultimate bearing capacities determined according to the Terzaghi's method to FEM varies between 1.5 and 4. In cases, in which the failure surfaces obtained from both methods differ, this ratio can rise up to 11.

© 2023 DPU All rights reserved.

Keywords: Ultimate bearing capacity; Rock masses; Finite element method; Limit equilibrium method.

1. Introduction

Most rock masses are excellent foundation materials due to their bearing capacities of MPa. However, in the design of structures with high foundation loads such as high-rise buildings, dams and viaduct piers, the ultimate bearing capacities of rock masses should be accurately estimated.

* Corresponding author. Tel.: +491634677933

E-mail address: serdar.koltuk@campus.tu-berlin.de

<https://www.tu.berlin/ingenieurgeologie/ueber-uns>

Because of discontinuities existing in rock masses, the estimation of their bearing capacities is more difficult than soils. Various methods can be found in the literature to estimate the ultimate bearing capacity of shallow footings built on rock masses. These can be grouped under 4 main groups: Limit equilibrium, Slip-line, Limit analysis and Numerical methods [1-6]. In addition, the bearing capacities of certain rock types can be estimated empirically with the help of diagrams developed depending on the unconfined compressive strengths of intact rocks and the widths of openings existing in rock masses [7].

Terzaghi's limit equilibrium method [1] with the failure criterion of Mohr-Coulomb is widely used by engineers working in construction practice to estimate the ultimate bearing capacity of shallow foundations. Bowles [8] pointed out that intact rock samples are used in the laboratory to determine the shear strength parameters (c and ϕ) so that they do not account for the effect of discontinuities existing in rock masses. Therefore, he suggested that ultimate bearing capacities calculated according to the Terzaghi's approach should be reduced.

It is known that the shear strengths of most rocks are significantly affected by stress levels. Furthermore, rocks have significant tensile strengths compared to soils. Among the nonlinear failure criteria in the literature, the criterion of Hoek and Brown reliably simulates the deformation behavior of isotropic rocks [9,10]. Miranda et al. [11] combined the limit equilibrium method suggested by Wyllie [2] with the Hoek & Brown failure criterion.

The numerical methods have become widely used for solving complex engineering problems. In the last two decades, a growing use of numerical methods with the Hoek & Brown criterion have been observed to estimate the bearing capacities of rock masses.

In the recent studies, Javid, Fahimifar and Imani [12] investigated the effect of the interaction between two shallow strip footings on the ultimate bearing capacity using the Hoek & Brown criterion and two-dimensional numerical analyses. It has been seen that the ratio of the bearing capacity of a strip footing under the effect of a neighboring footing to the bearing capacity of the same isolated footing is about 1.3 to 1.6. Mansouri, Imani and Fahimifar [13] studied the ultimate bearing capacity of square- and rectangular-shaped footings with the help of the Hoek and Brown criterion and three-dimensional numerical analyses. It has been shown that the ultimate bearing capacities obtained from the 2D-analyses are higher than in the 3D-analyses. Shamloo and Imani [14] demonstrated that the effect of embedment depths on the bearing capacity of footings in rock masses cannot be taken into account correctly with the aid of equivalent surface loads. Using a self-developed adaptive finite element limit analysis code, Wu et al. [15] studied the ultimate bearing capacity of footings subjected to eccentric loads in rock masses with voids. To estimate the bearing capacity of strip footings on rock masses under three-dimensional effect, Chen, Zhu and Zhang [16] developed an analytical method using the failure criterion of Hoek & Brown. Ranjbarnia, Zarei and Goudarzy [17] introduced a probabilistic approach to estimate the bearing capacity of shallow foundations on rock masses. Das and Chakraborty [18] developed the design charts to estimate the bearing capacity of strip foundations with eccentric and inclined loads. Chen, Zhu and Zhang [19] demonstrated that ignoring the three-dimensional strength and the weight of rock mass would lead to the underestimation of the bearing capacity of rock masses.

In the present study, the ultimate bearing capacity of a strip footing on rock masses with different geotechnical properties was estimated using the finite element method (FEM) with the failure criterion of Hoek & Brown. The results of the FE-analyses are compared to those obtained from the limit equilibrium methods (LEM) suggested by Terzaghi [1] and Miranda et al. [11]. The present study shows engineers working in construction practice the limits of the use of limit equilibrium methods in determining the bearing capacity of rock masses.

2. Methodology

2.1. Numerical method

The software Plaxis-2D [20], which is based on the finite element method, is used to estimate the ultimate bearing capacity of a strip footing on rock masses.

Numerical model:

By utilizing the symmetry feature, only half of the numerical model shown in Figure 1 was created. Pre-analyses have shown that the size effect of the model on the numerical results can be ignored if the horizontal and vertical lengths of the model were not smaller than 12.5 m and 10 m, respectively. The lateral boundaries were fixed in the horizontal direction while the bottom boundary was fixed in the both directions. A rigid strip footing was modeled as a continuous load with a width of 0.5 m on the rock surface without an embedment depth. In order to model a rough footing, the horizontal movement of the footing was prevented. In the vertical direction, a deformation value leading to the ground failure was inputted. The ground water level was defined at the base of the model, and the moist unit weight of the rock masses was set to 24 kN/m³.

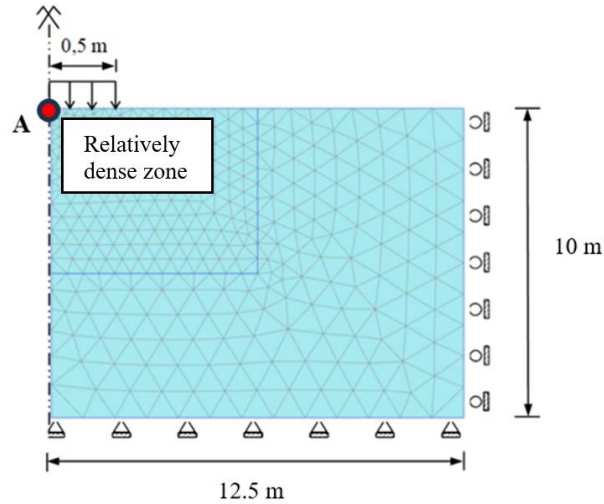


Fig. 1. FE-model in Plaxis 2D.

The effect of mesh density on the numerical results is well known. A dense mesh causes prolonged analysis time and an uneconomical solution while the use of a relatively loose mesh affects the accuracy of numerical results. In Plaxis-2D [20], the generation of FE-mesh is fully automated and a robust triangulation procedure. Pre-analyses have shown that a relatively dense mesh generated in the area of 5 m x 6 m under the foundation load in Fig. 1 allowed the numerical results to converge to a constant value. As a result, 942 triangular elements with 15 nodes and an average element size of 0.42 m were generated in the numerical model.

The analyses consisted of 2 stages. In the first step, the initial stress condition existing in the rock mass before the foundation load was reconstructed using the K0-procedure. The second stage was the plastic calculation stage, in which the reaction force of the rock mass corresponding to the deformation inputted in the vertical direction was determined. To determine the ultimate bearing capacity, the graph of the reaction force-deformation obtained for Point A was considered. In this graph, twice the maximum load that converges to a constant value, which was determined using the method of tangent intersection defined by Singh et al. [21], was assumed to be q_{ult} .

Constitutive model:

In Plaxis-2D [20], the stress-strain behaviors of rock masses were modeled using the criterion of Hoek and Brown. The empirical equation proposed by Hoek for intact rocks is expressed as follows [9,10]:

$$\sigma_1 = \sigma_3 + \sigma_{ci} \cdot \left(m_i \cdot \frac{\sigma_3}{\sigma_{ci}} + 1 \right)^{0.5} \quad (1)$$

where σ_{ci} is the unconfined compressive strength of intact rock, m_i is material constant for intact rock, which is determined experimentally. Depending on the rock type, it can take values between 2 and 35.

Later, the Hoek's criterion was developed by Brown for jointed rock masses, and it is called the Hoek-Brown failure criterion. Hoek-Brown [9,10] generalized failure criterion is expressed by Eq. (2):

$$\sigma_1 = \sigma_3 + \sigma_{ci} \cdot (m_b \cdot \frac{\sigma_3}{\sigma_{ci}} + s)^a \quad (2)$$

where the parameters “ m_b , s and a ” are the material constants for rock mass and can be calculated by using Eq. (3)-(5):

$$s = e^{\left(\frac{GSI-100}{9-3D}\right)} \quad (3)$$

$$a = 0,5 + 0,167 \cdot (e^{\frac{-GSI}{15}} - 0,0013) \quad (4)$$

$$m_b = m_i \cdot e^{\left(\frac{GSI-100}{28-14D}\right)} \quad (5)$$

where GSI is Geological Strength Index, D is disturbance factor.

The geological strength index of the rock mass is determined visually depending on the structure of rock masses and their surface properties. The GSI-values vary between 0 for rocks that have decomposed into the soil and 100 for intact rocks with unweathered surfaces. A chart for determining GSI -value is given by Hoek and Brown [10]. The value of the disturbance factor varies between 0 (for undisturbed rocks) and 1 (for disturbed rocks by excavations or explosion etc.).

The Hoek & Brown failure criterion should be used for intact rocks, rock masses with several discontinuities and heavily jointed rock masses (Group I and III) which have similar surface properties and can be considered isotropic [5]. Furthermore, this failure criterion was based on the brittle failures observed in triaxial tests on intact rocks. Therefore, it should not be used for principal stress levels at which ductile failure appears [9,10]. Furthermore, it should not be used if the size of the rock blocks is larger than the structure or they have same size or if one group of discontinuities existing in the rock mass is weaker than the others [5,12].

Eight parameters must be inputted in Plaxis-2D [20]. These parameters are: 1) Young's modulus of rock mass E_{rm} , 2) Poisson's ratio of rock mass ν , 3) unconfined compressive strength of intact rock σ_{ci} , 4) material constant for intact rock m_i , 5) geological strength index of rock mass GSI, 6) disturbance factor D, 7) dilation angle of rock mass φ_{max} for zero confining pressure, and 8) confining pressure σ_φ at the depth, where the dilation angle is equal to zero.

The Young's modulus E_{rm} can be calculated based on the GSI and D values with the help of the following simplified equation:

$$E_{rm}(MPa) = 10^5 \cdot \left(\frac{1-\frac{D}{2}}{1+e^{\left(\frac{75+25D-GSI}{11}\right)}} \right) \quad (6)$$

Poisson's ratio ν varies between 0.1 and 0.4 depending on the rock type. In this study, the value of ν was taken as 0.2. Since the foundation used in the numerical model has no embedment depth, there will be no excavation, and thus no disturbance occurs in the rock mass. Therefore, the value of D was taken as 0. In the analyses, the unconfined compressive strength σ_{ci} of intact rock was considered with 4 different values $\sigma_{ci} = 1, 5, 25, 100$ MPa, the material constant m_i of intact rock with 3 different values $m_i = 2.5, 10, 20$, and the geological strength index of rock mass GSI with 5 different values $GSI = 10, 30, 50, 70, 90$.

Rock masses show dilatation behaviour when they are subjected to shearing under low confining pressures. In this study, the value of ϕ_{\max} was set to the friction angle of the rock mass, and its value was decreased linearly to 0 at a certain depth (associated flow) [20]. The confining stress at the depth where the dilation angle is zero was set to $\sigma_{\phi=0} = 1000 \text{ kN/m}^2$ taking into account that the vertical length of the model and the unit weight of rock masses were 10 m and 24 kN/m^3 , respectively.

2.2. Limit equilibrium method

In the present study, the limit equilibrium methods suggested by Terzaghi [1] and Miranda et al. [11] are used to estimate the ultimate bearing capacities of rock masses.

Terzaghi's method:

Terzaghi [1,22] proposed the failure surface under a strip foundation whose width B is very small compared to its length (see Figure 2a). This failure surface consists of 3 zones: 1) a triangular zone ACD, 2) radial shear zones AFD and CDE with logarithmic spirals DE and DF, 3) Rankine triangular passive zones AFH and CEG. The ultimate bearing capacity q_{ult} is obtained by considering the equilibrium of the triangular wedge ACD and using the failure criterion of Mohr-Coulomb as follow:

$$q_{ult} = c \cdot N_c + \gamma \cdot D_f \cdot N_q + 0,5 \cdot \gamma \cdot B \cdot N_{\gamma} \quad (7)$$

where c is cohesion, γ is unit weight of rock mass, D_f is foundation depth, B is foundation width, N_c , N_q and N_{γ} are bearing capacity factors that can be calculated by using Eqs. (8)-(10) depending on the friction angle ϕ of the rock mass:

$$N_c = 5 \cdot \tan^4 \left(45 + \frac{\phi}{2} \right) \quad (8)$$

$$N_q = \tan^6 \left(45 + \frac{\phi}{2} \right) \quad (9)$$

$$N_{\gamma} = N_q + 1 \quad (10)$$

Miranda's method:

Miranda et al. [11] combined Wyllie's limit equilibrium method [2] with the Hoek & Brown failure criterion. In Figure 2b, the wedge A is the active wedge and the wedge B is the passive wedge, which are representing the failure zones under a strip footing. In this method, the weights of the rock masses and the shear stresses that develop at the interface of both wedges are neglected. The wedges A and B are assumed to be in compression as in triaxial shear tests.

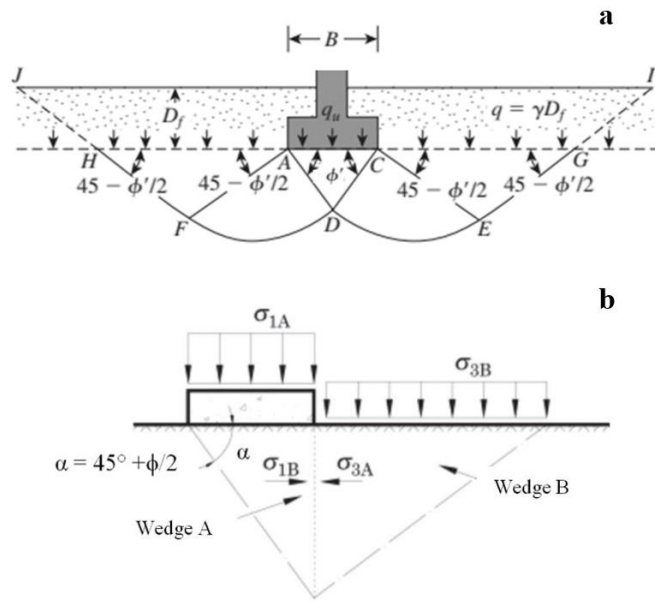


Fig. 2. Ultimate bearing capacity of rock masses according to (a) Terzaghi [22]; (b) Wyllie [2].

Assuming that there is no load on the rock surface outside the foundation area, the major and minor principal stresses in the wedge B will act in the horizontal direction (σ_{1B}) and in the vertical direction ($\sigma_{3B} = 0$), respectively. When the wedge A collapses, the minor principal stress σ_{3A} acting on the wedge A will be equal to the major principal stress σ_{1B} in the wedge B. Therefore, the major principal stress σ_{1A} in the wedge A will correspond to the ultimate bearing capacity q_{ult} of the footing.

In the Miranda's method, by substituting $\sigma_3 = \sigma_{3B} = 0$ in Eq. (2), the following equation is firstly obtained for the wedge B:

$$\sigma_{1B} = \sigma_{ci} \cdot s^a \quad (11)$$

Then, substituting $\sigma_3 = \sigma_{3A} = \sigma_{ci} \cdot s^a$ in Eq. (2), the following equation is obtained for the wedge A, which gives the ultimate bearing capacity q_{ult} :

$$q_{ult} = \sigma_{1A} = \sigma_{ci} \cdot [s^a + (m_b \cdot s^a + s)^a] \quad (12)$$

3. Results and discussion

3.1. Comparison of the failure surfaces obtained from the FEM and LEM

The numerical analyses carried out in this study gave three different failure surfaces, as shown in Fig. 3. These failure surfaces were called failure type A, B, and C. A summary of the failure types observed in the FE-analyses with associated flow is given in Table 1 depending on the Hoek & Brown parameters.

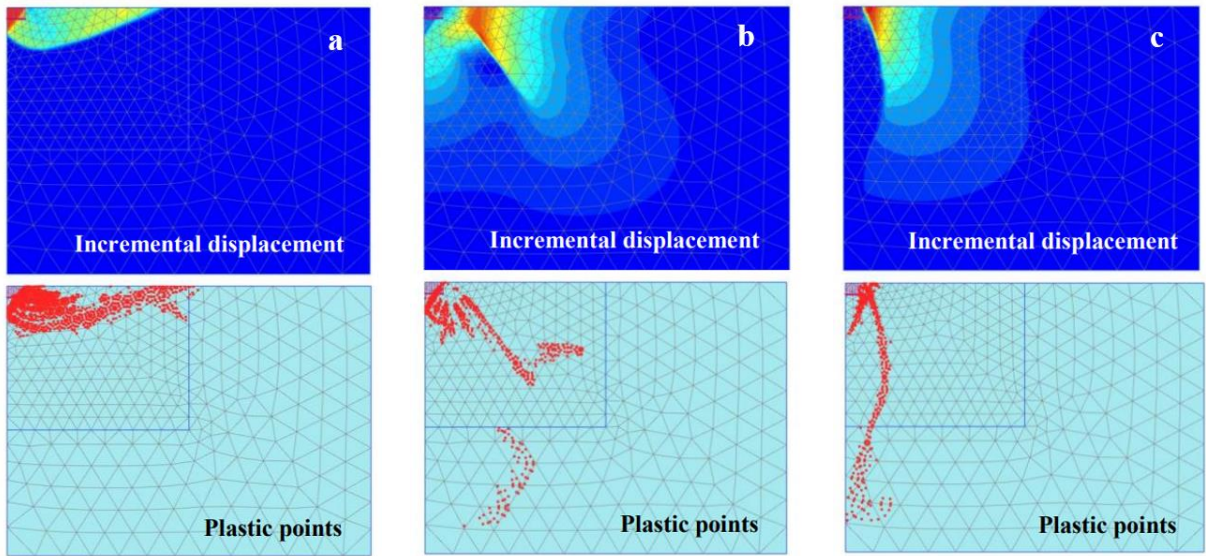


Fig. 3. Failure surfaces obtained from the FE-analyses: (a) failure type A, (b) failure type B, (c) failure type C.

A total of 60 analyses were performed, and the most of these analyses gave a failure zone shown in Fig. 3a (Failure type A). This failure zone corresponds to the general shear failure suggested by Terzaghi in Fig. 2a.

Only in 3 analyses ($\sigma_{ci} = 25$ MPa, $m_i = 10$, GSI = 30; $\sigma_{ci} = 25$ MPa, $m_i = 10$, GSI = 90 and $\sigma_{ci} = 100$ MPa, $m_i = 20$, GSI = 10), the Rankine triangular passive zone did not develop while the triangular zone and radial shear zone developed. The failure type B shown in Fig. 3b is a local shear failure.

In 12 analyses ($\sigma_{ci}=100$ MPa, $m_i = 10$, GSI = 10 – 90; $\sigma_{ci} = 100$ MPa, $m_i = 20$, GSI = 30 – 90; $\sigma_{ci} = 25$ MPa, $m_i = 20$, GSI = 70 – 90 and $\sigma_{ci} = 25$ MPa, $m_i = 10$, GSI = 50), the failure surface shown in Figure 3c appeared. Compared to the failure type A, a triangular wedge under the foundation appeared more deeply while the other zones of the Terzaghi's failure surface did not develop (punching failure). On the other hand, a new failure surface developed in vertical direction towards the inner part of the rock mass.

When the failure surfaces obtained from the FEM are examined, it is seen that the failure mechanism proposed by Wyllie [2,11] in Fig. 2b does not develop. Therefore, only the q_{ult} - values calculated according to the Terzaghi's approach were used for comparison in Figure 4.

Table 1. Failure types depending on the Hoek & Brown parameters.

Rock parameters		GSI (-)				
σ_{ci} (MPa)	m_i (-)	10	30	50	70	90
1	2.5	A	A	A	A	A
	10	A	A	A	A	A
	20	A	A	A	A	A
5	2.5	A	A	A	A	A
	10	A	A	A	A	A
	20	A	A	A	A	A
25	2.5	A	A	A	A	A
	10	A	B	C	A	B
	20	A	A	A	C	C
100	2.5	A	A	A	A	A
	10	C	C	C	C	C

20	B	C	C	C	C
----	----------	----------	----------	----------	----------

3.2. Comparison of the ultimate bearing capacities obtained from the FEM and LEM

In order to make a comparison between the results of the FEM and the Terzaghi's method, the equivalent parameters of the Mohr-Coulomb criterion (c and ϕ) corresponding to the Hoek-Brown parameters used in the FE-analyses should be determined. The parameters of Mohr-Coulomb are calculated by plotting a linear envelope on the nonlinear failure envelope of Hoek and Brown that provides the best fit for a given stress range.

Table 2. Equivalent Mohr-Coulomb parameters for $m_i = 2.5$.

GSI (-)	σ_{ci} (MPa)							
	1		5		25		100	
	c (MPa)	ϕ (°)	c (MPa)	ϕ (°)	c (MPa)	ϕ (°)	c (MPa)	ϕ (°)
10	0.01	9.5	0.05	9.5	0.25	9.5	1.00	9.5
30	0.02	15.0	0.11	15.0	0.52	15.0	2.09	15.0
50	0.03	19.5	0.17	19.5	0.83	19.5	3.34	19.5
70	0.06	24.0	0.31	24.0	1.54	24.0	6.16	24.0
90	0.17	26.5	0.83	26.5	4.14	26.5	16.57	26.5

Table 3. Equivalent Mohr-Coulomb parameters for $m_i = 10$.

GSI (-)	σ_{ci} (MPa)							
	1		5		25		100	
	c (MPa)	ϕ (°)	c (MPa)	ϕ (°)	c (MPa)	ϕ (°)	c (MPa)	ϕ (°)
10	0.02	18.0	0.10	18.0	0.48	18.0	1.92	18.0
30	0.04	24.5	0.18	24.5	0.88	24.5	3.52	24.5
50	0.05	30.5	0.25	30.5	1.25	30.5	4.98	30.5
70	0.07	36.5	0.36	36.5	1.78	36.5	7.11	36.5
90	0.13	41.5	0.64	41.5	3.21	41.5	12.83	41.5

Table 4. Equivalent Mohr-Coulomb parameters for $m_i = 20$.

GSI (-)	σ_{ci} (MPa)							
	1		5		25		100	
	c (MPa)	ϕ (°)	c (MPa)	ϕ (°)	c (MPa)	ϕ (°)	c (MPa)	ϕ (°)
10	0.03	23.5	0.13	23.5	0.65	23.5	2.59	23.5
30	0.05	30.5	0.23	30.5	1.12	30.5	4.49	30.5
50	0.06	36.5	0.31	36.5	1.54	36.5	6.14	36.5
70	0.08	42.5	0.41	42.5	2.05	42.5	8.22	42.5
90	0.13	48.0	0.63	48.0	3.16	48.0	12.65	48.0

In this study, the equivalent Mohr-Coulomb parameters were calculated using the Software RocData [23]. Here, the "General" option was selected as the failure envelope interval, and the equivalent Mohr-Coulomb parameters were determined for the stress range of $0 < \sigma_3 < 0.25 \cdot \sigma_{ci}$, in which a brittle failure appears. The equivalent parameters of Mohr-Coulomb (c and ϕ) corresponding to the parameters of Hoek-Brown are given in Tables 2-4.

In Tables 5-7, the ultimate bearing capacities obtained from the finite element method (FEM) and limit equilibrium methods (LEM) are given depending on the various rock mass properties. As it can be seen from Tables

5-7, Terzaghi's approach [1] gives the highest values of ultimate bearing capacity while the lowest values are obtained from the Miranda's method. Only in 3 analyses ($\sigma_{ci} = 25$ MPa, $m_i = 20$, GSI = 90; $\sigma_{ci} = 100$ MPa, $m_i = 20$, GSI = 90 and $\sigma_{ci} = 100$ MPa, $m_i = 20$, GSI = 70), the q_{ult} -values calculated according to Miranda's method are negligibly larger than the values in the FEM.

Table 5. Ultimate bearing capacities obtained from the FEM and LEM for $m_i = 2.5$.

GSI (-)	σ_{ci} (MPa)											
	1			5			25			100		
	Ultimate Bearing Capacity q_{ult} (MPa)											
	FEM	Miranda et al.	Terzaghi	FEM	Miranda et al.	Terzaghi	FEM	Miranda et al.	Terzaghi	FEM	Miranda et al.	Terzaghi
10	0.07	0.01	0.14	0.24	0.05	0.54	0.85	0.25	2.51	2.70	1	9.78
30	0.25	0.07	0.37	0.93	0.35	1.58	3.95	1.75	7.58	14.65	7	29.94
50	0.53	0.23	0.78	2.35	1.15	3.50	10.80	5.75	17.01	42.38	23	64.51
70	1.20	0.63	1.93	5.70	3.15	8.85	27.75	15.75	43.51	110.75	63	173.22
90	2.95	1.73	5.87	14.40	8.65	28.21	71.45	43.25	139.94	286.00	173	559.39

Table 6. Ultimate bearing capacities obtained from the FEM and LEM for $m_i = 10$.

GSI (-)	σ_{ci} (MPa)											
	1			5			25			100		
	Ultimate Bearing Capacity q_{ult} (MPa)											
	FEM	Miranda et al.	Terzaghi	FEM	Miranda et al.	Terzaghi	FEM	Miranda et al.	Terzaghi	FEM	Miranda et al.	Terzaghi
10	0.28	0.02	0.43	0.84	0.10	1.81	2.85	0.50	8.74	3.30	2	34.54
30	0.75	0.13	1.21	2.65	0.65	5.38	8.63	3.25	26.13	25.02	13	103.87
50	1.40	0.38	2.72	5.74	1.90	12.10	20.20	9.50	58.85	73.47	38	234.11
70	2.80	1.01	6.21	12.21	5.05	27.92	54.85	25.25	136.24	140.38	101	542.61
90	6.05	2.66	16.73	28.60	13.30	77.88	120.11	66.50	384.26	373.10	266	1532.15

Table 7. Ultimate bearing capacities obtained from the FEM and LEM for $m_i = 20$.

GSI (-)	σ_{ci} (MPa)											
	1			5			25			100		
	Ultimate Bearing Capacity q_{ult} (MPa)											
	FEM	Miranda et al.	Terzaghi	FEM	Miranda et al.	Terzaghi	FEM	Miranda et al.	Terzaghi	FEM	Miranda et al.	Terzaghi
10	0.55	0.03	0.88	1.81	0.15	3.75	5.35	0.75	17.98	12.45	3	71.0
30	1.34	0.17	2.45	4.70	0.85	10.88	18.15	4.25	52.74	31.05	17	209.21
50	2.61	0.51	5.48	9.80	2.55	24.55	34.20	12.75	119.81	55.12	51	476.45
70	4.65	1.34	12.51	20.10	6.70	56.21	40.89	33.50	274.42	127.14	134*	1092.52
90	9.10	3.47	32.32	42.50	17.35	146.65	81.79	86.75*	718.54	386.06	347*	2863.07

In the most FE-analyses, the vertical deformations occurring in the rock masses under the loads corresponding to

the ultimate bearing capacity of the footing were smaller than 2 cm. Only in 12 analyses ($\sigma_{ci} = 100$ MPa, $m_i = 2.5$, GSI = 70-90; $\sigma_{ci} = 100$ MPa, $m_i = 2.5-10$, GSI = 50; $\sigma_{ci} = 100$ MPa, $m_i = 2.5-20$, GSI = 30; $\sigma_{ci} = 100$ MPa, $m_i = 2.5-10$, GSI = 10; $\sigma_{ci} = 25$ MPa, $m_i = 10-20$, GSI = 10 and $\sigma_{ci} = 25$ MPa, $m_i = 20$, GSI = 30), the vertical deformations varied between 2 cm – 9 cm.

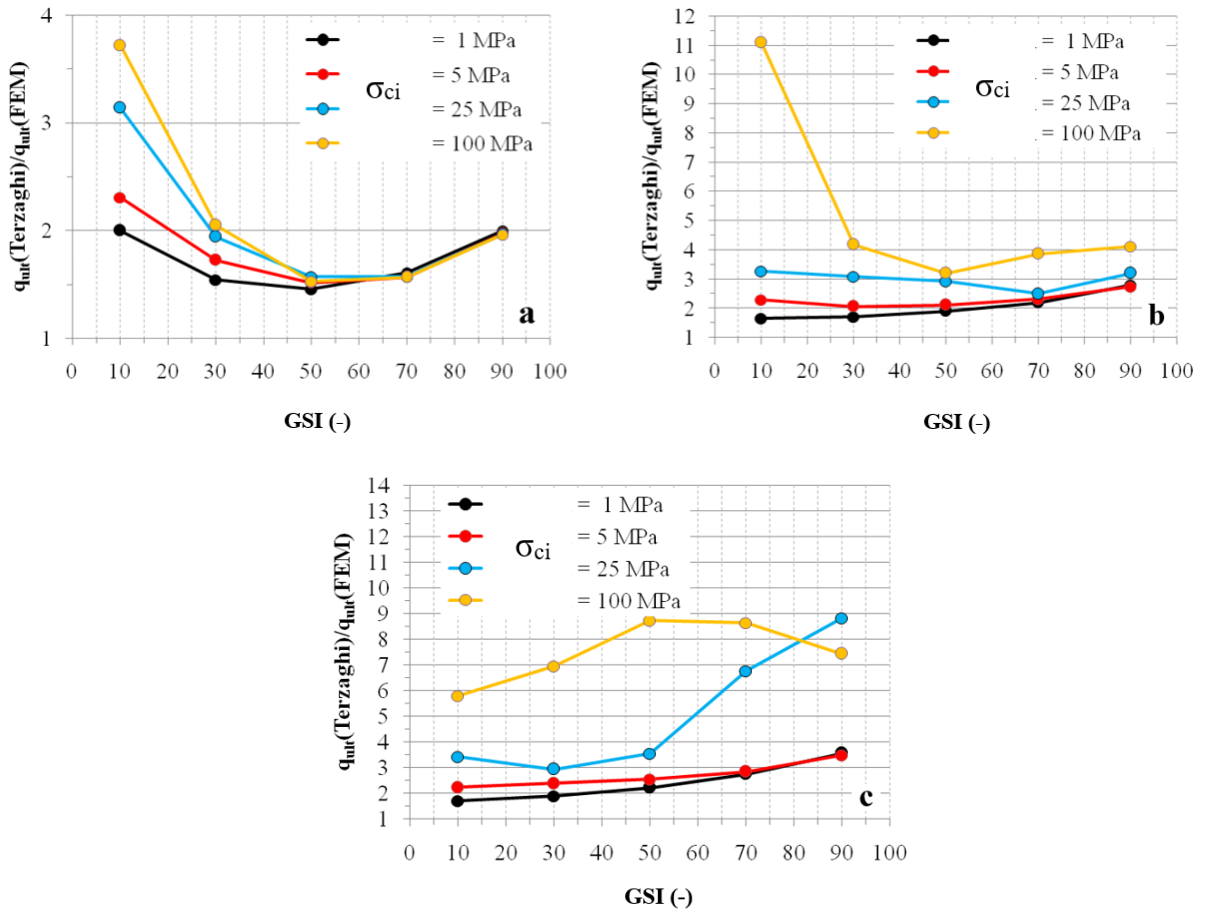


Fig. 4. Comparison of the ultimate bearing capacities obtained from the FEM and Terzaghi's method: a) $m_i = 2.5$, b) $m_i = 10$, c) $m_i = 20$.

In cases in which the unconfined compressive strength of the intact rock σ_{ci} is relatively low ($\sigma_{ci} \leq 5$ MPa with $m_i = 2.5-20$ and $\sigma_{ci} \leq 25$ MPa with $m_i < 20$), the ratio of the ultimate bearing capacity calculated according to the Terzaghi's approach $q_{ult}(Terzaghi)$ to those in the FEM $q_{ult}(FEM)$ varies between 1.5 and 4. In cases in which the value of σ_{ci} is higher ($\sigma_{ci} = 25$ MPa with $m_i = 20$ and GSI > 50 as well as $\sigma_{ci} > 25$ MPa with $m_i > 2.5$), the ratio of $q_{ult}(Terzaghi)/q_{ult}(FEM)$ varies between 3 and 11 (Fig. 4). This increase can be explained by the developing of more different failure surfaces shown in Figures 3b and 3c than the general shear failure suggested by Terzaghi in Fig. 2a.

4. Conclusions

In the present study, the ultimate bearing capacity of a strip footing constructed on rock masses with different material properties ($\sigma_{ci} = 1$ to 100 MPa, $m_i = 2.5$ to 20 and GSI = 10 to 90) are estimated using the finite element

method (FEM) with Hoek & Brown failure criterion. The results of the FE-analyses are compared to those determined according to the limit equilibrium method (LEM) suggested by Terzaghi [1] and Miranda et al. [11]. The analyses give the following results:

- The Terzaghi's approach gives the highest values of ultimate bearing capacities while the lowest values are obtained from the approach of Miranda;
- In most cases, especially in cases in which the unconfined compressive strength of the intact rock is relatively low ($\sigma_{ci} \leq 5$ MPa with $m_i = 2.5-20$, $\sigma_{ci} \leq 25$ MPa with $m_i < 20$), the FEM and the Terzaghi's method give the similar failure surfaces. In these cases, the ratio of the ultimate bearing capacities obtained from the Terzaghi's method q_{ult} (Terzaghi) to those in the finite element method q_{ult} (FEM) varies between 1.5 and 4.
- In cases in which the value of σ_{ci} is relatively high ($\sigma_{ci} = 25$ MPa with $m_i = 20$ and $GSI > 50$ as well as $\sigma_{ci} > 25$ MPa with $m_i > 2.5$), the ratio of $q_{ult(Terzaghi)}/q_{ult(FEM)}$ can rise up to 11;
- The relatively high ratios of $q_{ult(Terzaghi)}/q_{ult(FEM)}$ appear especially for rock masses with high unconfined compressive strengths ($\sigma_{ci} \geq 25$ MPa), which can be explained that the rocks with relatively low compressive strengths behave like soils.

Finally, it should be mentioned that the loads corresponding to the ultimate bearing capacity of footings can lead to high deformations, which may be not allowed. The most important advantage of the FEM compared to the LEM is that it enables the estimation of deformations of rock masses under foundation loads. However, the correct estimation of the dilatant behaviour of rock masses has an essential role in accuracy of the FE-results. Thus, sensitivity analyses are necessary with respect to the value of maximum dilatancy angle on rock surface and its variation with depth.

Acknowledgements

This research received no external funding, and there is no conflict of interest with any person or institution.

References

- [1] K. Terzaghi, *Theoretical soil mechanics*. New York, USA: John Wiley & Sons, 1943.
- [2] D.C. Wyllie, *Foundations on Rock*. London: E & FN Spon and USA and Canada: Routledge, 1999.
- [3] A. Serrano, C. Ollala, and J. Gonzalez, "Ultimate bearing capacity of rock masses based on the modified Hoek Brown criterion," *International Journal of Rock Mechanics & Mining Sciences*, vol. 37, no. 6, pp. 1013–1018, September 2000, doi:10.1016/S1365-1609(00)00028-9.
- [4] X. Yang and J.H. Yin, "Upper bound solution for ultimate bearing capacity with a modified Hoek-Brown failure Criterion," *International Journal of Rock Mechanics & Mining Sciences*, vol. 42, no. 4, pp. 550–560, June 2005, doi: 10.1016/j.ijrmms.2005.03.002.
- [5] R.S. Merifeld, A.V. Lyamin, and S.W. Sloan, "Limit analysis solutions for the bearing capacity of rock masses using the generalised Hoek-Brown criterion," *International Journal of Rock Mechanics & Mining Sciences*, vol. 43, no. 6, pp. 920–937, September 2006, doi: 10.1016/j.ijrmms.2006.02.001.
- [6] Z. Saada, S. Maghous, and D. Garnier, "Bearing capacity of shallow foundations on rocks obeying a modified Hoek–Brown failure criterion," *Computers and Geotechnics*, vol. 35, no. 2, pp. 144–154, March 2008, doi: 10.1016/j.compgeo.2007.06.003.
- [7] Eurocode 7, *Geotechnical Design–Part 1: General Rules*. CEN, Brussels: 2004.
- [8] J.E. Bowles, *Foundation Analysis and Design*. Singapore: McGraw Hill, 1996.
- [9] E. Hook, C. Carranza-Torres, and B. Corkun, "Hoek-Brown failure criterion-2002 edition," in *Proc. of the North American rock mechanics society meeting*, 2002, October 267-273, 2002. [Online]. Available: <https://www.rocsience.com/assets/resources/learning/hoek/Hoek-Brown-Failure-Criterion-2002.pdf>.
- [10] E. Hoek and E.T. Brown, "The Hoek-Brown failure criterion and GSI-2018 edition," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 11, no. 3, pp. 445–463, June 2019, doi: 10.1016/j.jrmge.2018.08.001.
- [11] T. Miranda, F. Martins, and N. Araujo, "Design of spread foundations on rock masses according to Eurocode 7," in *12th ISRM Congress*, 2011, pp. 1959–1962, doi: 10.1201/b11646-373.
- [12] A.H. Javid, A. Fahimifar, and M. Imani, "Numerical investigation on the bearing capacity of two interfering

- strip footings resting on a rock mass,” *Computers and Geotechnics*, vol. 69, pp. 514-528, September 2015, doi: 10.1016/j.compgeo.2015.06.005.
- [13] M. Mansouri, M. Imani, and A. Fahimifar, “Ultimate bearing capacity of rock masses under square and rectangular footings,” *Computers and Geotechnics*, 111, pp. 1-9, July 2019, doi: 10.1016/j.compgeo.2019.03.002.
- [14] S. Shamloo and M. Imani, “Upper bound solution for the bearing capacity of rock masses considering the embedment depth,” *Ocean Engineering*, vol. 218, no. 6, December 2020, doi: 10.1016/j.oceaneng.2020.108169.
- [15] G. Wu, M. Zhao, R. Zhang, and G. Liang, “Ultimate bearing capacity of eccentrically loaded strip footings above voids in rock masses,” *Computers and Geotechnics*, vol. 218, December 2020, doi: 10.1016/j.compgeo.2020.103819.
- [16] H. Chen, H. Zhu, and L. Zhang, “An analytical approach to the ultimate bearing capacity of smooth and rough strip foundations on rock mass considering three-dimensional (3D) strength,” *Computers and Geotechnics*, vol. 149, no. 3, September 2022, doi: 10.1016/j.compgeo.2022.104865.
- [17] M. Ranjbarnia, F. Zarei, and M. Goudarzy, “Probabilistic Analysis of Bearing Capacity of Square and Strip Foundations on Rock Mass by the Response Surface Methodology,” *Rock Mechanics and Rock Engineering*, 56, pp. 343–362, vol. 56, October 2022, doi: 10.1007/s00603-022-03090-5.
- [18] S. Das and D. Chakraborty, “Effect of eccentric and inclined loading on the bearing capacity of strip footing placed on rock mass,” *Journal of Mountain Science*, vol. 21, pp. 292–312, January 2024, doi: 10.1007/s11629-023-8312-2.
- [19] H. Chen, H. Zhu, and L. Zhang, “Semi-analytical solution for ultimate bearing capacity of smooth and rough circular foundations on rock considering three-dimensional strength,” *International Journal for Numerical and Analytical Methods in Geomechanics*, February 2024, doi: 10.1002/nag.3699.
- [20] Plaxis 2D, Delft: Plaxisbv, 2019.
- [21] M. Singh, M.N. Viladkar, P.S. Shekhawat, K. Tripathi, and M. Amin, “Bearing capacity of strip footings on jointed rock mass,” *Arabian Journal of Geosciences*, vol. 15, September 2022, doi: 10.1007/s12517-022-10841-9.
- [22] B.M. Das, *Principles of Foundation Engineering*. Boston, USA: Cengage Learning, 2014.
- [23] RocData, version 5.0, Toronto: Rocscience Inc., 2019.



RESEARCH ARTICLE

Receive Date: 06.12.2023

Accepted Date: 20.03.2024

Synthesis, characterization, anti-microbial activity studies of salicylic acid and 2-aminopyridine derivatives salts and their Cu(II) complexes

Halil İlkimen^{a,*}, Aysel Gülbandır^b

^aDepartment of Chemistry, Faculty of Art and Sciences, Kütahya Dumlupınar University, 43100 Kütahya, Türkiye, ORCID: 0000-0003-1747-159X

^bDepartment of Food Engineering, Faculty of Agriculture, Eskişehir Osmangazi University, 26000 Eskişehir, Türkiye, ORCID: 0000-0001-9075-9923

Abstract

Four salts (**1-4**) obtained between salicylic acid (H_2salic) and 2-amino-Xpyridine {X = (2ap), 3-methyl (2a3mp), 4-methyl (2a4mp) and 5-methyl (2a5mp)} and the Cu(II) complex of H_2salic (**5**) by methods available in the literature and new Cu(II) complexes (**6-9**) of the salts (**1-4**) has been prepared. The Cu(II) complexes (**6-9**) were suggested by elemental analysis, FT-IR, AAS, UV-Vis and magnetic susceptibility techniques. The spectroscopic research results indicated that complexes **6-9** have tetrahedral geometries. Additionally, antimicrobial activities of free ligands (H_2salic , 2ap, 2a3mp, 2a4mp and 2a5mp), **1-9** were studied against *Candida albicans* (F89) yeast, *Staphylococcus aureus* (NRRL B-767), *Pseudomonas aeruginosa* (ATCC 27853), *Bacillus subtilis*, *Listeria monocytogenes* (ATCC 7644), *Escherichia coli* (ATCC25922) and *Enterococcus faecalis* (ATCC 29212) bacteria. The results were compared with the control compounds (Fluconazole, Vancomycin, Cefepime and Levofloxacin). All compounds showed activity against bacteria and yeasts.

© 2023 DPU All rights reserved.

Keywords: 2-Aminopyridine, salicylic acid, salt, Cu(II) complex, anti-bacterial and anti-fungal activities.

1. Introduction

2-Aminopyridines have garnered particular interest due to their diverse pharmacological properties linked to their inclusion in certain compounds. Studies have demonstrated that the presence of a tiny 2-aminopyridine molecule enhances the target molecule's therapeutic qualities, regardless of how complex the molecule is more heterocycles present in its structure or a simple molecule with a few groups on it. Numerous medications, including piroxicam, tenoxicam, sulfasalazine with anti-inflammatory qualities, delavirdine as an anti-HIV medication, sulfapyridine as an antibacterial medication, and tripelenamine as an antihistaminic medication, are currently available on the market and include traces of 2-aminopyridine. The antitumoral, anti-alzheimer, antidiabetic, antimicrobial, antiviral, analgesic, anti-inflammatory, antiparasitic, antimalarial, antihistamine, anticonvulsant, Renin, n-NOS, CXCR1/2, JNK1, PKC, and Syk inhibitors have thus been shown to be present in both simple and complex compounds

containing grafted 2-aminopyridine moiety [1].

Many studies are carried out with salicylic acid (H_2salic) and its derivatives with electron-donating oxygen atoms ($-COOH$ and OH) and its protonated forms ($Hsalic^-$ and $salic^{2-}$). Salicylic acid are one of the well-known hydroxybenzoic acids that have antimicrobial, anti-inflammatory, anti-cancer, anti-tumor, anti-proliferative, anti-viral and analgesic properties [2]. Salicylic acid can bind to metal ion from both carboxylic and hydroxyl group in various modes such as monodentate, bidentate, tridentate, pentadentate and bridging [3-10]. In addition, copper(II), manganese(II) and zinc(II) complexes of salicylic acid and its derivatives have potential for treating cancer [11,12]. Proton transfer salts and metal complexes containing salicylic acid and organic bases have been synthesized in the literature [13-18]. The salts of 2-aminopyridine derivatives and salicylic acid have been synthesized, but the metal complex has not been synthesized [19-27].

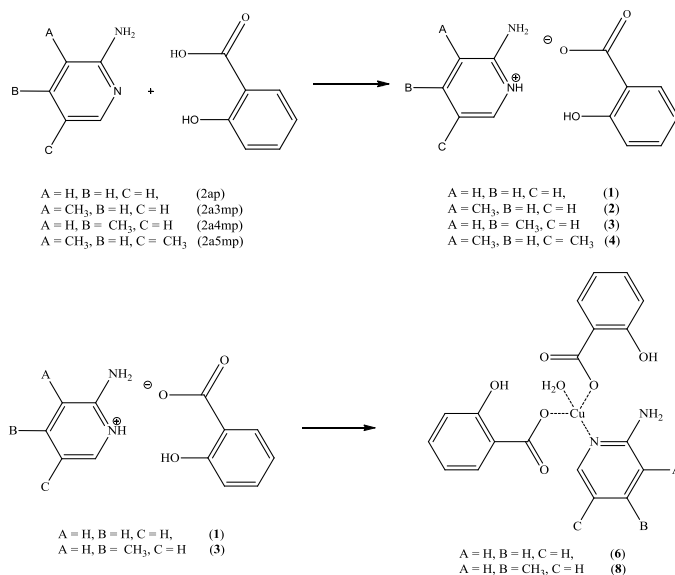
2. Experimental

2.1. Preparation of 1-4 and 6-9.

The Cu(II) complex of H_2salic ($[Cu(Hsalic)_2(H_2O)]$, **5**) was obtained by the method available in the literature [11].

5 mmol H_2salic (2.3123 g) and 5 mmol 2-aminopyridine (**1** for **2ap**, **2** for **2a3mp**, **3** for **2a4mp** and **4** for **2a5mp**) dissolved in 100 mL of ethanol. The white amorphous solids were procured by stirring for three days (70% yield for **1**, 80% yield for **2**, 75% yield for **3** and 80% yield for **4**) (Fig. 1).

5 mmol Copper(II) acetate monohydrate and 5 mmol salt **{1** for **6**, **2** for **7**, **3** for **8** and **4** for **9**) was dissolved in ethanol:water solution (2:1) (75 mL) with stirring one week. Green amorphous solids (75% yield for **6**, 65% yield for **7**, 70% yield for **8** and 60% yield for **9**) were obtained from the mixtures (Fig. 1).



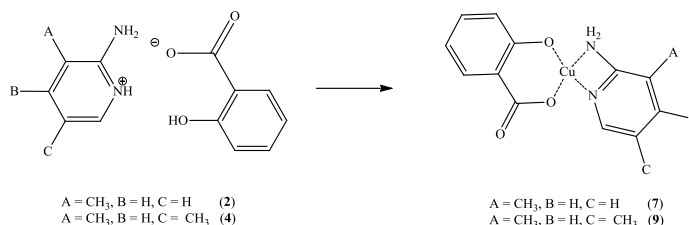


Fig. 1. The structures of compounds **1-4** and **6-9**.

2.2. Antimicrobial study

The antimicrobial properties of the substances were evaluated using a microbroth dilution susceptibility test. Dimethyl sulfoxide was used in the preparation of stock solutions. 4 mg of all compounds were taken and each dissolved in 2 mL of dimethyl sulfoxide. 10^8 Colony Forming Units/mL in double-strength Mueller-Hinton broth. Subsequently, 100 μ L of each microbial suspension was added to the wells. A control well without any microorganisms was included for comparison. The growth medium and sterile distilled water served as positive controls. Following 18-24 hours of incubation at 37 °C, the well displaying no turbidity first was identified as the Minimum Inhibitory Concentration (MIC).

3. Results and discussion

3.1. Elemental analysis and AAS results

According to the elemental analysis and AAS results of **6-9**, the metal:H₂salic:aminopyridine ratio was found to be 1:1:2 for **6** and **8** and 1:1:1 for **7** and **9** (Table 1).

Table 1. Elemental analysis and ICP-OES results of the studied substances.

Compound	Formula	Found% Anal. Cald.%			
		C	H	N	Cu ²⁺
1	C ₁₂ H ₁₂ N ₂ O ₃	62.10(62.06)	5.20(5.21)	12.10(12.06)	-
2	C ₁₃ H ₁₄ N ₂ O ₃	63.45(63.40)	5.70(5.73)	11.41(11.38)	-
3	C ₁₃ H ₁₄ N ₂ O ₃	63.42(63.40)	5.71(5.73)	11.39(11.38)	-
4	C ₁₃ H ₁₄ N ₂ O ₃	63.44(63.40)	5.75(5.73)	11.35(11.38)	-
6	C ₁₉ H ₁₈ CuN ₂ O ₇	50.75(50.72)	4.05(4.03)	6.20(6.23)	14.10(14.12)
7	C ₁₃ H ₁₂ CuN ₂ O ₃	50.75(50.73)	3.90(3.93)	9.15(9.10)	20.60(20.65)
8	C ₂₀ H ₂₀ CuN ₂ O ₇	50.76(51.78)	4.30(4.35)	6.00(6.04)	13.75(13.70)
9	C ₁₃ H ₁₂ CuN ₂ O ₃	50.78(50.73)	3.96(3.93)	9.08(9.10)	20.60(20.65)

3.2. Thermal analyses of **6-9**.

TG-DTG and DTA curves and values of **6-9** are given in Figs 3-5, respectively, and Table 2. Results of thermal analyses are similar to Cu(II) complexes of salicylic acid in the literature [28].

Compounds **6**, **7** and **9** thermally decomposed in two steps. The endothermic first stage corresponds to the loss of H₂O, C₃H₃ and 2a5mp units and the exothermic second stage corresponds to the loss of the 2Hsalic+2ap,

C₁₃H₁₂N₂O₃ and Hsalic units, respectively.

Compound **8** thermally decomposed in three steps. The endothermic first stage corresponds to the loss of one mole of water molecules. The endothermic second stage is consistent with the loss of 2a4mp. The exothermic third stage shows the loss of Hsalic. The final product left undecomposed is CuO for **6-9**.

3.3. FT-IR results

The IR data of **1-9** are given in Table 3. There are broad vibration peaks between 3500 and 3437 cm⁻¹ attributed to the ν(O-H) vibrations of **1-5**, **6** and **8**. Bands appearing at 3285 and 3202 cm⁻¹ for **1**, 3377 and 3325 cm⁻¹ for **2**, 3316 and 3294 cm⁻¹ for **3**, 3320 and 3293 cm⁻¹ for **4**, 3401 and 3341 cm⁻¹ for **6**, 3405 and 3320 cm⁻¹ for **7**, 3322 and 3291 cm⁻¹ for **8** and 3420 and 3330 cm⁻¹ for **9**, are assigned to NH₂ vibrations. The ν(N⁺-H) peaks observed in the range 3529-2745 cm⁻¹ for **1-4** were not observed in the complexes (**6-9**). The difference (Δν) between the extensions of the asymmetric/symmetric vibrations of the COO⁻ group shows how it coordinates to the metal ion. The differences of (**5-9**) were calculated 199 (1601 and 1437 cm⁻¹), 208 (1674 and 1466 cm⁻¹), 213 (1662 and 1449 cm⁻¹), 219 (1667 and 1448 cm⁻¹) and 200 (1636 and 1436 cm⁻¹), respectively. These results indicate that the carboxylate group is monodentate bound to the metal ion [29]. The peaks at the range of **1-9**, 3046-3104 cm⁻¹, 2783-2981 cm⁻¹, 1408-1647 cm⁻¹, 1075-1390 cm⁻¹, 748-757 cm⁻¹, 533-593 cm⁻¹ and 426-450 cm⁻¹ are assigned ν(C-H)_{ar.}, ν(C-H)_{alp.}, ν(C=N)/ν(C=C) (except **5**), ν(C-O), ν(py) (except **5**), ν(M-O) (except **1-4**) and ν(M-N) (except **1-5**), respectively.

Table 2. Thermal analyses results of compounds **6-9**.

Compound	Temperature (°C)	DTG _{max} (°C)	Leaving Group	Found (%)	Calculated (%)
6	30-148	131	H ₂ O	4.00	4.00
	148-550	205, 239, 322	2Hsalic+2ap	81.88	81.80
	-	-	Cu	14.12	14.20
7	30-275	255	C ₃ H ₃	13.02	13.00
	275-650	288	C ₁₃ H ₁₂ N ₂ O ₃	66.33	66.30
	-	-	Cu	20.65	20.70
8	30-150	120	H ₂ O	3.88	4.10
	150-215	201	2a4mp	23.29	23.30
	215-700	259, 304, 340	2Hsalic	59,13	58,70
			Cu	13.70	13.90
9	30-317	267, 289, 314	2a5mp	35.13	35.20
	317-600	440	Hsalic	44,22	44.1
			Cu	20.65	20.70

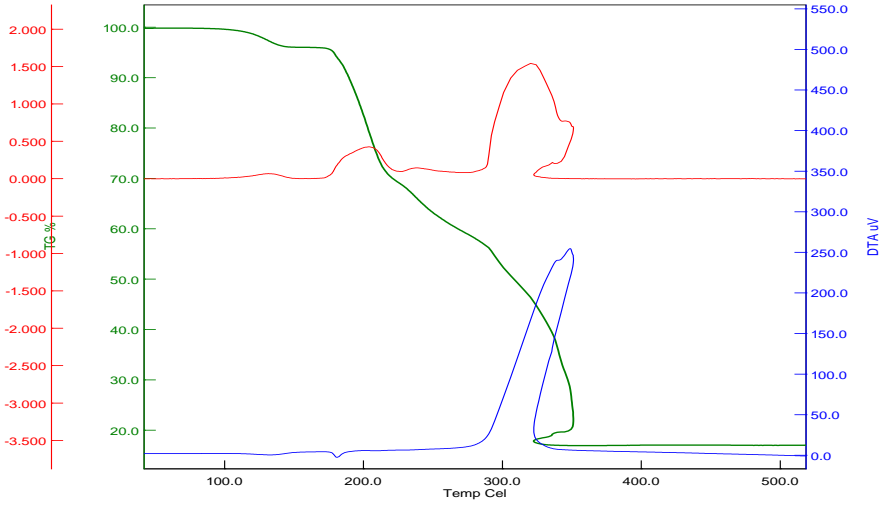


Fig. 2. Thermal analysis results of 6.

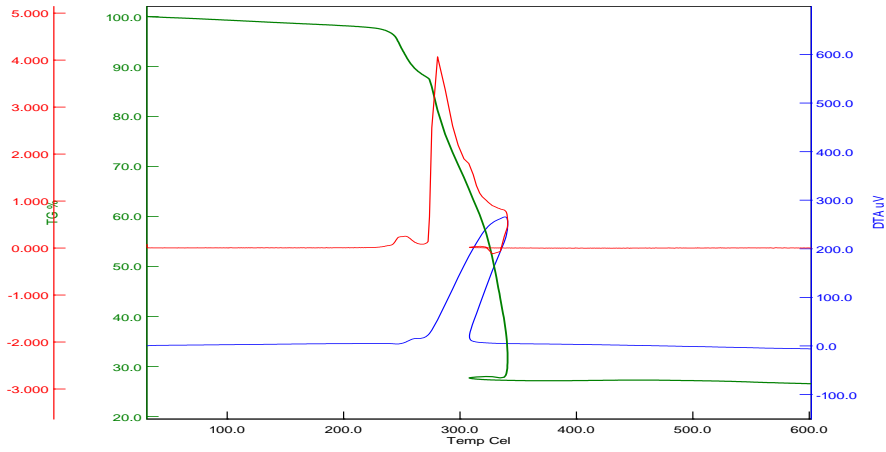


Fig. 3. Thermal analysis results of 7.

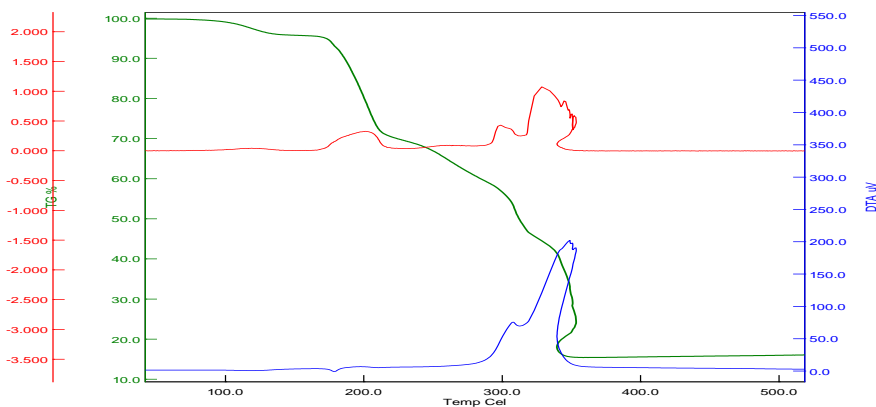


Fig. 4. Thermal analysis results of **8**.

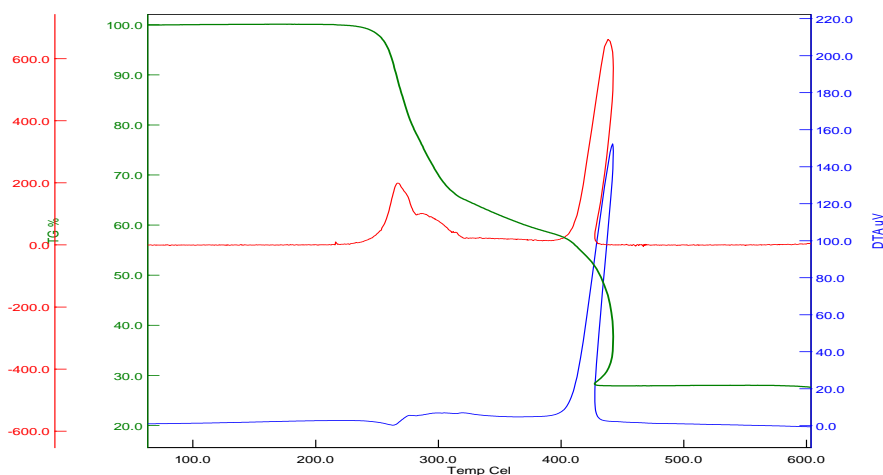


Fig. 5. Thermal analysis results of **9**.

Table 3. IR data of all compounds (cm^{-1}).

	1	2	3	4	5	6	7	8	9
$\nu(\text{O-H})$	3450(br)	3500(br)	3420(br)	3425(br)	3550(br)	3437(br)	-	3437(br)	
$\nu(\text{N-H})$	3285(m) 3202(m)	3377(m) 3325(m)	3316(m) 3294(m)	3320(m) 3293(m)	-	3401(m) 3341(m)	3405(m) 3320(m)	3322(m) 3291(m)	3420(m) 3330(m)
$\nu(\text{C-H})_{\text{ar}}$	3046(w)	3087(w)	3082(w)	3056(w)	3063(w)	3104(w)	3055(w)	3082(w)	3060(w)
$\nu(\text{C-H})_{\text{alip}}$	-	2977(w) 2918(w) 2876(w)	2981(w) 2870(w) 2821(w)	2956(w) 2912(w) 2878(w)	-	-	2974(w) 2923(w) 2855(w)	2971(w) 2881(w) 2794(w)	2959(w) 2891(w) 2783(w)
$\nu(\text{N}^+-\text{H})$	2706(w) 2545(w)	2709(w) 2529(w)	2717(w) 2523(w)	2745(w) 2560(w)	-	-	-	-	-
$\nu(\text{C=O})$	1670(s)	1667(s)	1666(s)	1659(s)	1601(s) 1402(s)	1674(s) 1466(s)	1662(s) 1449(s)	1667(s) 1448(s)	1636(s) 1436(s)
$\nu(\text{C=N})$	1647(s)	1635(s)	1642(s)	1623(s)	1557(s)	1627(s)	1640(s)	1609(s)	1601(s)
$\nu(\text{C=C})$	1613(s)	1575(s)	1606(s)	1578(s)	1487(s)	1596(s)	1602(s)	1584(s)	1562(s)

	1588(s)	1480(s)	1579(s)	1553(s)	1470(s)	1568(s)	1560(s)	1553(s)	1516(s)
	1553(s)	1451(s)	1552(s)	1478(s)	1434(s)	1526(s)	1505(s)	1525(s)	1501(s)
	1479(s)		1478(s)	1446(s)		1488(s)	1471(s)	1479(s)	1484(s)
	1451(s)		1448(s)			1443(s)	1408(s)		1456(s)
v(C-O)	1374(s)	1380(s)	1378(s)	1378(s)	1331(s)	1379(s)	1388(s)	1379(s)	1390(s)
	1248(s)	1250(s)	1248(s)	1252(s)	1240(s)	1225(s)	1261(s)	1249(s)	1250(s)
	1105(s)	1075(s)	1138(s)	1135(s)	1155(s)	1143(s)	1140(s)	1138(s)	1153(s)
v(py)	753(s)	757(s)	756(s)	750(s)	-	748(s)	753(s)	757(s)	753(s)
v(M-O)	-	-	-	-	582(w)	533(w)	593(w)	551(w)	562(w)
v(M-N)	-	-	-	-	-	426(w)	435(w)	450(w)	434(w)

3.4. Results of UV/Vis measurements

The electronic spectra of compounds **1-4** (Fig. 6) and **6-9** (Fig. 7) were registered in DMSO. π - π^* and n - π^* transitions are shown 317 nm ($38640 \text{ Lmol}^{-1}\text{cm}^{-1}$) and 309 nm ($36410 \text{ Lmol}^{-1}\text{cm}^{-1}$) for **1**, 320 nm ($38640 \text{ Lmol}^{-1}\text{cm}^{-1}$) and 314 nm ($37030 \text{ Lmol}^{-1}\text{cm}^{-1}$) for **2**, 324 nm ($43400 \text{ Lmol}^{-1}\text{cm}^{-1}$) and 315 nm ($39720 \text{ Lmol}^{-1}\text{cm}^{-1}$) for **3**, 308 nm ($48170 \text{ Lmol}^{-1}\text{cm}^{-1}$) and 302 nm ($43400 \text{ Lmol}^{-1}\text{cm}^{-1}$) for **4**, 414 nm ($01170 \text{ Lmol}^{-1}\text{cm}^{-1}$) and 317 nm ($40390 \text{ Lmol}^{-1}\text{cm}^{-1}$) for **6**, 412 nm ($01390 \text{ Lmol}^{-1}\text{cm}^{-1}$) and 318 nm ($45160 \text{ Lmol}^{-1}\text{cm}^{-1}$) for **7**, 413 nm ($01320 \text{ Lmol}^{-1}\text{cm}^{-1}$) and 316 nm ($32490 \text{ Lmol}^{-1}\text{cm}^{-1}$) for **8**, 413 nm ($01320 \text{ Lmol}^{-1}\text{cm}^{-1}$) and 316 nm ($32490 \text{ Lmol}^{-1}\text{cm}^{-1}$) for **9**. The d-d transitions are shown at 769 nm ($170 \text{ Lmol}^{-1}\text{cm}^{-1}$) for **6**, 750 nm ($160 \text{ Lmol}^{-1}\text{cm}^{-1}$) for **7**, 750 nm ($180 \text{ Lmol}^{-1}\text{cm}^{-1}$) for **8** and 750 nm ($180 \text{ Lmol}^{-1}\text{cm}^{-1}$) for **9** [30].

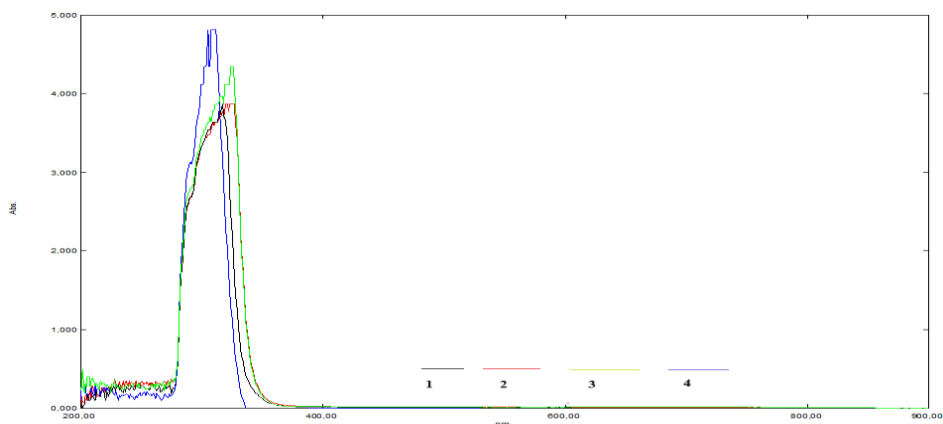


Fig. 6. UV-Vis spectra of compound **1-4**.

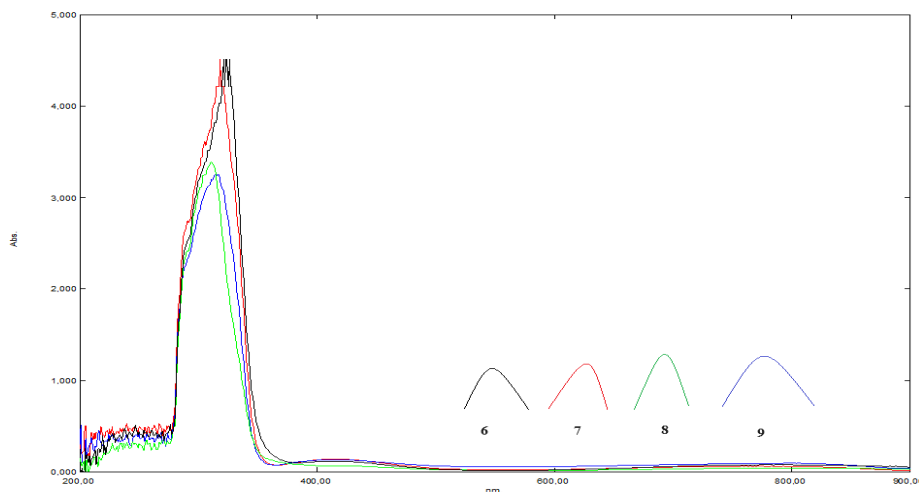


Fig. 7. UV-Vis spectra of compound 6-9.

3.5. Magnetic susceptibilities

Magnetic susceptibility results of Cu(II) complexes (6-9) were found between 1.60, 1.63, 1.61 and 1.65 BM, respectively. These values say that there are unpaired electrons in the complexes. The magnetic moment for the Cu(II) ion obtained in the tetrahedral geometry is also consistent with this value [31,32].

3.6. Antimicrobial results of compounds.

The antimicrobial activity of Fluconazole, Levofloxacin, Cefepime, Vancomycin, copper(II) acetate, free ligands and 1-9 were investigated by microdilution method. All compounds showed activity against bacteria and yeast. MIC values of anti-fungal and anti-microbial agents, all compounds are given in Table 4. Activity values are similar to 2-aminopyridine found in the literature [33-36].

The antifungal drug and substances have activity against *C. albicans* when MIC values are compared; copper(II) acetate observed greater activity than according to Fluconazole while H₂salic, 2ap, 2a3mp, 2a5mp, 2, 4, 6 and 7 showed equal effective. Other compounds were found to have a lower degree of action.

All antibacterial drugs and substances have activity against *L. monocytogenes*; when MIC values are compared; all compounds indicated greater activity than (except 2 and 6) according to Vancomycin {H₂salic, 2a3mp and 7 > copper(II) acetate, 2ap, 2a4mp, 2a5mp, 1, 3-5, 8 and 9 > 2 and 6}. H₂salic, 2a3mp and 7 exhibited comparable efficacy in relation to Levofloxacin and Cefepime, while demonstrating lower effectiveness compared to other compounds.

B. subtilis; all compounds showed greater activity than according to Vancomycin {2a5mp > 5 and 8 > copper(II) acetate, H₂salic, 2ap, 2a3mp, 2a4mp, 2, 4, 6, 7 and 8 > 1 and 3}. 2a5mp, 5 and 8 showed greater activity than according to Levofloxacin and Cefepime while copper(II) acetate, H₂salic, 2ap, 2a3mp, 2a4mp, 2, 4, 6, 7 and 8 showed equally effective. Other compounds were found to have a lower degree of according to Levofloxacin and Cefepime.

E. coli; 2ap, 2a3mp, 2a4mp and 7 indicated greater activity than according to Cefepime while copper(II) acetate, H₂salic, 2a5mp, 4, 5, 8 and 9 showed similar effective. Compounds 1-3 and 6 were found to have a lower degree of according to Cefepime. 2ap indicated greater activity than according to Vancomycin and Levofloxacin while 2a3mp, 2a4mp and 7 showed equal effective. The other compounds seen lower degree of according to Vancomycin and

Levofloxacin.

S. aureus: while **7** and **9** showed equally effective, the other compounds were found to have a lower degree of according to Levofloxacin and Vancomycin. **7** and **9** showed greater activity than according to Cefepime while the other compounds (except **3**) showed equally effective. Compound **3** was found to have a lower degree of according to Cefepime.

E. faecalis: **8** displayed analogous performance with respect to Cefepime, whereas other substances showed a diminished level of effectiveness {copper(II) acetate, H₂salic, 2ap, 2a3mp, 2a4mp, 2a5mp, **1**, **3-5**, **7** and **9** > **2** and **8**}. While **8** showed equally effective according to Levofloxacin and Vancomycin, the other compounds (except **2** and **6**) showed equally effective. Compounds **2** and **6** was found to have a lower degree of according to Cefepime.

P. aeruginosa: 2a4mp, H₂salic and **9** showed greater activity than according to Vancomycin while copper(II) acetate, 2ap, 2a3mp, 2a5mp, **4**, **6** and **7** equally effective. Compounds **1-3**, **5** and **8** were found to have a lower degree of according to Vancomycin. 2a4mp demonstrated superior efficacy compared to Cefepime and Levofloxacin, while **9** proved to be equally potent. Other compounds exhibited a lesser degree of effectiveness in relation to Cefepime and Levofloxacin.

Table 4. MIC values of compounds (µg/mL).

Compound	<i>C. albicans</i>	<i>L. monocytogenes</i>	<i>B. subtilis</i>	<i>E. faecalis</i>	<i>S. aureus</i>	<i>E. coli</i>	<i>P. aeruginosa</i>
Vankomisin	-	125.00	250.00	62.50	31.25	31.25	62.50
Levoflaksasn	-	31.25	62.50	62.50	31.25	31.25	31.25
Sefepim	-	31.25	62.50	31.25	62.50	62.50	31.25
Fluconazole	62.50	-	-	-	-	-	-
copper(II) acetate	31.25	31.25	62.50	62.50	62.50	62.50	62.50
H ₂ salic	62.50	31.25	62.50	62.50	62.50	62.50	31.25
2ap	62.50	62.50	62.50	62.50	62.50	15.60	62.50
2a3mp	62.50	31.25	62.50	62.50	62.50	31.25	62.50
2a4mp	125.00	62.50	62.50	62.50	62.50	31.25	15.60
2a5mp	62.50	62.50	7.80	62.50	62.50	62.50	62.50
1	125.00	62.50	125.00	62.50	62.50	125.00	125.00
2	62.50	125.00	62.50	125.00	62.50	125.00	125.00
3	125.00	62.50	125.00	62.50	125.00	125.00	125.00
4	62.50	62.50	62.50	62.50	62.50	62.50	62.50
5	125.00	62.50	31.25	62.50	62.50	62.50	125.00
6	62.50	125.00	62.50	125.00	62.50	125.00	62.50
7	62.50	31.25	62.50	62.50	31.25	31.25	62.50
8	125.00	62.50	31.25	31.25	62.50	62.50	125.00
9	125.00	62.50	62.50	62.50	31.25	62.50	31.25

4. Conclusions

New four salts (**1-4**) between salicylic acid (H₂salic) and 2-amino-X-pyridine {X = (2ap), 3-methyl (2a3mp), 4-methyl (2a4mp) and 5-methyl (2a5mp)}, the Cu(II) complex of H₂salic (**5**) by methods available in the literature and new Cu(II) complexes (**6-9**) of **1-4** have been synthesized. The structures of **6-9** were suggested by elemental

analysis, AAS, FT-IR, UV-Vis and magnetic susceptibility studies. While the metal:acid:base ratio was 1:2:1 for **6** and **8** 1:1:1 for **7** and **9**. The spectroscopic research results indicated that complexes **6–9** have tetrahedral geometries. All compounds showed activity against bacteria and yeast. Compounds copper(II) acetate for *C. Albicans*, copper(II) acetate, H₂salic, 2a3mp and **7** for *L. monocytogenes*, 2a5mp for *B. subtilis*, **7** and **9** for *S. aureus*, 2ap for *E. Coli*, **8** for *E. Faecalis*, and 2a4mp for *P. aeruginosa* have the best activity.

Acknowledgements

The authors acknowledge the assistance provided by the Kütahya Dumlupınar University Research Fund (Grant No. 2019/12 and 2020/02).

References

- [1] Marinescu, M. (2017). “2-Aminopyridine – a classic and trendy pharmacophore”. *International Journal of Pharma and Bio Sciences*, vol. 8, no. 2, pp. 338-355, doi: 10.22376/ijpbs.2017.8.2.p338-355.
- [2] Rosheen, S. S. and Utreja, D. (2023). “Salicylic acid: synthetic strategies and their biological activities”. *ChemistrySelect*, vol. 8, pp. e202204614, doi: 10.1002/slct.202204614.
- [3] Rao, C. N. R., Natarajan, S. and Vaidhyanathan, R. (2004). “Metal carboxylates with open architectures”. *Angewandte Chemie International Edition*, vol. 43, pp. 1466-1496, doi: 10.1002/anie.200300588.
- [4] Yin, M. C., Ai, C. C., Yuan, L. J., Wang, C. W. and Sun, J. T. (2004). “Synthesis, structure and luminescent property of a binuclear terbium complex [Tb₂(Hsal)₈(H₂O)₂][(Hphen)₂].2H₂O”. *Journal of Molecular Structure*, vol. 691, no. 1-3, pp. 33-37, doi: 10.1016/j.molstruc.2003.10.032.
- [5] Shake, A. R., Tsai, H. L., Webb, R. J., Foltz, K., Christou, G. and Hendrickson, D.N. (1994). “High-spin molecules: Iron(III) incorporation into [Mn₁₂O₁₂(O₂CMe)₁₆(H₂O)₄] to yield [Mn₈Fe₄O₁₂(O₂CMe)₁₆(H₂O)₄] and its influence on the S = 10 ground state of the former”. *Inorganic Chemistry*, vol. 33, no. 26, pp. 6020-6028, doi: 10.1021/ic00104a009.
- [6] Murugavel, R., Karambelkar, V. V., Anantharaman, G. and Walawalkar, M. G. (2001). “Synthesis, spectral characterization, and structural studies of 2-aminobenzoate complexes of divalent alkaline earth metal ions: X-ray crystal structures of [Ca(2-aba)₂(OH₂)₃]_∞, [Sr(2-aba)₂(OH₂)₂·H₂O]_∞, and [Ba(2-aba)₂(OH₂)_∞ (2-abaH = 2-NH₂C₆H₄COOH)”. *Inorganic Chemistry*, vol. 39, pp. 1381-1390., doi: 10.1021/ic990895k.
- [7] Murugavel, R., Baheti, K. and Anantharaman, G. (2001). “Reactions of 2-mercaptobenzoic acid with divalent alkaline earth metal ions: synthesis, spectral studies, and single-crystal X-ray structures of calcium, strontium, and barium complexes of 2,2'-dithiobis(benzoic acid)”. *Inorganic Chemistry*, vol. 40, pp. 6870-6878, doi: 10.1021/ic010519b.
- [8] Murugavel, R., Krishnamurthy, D. and Sathiyendiran, M. (2002). “Anionic metal-organic and cationic organic layer alternation in the coordination polymers [{M(BTEC)(OH₂)₄]{C₄H₁₂N₂}.4H₂O]_n (M = Co, Ni, and Zn; BTEC = 1,2,4,5-benzenetetracarboxylate)”. *Journal of the Chemical Society*, pp. 34-39, doi: 10.1039/B105687P.
- [9] Murugavel, R. and Banerjee, S. (2003). “First alkaline earth metal 3-aminobenzoate (3-aba) complex: 1-D polymeric [Ca(3-aba)₂(H₂O)₂]_n assembly”. *Inorganic Chemistry Communications*, vol. 6, pp. 810-814, doi: 10.1016/S1387-7003(03)00112-6.
- [10] Prabusankar, G. and Murugavel, R. (2004). “Hexameric organotin carboxylates with cyclic and drum structures”. *Organometallics*, vol. 23, pp. 5644-5647, doi: 10.1021/om049584u.
- [11] Han, B. and Hoang, B.X. (2018). “Metal complexes as pharmaceuticals for treatment and prevention of cancer and inflammatory diseases”. *US20180353539A1*.
- [12] Connor et al. M. O'. (2012). “Copper(II) complexes of salicylic acid combining superoxide dismutase mimetic properties with DNA binding and cleaving capabilities display promising chemotherapeutic potential with fast acting in vitro cytotoxicity against cisplatin sensitive and resistant”. *Journal of Medicinal Chemistry*, vol. 55, pp. 1957-1968, doi: 10.1021/jm201041d.
- [13] Ravaeva, M. Y., Cheretaev, I. V. and Chuyan, E. N. (2021). “Metal salicylates Co²⁺, Zn²⁺, Ni²⁺, Mn²⁺, Li⁺ and Mg²⁺: properties and effect on pain sensitivity”. *Journal of Physics: Conference Series*, vol. 1967, pp. 012033, doi: 10.1088/1742-6596/1967/1/012033.
- [14] Diaz, A. M., Villalonga, R. and Cao, R. (2009). “Antioxidative properties of copper(II) complexes”. *Journal of Coordination Chemistry*, vol. 62, no. 1, pp. 100-107, doi: 10.1080/00958970802474755.
- [15] Geraghty, M., Sheridan, V., McCann, M., Devereux, M. and McKee, V. (1999). “Synthesis and anti-Candida activity of copper(II) and manganese(II) carboxylate complexes X-ray crystal structures of [Cu(sal)(bipy)]C₂H₅OH.H₂O and [Cu(norb)(phen)₂].6.5H₂O (salH₂ = salicylic acid; norbH₂ = cis-5-norbornene-endo-2,3-dicarboxylic acid; bipy = 2,2'-bipyridine; phen = 1,10-phenanthroline)”. *Polyhedron*, vol. 18, no. 22, pp. 2931-2939, doi: 10.1016/S0277-5387(99)00201-6.
- [16] Iebudak, H. et al. (2003). “Syntheses, characterization and crystal structures of novel amine adducts of metal saccharinates, orotates and salicylates”. *Journal of Molecular Structure*, vol. 657, no. 1-3, pp. 255-270, doi: 10.1016/S0022-2860(03)00404-6.
- [17] Devereux, M. et al. (2007). “Synthesis, catalase, superoxide dismutase and antitumor activities of copper(II) carboxylate complexes incorporating benzimidazole, 1,10-phenanthroline and bipyridine ligands: X-ray crystal structures of [Cu(BZA)₂(bipy)(H₂O)], [Cu(SalH)₂(BZDH)₂] and [Cu(CH₃COO)₂(5,6-DMBZDH)₂] (SalH₂ = salicylic acid; BZAH = benzoic acid; BZDH = benzimidazole and 5,6-DMBZDH = 5,6-dimethylbenzimidazole)”. *Polyhedron*, vol. 26, no. 15, pp. 4073-4084, doi: 10.1016/j.poly.2007.05.006.

- [18] Han, B. and Hoang, B. X. (2018). "Metal complexes as pharmaceuticals for treatment and prevention of cancer and inflammatory diseases". United States, US20180353539 A1 2018-12-13.
- [19] Cluzan, R. and Nicod, B. (1971). "Analgesic and antiinflammatory effects of aminopyridine salicylates". France, FR2085646 A5 1971-12-31.
- [20] Sikkema, D. J. (2007). "A rigid rod polypyridobisimidazole: "M5", towards a new generation of structural composites". 1st Proceedings of the Aachen-Dresden International Textile Conference, pp. 36-56.
- [21] Montoro, D. F., Calatayud, D. J. and Vilar, D. A. (1979). "Salicylic acid derivatives". Spain, ES475267 A1 1979-05-01.
- [22] Lu, J. F., Ge, H. G. and Min, S. T. (2011). "Syntheses and structural characterization of silver(I) complexes with pyridine and carboxylate derivatives". *Russian Journal of Coordination Chemistry*, vol. 37, no. 1, pp. 36-39, doi: 10.1134/S107032841101009X.
- [23] Hanif, M., Khan, E., Khalid, M., Tahir, M. N., Morais, S. F. A. and Braga, A. A. C. (2020). "2-Amino-3-methylpyridinium, 2-amino-4-methylbenzothiazolium and 2-amino-5-chloropyridinium salts. Experimental and theoretical findings". *Journal of Molecular Structure*, vol. 1222, pp. 128914, doi: 10.1016/j.molstruc.2020.128914.
- [24] Hemamalini, M. and Fun, H. K. (2010). "2-Amino-4-methylpyridinium 2-hydroxybenzoate". *Acta Crystallographica Section E: Crystallographic Communications*, vol. E66, no. 8, pp. o2151-o2152, doi: 10.1107/S160053681002965X.
- [25] Prabha, D., Harish, P., Babu, B. and Moorthi, V. S. N. (2016). "Growth and characterization of an organic crystal and DFT studies of 2-amino 5-methyl pyridinium salicylate". *Oriental Journal of Chemistry*, vol. 32, no. 4, pp. 1937-1945, doi: 10.13005/ojc/320420.
- [26] Punithaveni, B., Thilagavathy, K., Muthukumarasamy, N., Nithyaprakash, D. and Saravanabhavan, M. (2018). "Structural, spectral, electrical, Z-scan and HOMO LUMO studies on new 2-amino-6-methylpyridinium 2-hydroxybenzoate crystal". *Materials Science-Poland*, vol. 36, no. 4, pp. 537-546, doi: 10.2478/msp-2018-0098.
- [27] Sivakumar, P., Sudhakar, S., Israel S. and Chakkaravarthi, G. (2016). "2-Amino-6-methylpyridinium 2-hydroxybenzoate". *IUCrData*, vol. 1, no. 5, pp. x160747-x60750, doi: 10.1107/S2414314616007471.
- [28] İlkimen, H. and Yenikaya, C. (2017). "Synthesis and characterization of mixed ligand Cu(II) complexes of salicylic acid derivatives with 2-aminobenzotriazol derivatives". *Pamukkale University Journal of Engineering Sciences*, vol. 23, no. 7, pp. 899-907, doi: 10.5505/pajes.2016.76735.
- [29] Nakamoto, K. (1997). "Infrared and raman spectra of inorganic and coordination compounds". 5th ed. NewYork: Wiley-Interscience, pp. 232.
- [30] Büyükkıdan N. and Ozer, S. (2013). "Synthesis and characterization of Ni(II) and Cu(II) complexes derived from novel phenolic Mannich bases". *Turkish Journal of Chemistry*, vol. 37, no. 1, pp. 101-110, doi: 10.3906/kim-1203-67.
- [31] Hathaway, B. J., Holah, D. G. and Underhill, A. E. (1962). "468. The preparation and properties of some bivalent transition-metal tetrafluoroborate-methyl cyanide complexes". *Journal of the Chemical Society*, vol. 24, no. 4, pp. 2444-2448, 1962, doi: 10.1039/JR9620002444.
- [32] Rehman, S. U., Ikram, M., Rehman, S., Islam, N. U., Jan, N. and Mex, J. (2011). "Synthesis and characterization of Ni(II), Cu(II) and Zn(II) tetrahedral transition metal complexes of modified hydrazine". *Revista de la Sociedad Química de Mexico*, vol. 55, no. 3, pp. 164-167, doi: 10.29356/jmcs.v55i3.815.
- [33] Yenikaya, C. et al. (2011). "Synthesis, characterization and biological evaluation of novel Cu(II) complexes with proton transfer salt of 2,6-pyridinedicarboxylic acid and 2-amino-4-methylpyridine". *Journal of Coordination Chemistry*, vol. 64, no. 19, pp. 3353-3365, doi: 10.1080/00958972.2011.620608.
- [34] İlkimen, H., Türken, N. and Gülbandılar, A. (2021). "Synthesis, characterization, antimicrobial and antifungal activity of studies of two novel aminopyridine-sulfamoylbenzoic acid salts and their Cu(II) complexes". *Journal of the Iranian Chemical Society*, vol. 18, pp. 1941-1946, doi: 10.1007/s13738-021-02157-4.
- [35] İlkimen, H., Salün, S. G., Gülbandılar, A. and Sarı, M. (2022). "The new salt of 2-amino-3-methylpyridine with dipicolinic acid and its metal complexes: Synthesis, characterization and antimicrobial activity studies". *Journal of Molecular Structure*, vol. 1270, pp. 133961, doi: 10.1016/j.molstruc.2022.133961.
- [36] İlkimen H. and Gülbandılar A. (2023). "Synthesis, characterization, antimicrobial and antifungal activity studies of four novel 2-aminopyridine and 2,4-dichloro-5-sulfamoylbenzoic acid salts and their Cu(II) complexes". *Kuwait Journal of Science*, vol. 50, no. 3A, pp. 1-11, doi: 10.48129/kjs.19163.



RESEARCH ARTICLE

Receive Date: 14.10.2023

Accepted Date: 08.03.2024

The effect of fingerprint enhancement methods applied on adhesive surfaces on DNA recovery: a preliminary study

Fatma Cavus Yonar^a, Yakup Gulekci^{b,*}

^aIstanbul University-Cerrahpasa, Institute of Forensic Sciences and Legal Medicine, 34500, Istanbul, Turkiye,
ORCID: 0000-0001-9643-6850

^bKutahya Health Sciences University, Faculty of Engineering and Natural Sciences, Department of Forensic Sciences, 43100 Kutahya, Turkiye,
ORCID: 0000-0001-5941-8434

Abstract

The presence of body fluids such as blood, saliva, semen or urine during fingerprint research on the evidence taken from the crime scene makes it necessary to protect biological materials to examine the evidence in multiple ways. Therefore, it is crucial that fingerprint development (FD) techniques do not disrupt the structure of biological materials during FD procedures. In this sense, it is essential to determine whether biological material or fingerprints should be the priority during the collection of evidentiary materials, to determine the systematic order and to determine whether the FD methods to be applied cause damage to the genetic material used in the identification of individuals and to evaluate them in terms of their evidentiary quality. This study investigated the effects of the application of trace detection methods on DNA profiling processes in evidence where fingerprints and biological samples are found at the same time. In this study, blood, saliva, semen and urine samples were taken from a male individual who signed an informed consent form at the laboratory stage. The samples were applied 50 μ L on the adhesive tape surface. After application, the samples were treated with crystal violet (CV) and sticky side (SS) fingerprint development chemicals suitable for the surface type. The prepared samples were dried under room conditions. After 1 day and 45 days under normal room conditions, silica-based DNA extraction was performed. After extraction, DNA quantification was performed using the fluorimetry method. In the study, biological samples with known DNA content were used to focus on DNA quantification. Among the fresh samples prepared in the study, DNA recovery was higher in the SS-treated urine, blood and saliva samples and in the CV-treated semen sample group compared to the other groups. This shows that chemical treatment of some biological samples on adhesive tape increases the efficiency of DNA recovery. When the 45-day waiting samples were compared with the control group samples, DNA recovery decreased in CV-treated urine and blood samples, while DNA recovery increased in SS-treated urine and blood samples. In semen samples, both CV and SS treatment negatively affected DNA recovery. In saliva samples, DNA recovery increased ~2-fold in the CV-treated sample group, while SS treatment caused a ~75% decrease in DNA recovery. The results show that the non-porous adhesive tape does not adversely affect the amount of DNA in terms of STR profiling of latent FD chemicals used on the surfaces and that adhesive tape treated with fingerprint enhancement chemicals can actually be used for advanced forensic genetic analyses for DNA extraction on surfaces.

© 2023 DPU All rights reserved.

Keywords: Fingerprint enhancement; Adhesive surface; DNA extraction; Biological fluids; Forensic genetics

*Corresponding author. Tel.: +90-506-792-93-59

E-mail address: yakup.gulekci@ksbu.edu.tr

© 2023 DPU All rights reserved.

1. Introduction

In crime dynamics, no matter how professional the perpetrators are, a suspect leaves something at the crime scene or the victim and takes something away from the crime scene or the victim. With the current technology, it is possible to identify and analyse the visible/invisible evidence left at the crime scene [1], [2]. Forensic scientists investigate human biological fluids (such as blood, saliva, semen, and urine) and comparable fingerprints as the most reliable types of evidence, which have the characteristics of certainty, uniqueness (being unique to the individual) and high distinctiveness, to provide objective evidence to the courts regarding the elucidation of crimes and the identification of criminals.

Fingerprints are one of the oldest forms of forensic evidence linking the crime scene to the offender and are based on the assumption that everyone has a unique set of patterns on the fingertip. The residues that make up the fingerprint pattern are a mixture of secretions from various glands in the skin (sudoriferous eccrine, apocrine, and sebaceous), but also a complex formation mixed with environmental substances that come into contact with the person's skin [3]. The condition and structure of the surface (such as rough-smooth, porous-non-porous, absorbent-non-absorbent) on which the fingerprint is found, whether there is any residue affecting the fingerprint (such as oil, blood), environmental factors (such as wetting, drying), the possible age of the fingerprint (duration of stay on the surface) make it necessary for the methods used in fingerprint research to be different [4]. Among the FD methods, Iodine Vapour, Ninhydrin, DFO (9- Diazaflueron), Indanedione, 5-MTN (5-(methylthio)ninhydrin), Termanin and Silver Nitrate are used on porous surfaces (such as raw wood, paper) and on non-porous surfaces (metal, glass, plastic, Cyanoacrylate and colouring methods (such as Rhodamine-6G, Ardrex, Nile Red, Yellow Basic), Sudan Black, Amido Black, Hungarian Red, SPR and especially Crystal Violet, Sticky Side chemicals are used on adhesive surfaces. The main goal of FD is to take advantage of the adhesion and colouring properties of the chemicals used for the sweat and sweat-containing substances in the fingerprint [5].

In forensic cases such as murder, sexual assault, and theft, efforts are made to determine whether there is a connection between the suspects and the incident by analysing genetic information obtained from biological materials (such as blood, semen, saliva, and urine) found at the crime scene. In criminal investigations, biological samples collected from the crime scene are identified by DNA analysis. The DNA molecule is found in the cell, which is the building block of human beings, and the DNA of all people except identical twins is different from each other. Another important feature of DNA is that it shows the same structural features in all human cells. In addition, it is known that DNA is inherited from parents and maintains its structure, except for some rare negative effects, such as mutations and external factors. Mutations, while sometimes considered negative, are indeed a natural part of evolution and contribute to genetic diversity. Additionally, sometimes external factors can induce adverse effects on DNA. These scientific facts have made DNA-based identification one of the most valid and precise methods [6].

In some incidents, findings can yield more than one type of evidence group, such as physical, chemical, biological, and trace evidence. One of the most critical findings that contain different evidence groups in criminal investigations is adhesive surfaces. Biological and trace evidence obtained from the adhesive surfaces of the tapes makes an essential contribution to the rapid resolution of forensic incidents. In particular, adhesive surfaces are actively used in the creation of terrorist attacks and bomb devices, in the packaging used in the transport of narcotic substances, in the detection of kidnapped persons or in the identification of the number plates attached to stolen motor vehicles. The importance of adhesive surfaces in terms of criminal investigations stems from the fact that these surfaces have any connection with the environment in which the event took place and that they contain DNA or fingerprints of the people who carried out the event [7], [8]. Judicial authorities may request forensic scientists to perform both fingerprint and DNA analyses on such evidence from the crime scene in criminal laboratories [9].

However, fingerprints obtained from crime scenes may be dirty or partial and may not be suitable for identification. DNA profiling of these prints, which are not suitable for identification, can be used to identify this evidence [10]. Thus, combining partial results of fingerprints and DNA can increase confidence in the identification of the suspect. However, the concern that FD may reduce DNA recovery also raises concerns about the simultaneous use of these complementary analyses [11]. Although this concern contributes to the development of analyses to determine the DNA yield on the surface of the finding by using latent trace development techniques on the finding, the data obtained may contain complex and variable results since biological samples containing non-standardised and uncertain amounts of DNA are used in these studies [12], [13]. Using a starting material with known amounts of DNA can eliminate the variability in DNA recovery and allows for statistical analysis between methods for DNA recovery [11]. Many studies have been conducted to evaluate the effects of FD techniques on subsequent DNA profiling, depending on the quantity and quality of DNA present in the fingerprint [14]–[18]. However, research on the recovery of DNA on surfaces containing different biological fluids and treated with FD chemicals is at the theoretical level and is very limited. Since criminal investigations are multi-faceted and the evidence obtained is valuable not only in terms of fingerprint research but also in terms of biological investigations, it is important to investigate whether the applied FD techniques disrupt the genetic material of the perpetrator of the incident. This aims to investigate the effects of FD methods applied to biological samples on adhesive surfaces on which fingerprints and biological fluids are found at the same time on the recovery of DNA from the target surface, and the results obtained are aimed to be used for the reorganisation of criminal analysis applications and obtaining results.

2. Materials and methods

2.1. Fingerprint development reagent and chemicals

The formation procedures of the fingerprint development/staining methods used in the study were performed according to the formulations of Bleay et al.[5], [19]:

Crystal Violet: 1 g Crystal Violet (CV) (Meck, Germany) was weighed and dissolved in 1000 mL distilled water to prepare a 1000 mL working solution. Sticky Side (SS) (Sirchie, USA) FD was used as a ready solution.

2.2. Preparation of biological samples

All procedures performed in this human participant study complied with the ethical standards of the institutional and/or national research committee, the 1964 Declaration of Helsinki and its subsequent amendments, or comparable ethical standards. 70 mL venous blood, 50 mL saliva, 100 mL urine, and 40 mL semen samples were obtained from a 35-year-old healthy man on the same day, who signed an informed consent form. Biological samples (venous blood, saliva, urine and sperm) were stored in a refrigerator (Vestel, Turkey) at +4 °C until the study was performed. No special instructions were given to the volunteers to obtain realistic data on forensic cases. The surface and consumables were exposed to UV light for 30 minutes before the experimental work to prevent possible contamination. 50 µl of biological sample was transferred onto adhesive tape (Pattex, Germany) using an automatic pipette (Eppendorf, Germany) (Figure 1). Samples of the volunteer applied on the adhesive surface and not treated with any FD were used as control samples. A blind sample for negative control was run with each group of samples to eliminate the risk of contamination.

2.3. Application of FD chemicals

In the study, 2 types of FDs were applied on the adhesive tape surface to which blood, saliva, urine and semen samples were transferred. CV and SS were applied 100 µl with an automatic pipette (Eppendorf, Germany) and allowed to dry in a fume hood.

The prepared samples were analysed at 2 different time intervals (day 1 and day 45) to determine the possible change in the DNA amount of biological samples treated with fingerprint chemicals over time. The prepared samples were kept for 45 days in sterile evidence storage cabinets where the evidence obtained from the crime scene was kept. At the end of the process, DNA extraction was performed from both freshly prepared samples and samples kept for 45 days.

2.4. DNA extraction and DNA quantification

The adhesive tape bearing the biological sample was utilised in its entirety as the starting material for the extraction. The entire adhesive tape, approximately 1.0 cm wide and 3.0 cm long, was placed in a 1.5 mL microcentrifuge tube. DNA extraction of FD chemicals-treated blood, saliva, urine, and semen samples was performed following the manufacturer's instructions and extracted using the QIAamp® DNA Micro Kit (Qiagen, Hilden, Germany). To a 1.5 mL microcentrifuge tube containing the sample, 20 µL proteinase K solution and 400 µL AL kit lysis buffer were added. After vortexing for 15 s, the tube was incubated at 60°C for 30 min for protein digestion. After a brief centrifugation, the resulting mixture was transferred to a 2 mL QIAamp micro-spin column and centrifuged at 8000 rpm for 1 min. After discarding the collection tube and replacing it with a new collection tube, 500 µL of AW1 solution was added to the column and then centrifuged at 8000 rpm for 1 min. After changing to a new collection tube, 500 µL of AW2 solution was added to the column, followed by centrifugation at 14000 rpm for 3 min. After transferring the QIAamp micro-spin column to a new 1.5 mL microcentrifuge tube, 20 µL of AE buffer was added to the column and incubated for 10 min at room temperature. Finally, DNA was extracted by centrifugation at 8000 rpm for 1 min [9], [20].

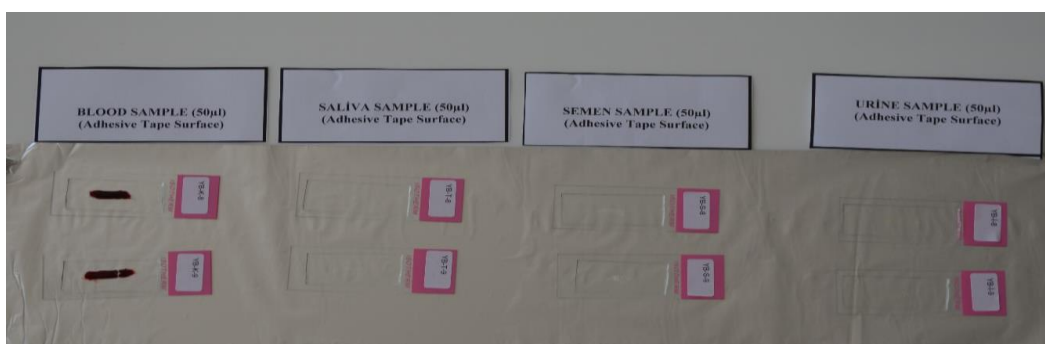


Fig. 1. Biological samples applied on adhesive tape in a volume of 50 µL.

Isolated DNA was quantified by the fluorimetric method using fluorescence technology, which provides high accuracy and sensitivity. In this method, a Qubit® Fluorimeter (Invitrogen by Thermo Fisher Scientific, USA) measuring at 260 nm wavelength was used, and the standard procedure steps for the Quant-iT™ dsDNA HS (Invitrogen by Thermo Fisher Scientific, USA) kit were followed [21]. DNA extraction and DNA quantification were performed for each sample type in 3 repeats.

2.5. Statistical analysis

All statistical evaluations were performed using IBM SPSS Statistics v25.0 software. The differences between the two study groups, such as those based on time, were examined using an independent sample t-test or a Wilcoxon signed-rank test, depending on whether the data followed a normal distribution. For differences between three or

more groups, such as different biological samples or fingerprint development chemicals, one-way ANOVA and post hoc tests were applied. A p-value < 0.05 was considered statistically significant.

3. Results and discussion

Biological samples of various body fluids and evidentiary quality are frequently encountered at the crime scene. Developing fingerprints contaminated with biological samples with different lifetimes and obtained using different methods is challenging for forensic scientists. Furthermore, the concern that FD procedures may adversely affect DNA recovery may prevent the simultaneous use of complementary analyses. Scientists are investigating the effect of chemical agents used to develop fingerprints on DNA yield in biological findings. In this sense, this study tried to determine how the techniques used in the development of latent fingerprints affect DNA recovery, to determine the order of application of fingerprint development processes with biological evidence on adhesive surfaces in terms of evidence security, and to identify and investigate the effects of these FD agents that affect forensic DNA analysis.

The presence of body fluids such as epithelial cells, blood, saliva, urine and semen, as well as fingerprints on the evidence taken from the crime scene, makes it necessary to protect biological materials in the development of latent fingerprints. Therefore, this study will contribute to the determination of the order of examination in the criminal laboratory to prevent the destruction of biological or fingerprint evidence and prevent contamination of evidence. According to studies, the vast majority of studies in the field of fingerprinting have focused on the development of methods to make fingerprints visible on various surfaces and conditions [22]–[24]. However, the limited number of studies discussing the impact of FD methods on other types of evidence often focus on the development of strategies for the recovery of the touch DNA contained in the fingerprint itself and transferred from fingerprints by epithelial cells [2], [25]. Accordingly, the number of studies on the extent to which FD methods change the structural properties of biological materials (such as blood and saliva) or the level of DNA degradation is also limited [11], [26].

DNA extraction and quantification were performed in repeats of 3 from each sample in this study. The average DNA amounts obtained by Qubit Fluorimeter after DNA extraction are given in Table 1. The amount of DNA observed from the blind samples studied in each group to observe the presence of contamination was <0.05 ng/μL. The degradation index was calculated by the ratio of the amount of DNA obtained from the control sample to the amount of DNA obtained from the FD-treated sample (Suppl 1). In addition, in contrast to the studies in the literature, this study focused on DNA quantification and used starting material with known amounts of DNA [26], [27]. The presence of contamination was checked by using a negative control (NC) to ensure internal control throughout the study. No DNA presence was observed in the negative controls used (<0.05 ng/μL). Among the fresh samples prepared in the study, DNA recovery was higher in the SS-treated urine, blood and saliva samples and in the CV-treated semen sample group compared to the other groups (Table 1). Surprisingly, DNA recovery in SS and CV-treated saliva samples, CV-treated semen samples, and SS-treated blood samples increased significantly after chemical treatment compared to the control group. This shows that chemical treatment of some biological samples on adhesive tape increases the efficiency of DNA recovery. When the 45-day waiting samples were compared with the control group samples, DNA recovery decreased in CV-treated urine and blood samples, while DNA recovery increased in SS-treated urine and blood samples. In semen samples, both CV and SS treatment negatively affected DNA recovery. In saliva samples, DNA recovery increased ~2-fold in the CV-treated sample group, while SS treatment caused a ~75% decrease in DNA recovery. It should be noted that oral hygiene or oral microbial content is a factor affecting saliva's DNA content especially.

Table 1. Time-dependent DNA quantification comparison independent sample T-test data obtained from adhesive tape surface treated with fingerprint development reagents.

Sample	Fingerprint development	Time (day)	P values
--------	-------------------------	------------	----------

chemicals		1 day		45 days		
		DNA quantity (ng/μL)*		DNA quantity (ng/μL)*		
		Mean	SD	Mean	SD	
Urine	Control	0,523	0,111	0,249	0,034	0,015
	Crystal violet	0,074	0,008	0,120	0,016	0,011
	Sticky side	0,485	0,019	0,342	0,027	0,002
Blood	Control	2,137	0,086	1,127	0,150	0,001
	Crystal violet	3,067	0,120	0,217	0,070	<0,001
	Sticky side	7,370	0,285	3,193	0,316	<0,001
Semen	Control	4,850	0,719	10,937	0,627	<0,001
	Crystal violet	6,790	0,403	10,403	0,425	<0,001
	Sticky side	4,730	0,205	6,177	0,265	0,002
Saliva	Control	3,830	0,519	8,820	0,390	<0,001
	Crystal violet	24,153	0,243	16,793	0,627	0,001
	Sticky side	47,973	1,830	2,430	0,415	<0,001

* Sample repeats n=3

In the comparison between groups (Figure 2), when biological samples treated with FD and kept for 1 to 45 days were compared according to sample type, DNA recovery decreased in urine, blood and saliva samples kept for 45 days, while DNA recovery increased in semen samples. These different observations can be explained by the amount of nucleated and DNA-containing cells that biological samples contain by their nature.

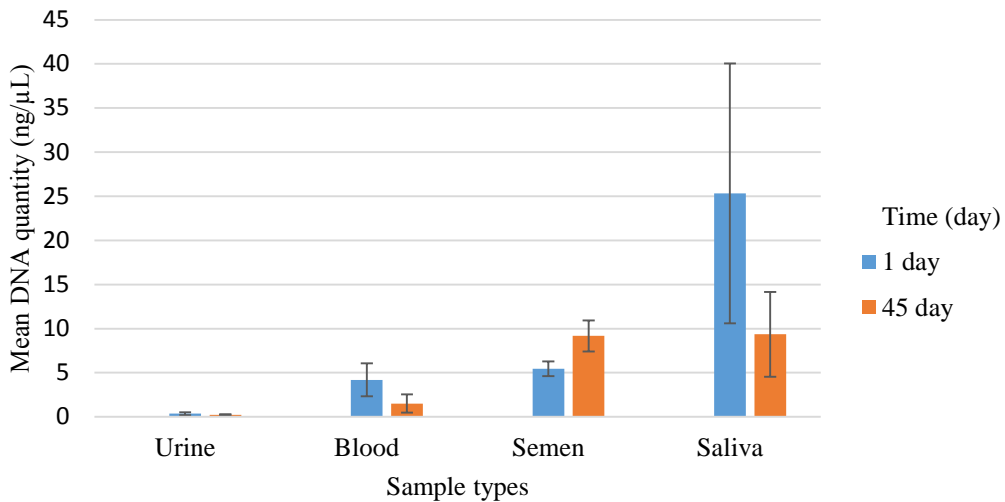


Fig. 2. Time-dependent DNA amount (ng/μL) obtained from different biological fluids regardless of FD difference.

When the DNA amounts (ng/μL) obtained from samples containing different biological fluids, treated with CV and SS FD techniques and kept for 1 to 45 days were evaluated in terms of chemical exposure (Figure 3), compared to the control group, both CV and SS treatment caused an increase in DNA recovery in 1-day samples, while CV

treatment caused an overall increase in DNA recovery in samples kept for 45 days, while SS treatment decreased DNA recovery. While more DNA amount was obtained with SS in fresh findings, the amount of DNA obtained at the end of the 45th day decreased significantly. This decrease in DNA recovery is not expected to significantly affect the quality of DNA profiles.

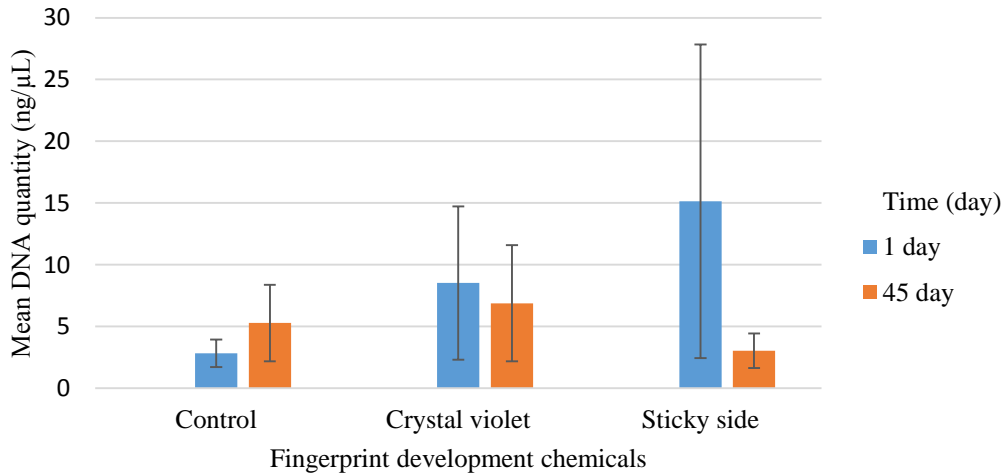


Fig. 3. Variation of DNA amount (ng/µL) depending on fingerprint development methods and time.

The SS method, which is effective on adhesive surfaces, has been used to enhance fingerprints on the adhesive surfaces of tapes [28]. SS combined with Un-du reagent has been observed to negatively affect the amount of DNA in blood fingerprints. However, if SS is applied alone, it is possible to obtain a profile [29]. In a study where bloodstains left on different surfaces at different time intervals were treated with fingerprint chemicals, it was noted that SS can be used for fingerprint development when it is ensured that there is sufficient DNA for analysis [30]. Au et al. developed bloody fingerprints on dark surfaces with white SS powder. They observed that SS reduced the amount of DNA [27]. In this study, similar to Au et al., it was found that the SS application and the 45-day waiting period negatively affected DNA recovery from biological samples.

CV, which becomes visible by attaching to fatty compounds in latent fingerprints, gives the fingerprint a purple appearance as a result of the process. In a study by Lennard et al., fingerprints and bloodstains were left on tape, and fingerprints were made visible by CV [28]. It was observed that CV did not have a negative effect on DNA typing. In particular, no negative effects were observed for DNA extraction, DNA quantification or typing in samples stored dry at room temperature [31]. Treatment of traces on adhesive surfaces with crystal violet did not affect STR analyses [28]. In a study in which dried bloodstains were treated with various reagents, it was observed that crystal violet did not reduce the amount of DNA [32]. PCR-based DNA typing of a single bloody fingerprint developed with crystal violet was successful [29]. The data obtained are parallel to this study. In the present study, CV application increased the DNA recovery rate in both fresh and aged samples. Furthermore, this study shows that in biological fluids other than semen (urine, blood and saliva), the amount of DNA required for identification can be obtained, although there is a decrease in DNA recovery over time.

Due to the sensitivity of biological samples, biological examinations are usually given priority in the examination of evidence. In such evidence, the biological sample taken in order not to damage the fingerprints on the evidence may be insufficient, and a sufficient amount of DNA cannot be obtained for genetic analyses. To obtain results from DNA analyses performed on the evidence, efforts to collect a substantial amount of biological samples often lead to

the compromising or damaging of fingerprint samples present on the same piece of evidence. For this reason, various precautions and caution should be taken when collecting fingerprints contaminated with other body fluids from the crime scene.

Based on the data obtained from the study, the extraction kit employed has demonstrated its capability to isolate a sufficient amount of DNA from adhesive surfaces treated with FD chemicals, suitable for forensic genotyping and phenotyping analyses. It was determined with the analyses that the latent FD chemicals used on adhesive tape surfaces, which are non-porous surfaces, do not affect DNA recovery at a level that would prevent STR profiling. According to the results obtained, it was determined that DNA recovery was high for all biological sample types obtained from adhesive tape in terms of surface type. These results indicate that forensic genetic analyses for DNA recovery can be performed on adhesive tape surfaces treated with fingerprint enhancement chemicals. Over the past two decades, most studies have used biological fluids, such as bloodstains or saliva, as DNA sources to examine the effect of fingerprint treatments on DNA analysis [33]–[35], while other more recent studies have used volunteer fingerprints as sources traces [13], [36]–[39]. On the contrary, there is no study in which blood, urine, semen and saliva samples were used together. In this sense, the results do not sufficiently overlap with similar findings in the literature because a study has not yet been to determine DNA recovery at two different time intervals after applying FD to four different body fluids on the adhesive surface.

4. Conclusion

It is essential to investigate whether the fingerprint development processes performed on biological evidence obtained in crime scene investigation disrupt the structure of biological materials and to determine which of the fingerprint development methods that provide the same function while providing fingerprint development can be used without disrupting the structure of DNA. This scientific gap in the literature also prevents criminal laboratories from establishing standardised methods for the genetic examination of FD-applied adhesive surfaces, which are also valuable biological evidence due to the DNA they contain. When considered worldwide, it is seen that there are few studies targeting this issue [2]. Moreover, these studies are far from being systematic, and they will unlikely standardise the data they present. Therefore, in the present study, basic steps towards system validation have been taken as a reference and guideline for all future studies in this field.

This study paves the way for an approach to determine how latent fingerprinting chemicals affect DNA recovery from different biological fluids. DNA was successfully extracted from all FD-treated adhesive tape surfaces containing different biological fluids, and DNA amounts were measured in all of them. Longer-term investigations are also required to address how other factors may affect DNA recovery, including the time between latent fingerprinting and DNA analysis, as well as the modification of initial DNA amounts by surface types and the determination of this change.

Acknowledgements

We sincerely thank both Cukurova University Biotechnology Centre Laboratory and Kutahya Health Sciences University, Application and Research Centre Laboratory. Their provision of laboratory facilities and a conducive working environment greatly facilitated the execution of this study.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this article. No financial or personal relationships with third parties exist that could unduly influence, or be perceived to influence, the content of this article.

Author contributions

The authors declare that they have all made equal contributions to the work presented in this article.

Funding

This study was financially supported by the Scientific Research Fund Kutahya Health Sciences University, under Grant Number FBA-2021-72.

Ethical approval

All experimental protocols were conducted in accordance with relevant guidelines and regulations and were approved by the Ethics Committee of Kutahya Health Sciences University (Approval No: 2020/09-05).

References

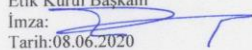
- [1] C. Neumann and H. Stern, "Forensic Examination of Fingerprints: Past, Present, and Future," *Chance*, vol. 29, no. 1, pp. 9–16, Jan. 2016, doi: 10.1080/09332480.2016.1156353.
- [2] P. Kumar, R. Gupta, R. Singh, and O. P. Jasuja, "Effects of latent fingerprint development reagents on subsequent forensic DNA typing: A review," *J. Forensic Leg. Med.*, vol. 32, pp. 64–69, May 2015, doi: 10.1016/j.jflm.2015.03.002.
- [3] S. P. Singh, "Development of Latent Finger Prints on Human Skin: A Review," *Int. J. Eng. Res. Technol.*, vol. 9, no. 6, pp. 1192–1198, 2020.
- [4] B. Yamashita and M. French, "Chapter 7: Latent Print Development," in *Fingerprint Sourcebook*, US Dept. of Justice, Office of Justice Programs, National Institute of Justice, 2010, pp. 1–68.
- [5] S. M. Bleay, R. S. Croxton, and M. De Puit, *Fingerprint development techniques: theory and application*, First. Wiley Press, 2018.
- [6] ICRC 2009: International Committee of Red Cross (ICRC), *Missing People, DNA Analysis and Identification of Human Remains: A Guide to Best Practice in Armed Conflicts and Other Situations of Armed Violence*, Second Edition. 2009.
- [7] Y. Gülekçi, "Parmak İzi Araştırmaları," in *Suç Araştırmalarında Kriminal Yaklaşımlar*, First., Akademisyen Press, 2021, pp. 185–214.
- [8] I. Feine *et al.*, "Acetone facilitated DNA sampling from electrical tapes improves DNA recovery and enables latent fingerprints development," *Forensic Sci. Int.*, vol. 276, pp. 107–110, Jul. 2017, doi: 10.1016/j.forsciint.2017.04.023.
- [9] H. Lee, J. Yim, and Y. Eom, "Effects of fingerprint development reagents on subsequent DNA analysis," *Electrophoresis*, vol. 40, no. 14, pp. 1824–1829, Jul. 2019, doi: 10.1002/elps.201800496.
- [10] R. Shalhoub *et al.*, "The recovery of latent fingermarks and DNA using a silicone-based casting material," *Forensic Sci. Int.*, vol. 178, no. 2–3, pp. 199–203, Jul. 2008, doi: 10.1016/j.forsciint.2008.04.001.
- [11] M. Carlin, R. Nickel, K. Halstead, J. Viray, A. Hall, and A. Ehrlich, "Quantifying DNA loss in laboratory-created latent prints due to fingerprint processing," *Forensic Sci. Int.*, vol. 344, p. 111595, Mar. 2023, doi: 10.1016/j.forsciint.2023.111595.
- [12] A. Gosch and C. Courts, "On DNA transfer: The lack and difficulty of systematic research and how to do it better," *Forensic Sci. Int. Genet.*, vol. 40, pp. 24–36, May 2019, doi: 10.1016/j.fsigen.2019.01.012.
- [13] J. Sewell *et al.*, "Recovery of DNA and fingerprints from touched documents," *Forensic Sci. Int. Genet.*, vol. 2, no. 4, pp. 281–285, Sep. 2008, doi: 10.1016/j.fsigen.2008.03.006.
- [14] M. . Schulz and W. Reichert, "Archived or directly swabbed latent fingerprints as a DNA source for STR typing," *Forensic Sci. Int.*, vol. 127, no. 1–2, pp. 128–130, Jun. 2002, doi: 10.1016/S0379-0738(02)00092-0.
- [15] D. Färber, A. Seul, H. Weisser, and M. Bohnert, "Recovery of Latent Fingerprints and DNA on Human Skin*," *J. Forensic Sci.*, vol. 55, no. 6, pp. 1457–1461, Nov. 2010, doi: 10.1111/j.1556-4029.2010.01476.x.
- [16] S. Norlin, M. Nilsson, P. Heden, and M. Allen, "Evaluation of the impact of different visualization techniques on DNA in fingerprints," *J. Forensic Identif.*, vol. 63, no. 2, pp. 189–204, 2013.
- [17] A. D. Solomon, M. E. Hytinen, A. M. McClain, M. T. Miller, and T. Dawson Cruz, "An Optimized DNA Analysis Workflow for the Sampling, Extraction, and Concentration of DNA obtained from Archived Latent Fingerprints," *J. Forensic Sci.*, vol. 63, no. 1, pp. 47–57, Jan. 2018, doi: 10.1111/1556-4029.13504.
- [18] Z. Subhani, B. Daniel, and N. Frascione, "DNA Profiles from Fingerprint Lifts—Enhancing the Evidential Value of Fingermarks Through Successful DNA Typing," *J. Forensic Sci.*, vol. 64, no. 1, pp. 201–206, Jan. 2019, doi: 10.1111/1556-4029.13830.
- [19] S. M. Goschay *et al.*, "Finger mark development techniques within scope of ISO 17025," in *Fingerprint Source Book, Home Office Centre for Applied Science and Technology (CAST)*, 2012, pp. 233–289.
- [20] Qiagen, "QIAamp DNA Micro Handbook," *Qiagen Sample and Assay Technologies*, 2013. https://www.qiagen.com/us/resources/resourcedetail?id=085e6418-1ec0-45f2-89eb-62705f86f963&lan_g=en (accessed Aug. 27, 2023).
- [21] Thermo Fisher Scientific Inc., "Quant-iT™ dsDNA Assay Kits, high sensitivity (HS)," *Invitrogen™*, 2006. <https://www.thermofisher.com/order/catalog/product/Q33120> (accessed Aug. 27, 2023).

- [22] H. C. Lee and R. E. Gaensslen, "Methods of latent fingerprint development," in *Advances in Fingerprint Technology*, CRC Press, 2001, pp. 105–175.
- [23] G. S. Bumbrah, R. M. Sharma, and O. P. Jasuja, "Emerging latent fingerprint technologies: a review," *Res. Reports Forensic Med. Sci.*, vol. 6, pp. 39–50, 2016.
- [24] M. Wang, M. Li, A. Yu, Y. Zhu, M. Yang, and C. Mao, "Fluorescent Nanomaterials for the Development of Latent Fingerprints in Forensic Sciences," *Adv. Funct. Mater.*, vol. 27, no. 14, Apr. 2017, doi: 10.1002/adfm.201606243.
- [25] P. Kanokwongnuwut, K. P. Kirkbride, H. Kobus, and A. Linacre, "Enhancement of fingermarks and visualizing DNA," *Forensic Sci. Int.*, vol. 300, pp. 99–105, Jul. 2019, doi: 10.1016/j.forsciint.2019.04.035.
- [26] S. Gino and M. Omedei, "Effects of the most common methods for the enhancement of latent fingerprints on DNA extraction from forensic samples," *Forensic Sci. Int. Genet. Suppl. Ser.*, vol. 3, no. 1, pp. e273–e274, Dec. 2011, doi: 10.1016/j.fsigss.2011.08.133.
- [27] C. Au, H. Jackson-Smith, I. Quinones, B. J. Jones, and B. Daniel, "Wet powder suspensions as an additional technique for the enhancement of bloodied marks," *Forensic Sci. Int.*, vol. 204, no. 1–3, pp. 13–18, Jan. 2011, doi: 10.1016/j.forsciint.2010.04.044.
- [28] C. Lennard, "Fingerprint detection and identification: current research efforts," *Aust. J. Forensic Sci.*, vol. 46, no. 3, pp. 293–303, Jul. 2014, doi: 10.1080/00450618.2013.839743.
- [29] T. Spear, N. Khoshkebari, J. Clark, and M. Murphy, "Summary of experiments investigating the impact of fingerprint processing and fingerprint reagents on PCR-based DNA typing profiles," 2015.
- [30] C. Roux, K. Gill, J. Sutton, and C. Lennard, "A further study to investigate the effect of fingerprint enhancement techniques on the DNA analysis of bloodstains," *J. Forensic Identif.*, vol. 49, no. 4, pp. 357–376, 1999.
- [31] C. Stein, S. Kyeck, and C. Henssge, "DNA Typing of Fingerprint Reagent Treated Biological Stains," *J. Forensic Sci.*, vol. 41, no. 6, pp. 1012–1017, Nov. 1996, doi: 10.1520/JFS14039J.
- [32] H. C. Lee; *et al.*, "Effect of Presumptive Test, Latent Fingerprint and Some Other Reagents and Materials on Subsequent Serological Identification, Genetic Marker and DNA Testing in Bloodstains," *J. Forensic Identif.*, vol. 39, no. 6, pp. 339–358, 1989.
- [33] M. Azoury, A. Zamir, C. Oz, and S. Wiesner, "The Effect of 1,2-Indanedione, a Latent Fingerprint Reagent on Subsequent DNA Profiling," *J. Forensic Sci.*, vol. 47, no. 3, pp. 586–588, May 2002, doi: 10.1520/JFS2001150.
- [34] C. Frégeau, O. Germain, and R. Fournay, "Fingerprint Enhancement Revisited and the Effects of Blood Enhancement Chemicals on Subsequent *Profiler Plus*TM Fluorescent Short Tandem Repeat DNA Analysis of Fresh and Aged Bloody Fingerprints," *J. Forensic Sci.*, vol. 45, no. 2, pp. 354–380, Mar. 2000, doi: 10.1520/JFS14688J.
- [35] A. Zamir, C. Oz, and B. Geller, "Threat Mail and Forensic Science: DNA Profiling from Items of Evidence After Treatment with DFO," *J. Forensic Sci.*, vol. 45, no. 2, pp. 445–446, Mar. 2000, doi: 10.1520/JFS14704J.
- [36] M. K. Balogh, J. Burger, K. Bender, P. M. Schneider, and K. W. Alt, "STR genotyping and mtDNA sequencing of latent fingerprint on paper," *Forensic Sci. Int.*, vol. 137, no. 2–3, pp. 188–195, Nov. 2003, doi: 10.1016/j.forsciint.2003.07.001.
- [37] P. Leemans, A. Vandeput, N. Vanderheyden, J.-J. Cassiman, and R. Decorte, "Evaluation of methodology for the isolation and analysis of LCN-DNA before and after dactyloscopic enhancement of fingerprints," *Int. Congr. Ser.*, vol. 1288, pp. 583–585, Apr. 2006, doi: 10.1016/j.ics.2005.09.079.
- [38] P. Tozzo, A. Giuliadori, D. Rodriguez, and L. Caenazzo, "Effect of Dactyloscopic Powders on DNA Profiling From Enhanced Fingerprints," *Am. J. Forensic Med. Pathol.*, vol. 35, no. 1, pp. 68–72, Mar. 2014, doi: 10.1097/PAF.000000000000081.
- [39] A. Zamir, E. Springer, and B. Glattstein, "Fingerprints and DNA: STR Typing of DNA Extracted from Adhesive Tape after Processing for Fingerprints," *J. Forensic Sci.*, vol. 45, no. 3, pp. 687–688, May 2000, doi: 10.1520/JFS14749J.

Appendix

T.C
KÜTAHYA SAĞLIK BİLİMLERİ ÜNİVERSİTESİ REKTÖRLÜĞÜ
GİRİŞİMSEL OLMAYAN KLİNİK ARAŞTIRMALAR ETİK KURULU
KARAR FORMU

ARAŞTIRMANIN AÇIK ADI		Parmak İzi Geliştirme Yöntemleri Uygulanan Biyolojik Materyallerin Bütünlüğünün Değerlendirilmesi ve Söz Konusu Yöntemlerin DNA Analizine Etkilerinin Araştırılması
ETİK KURUL BİLGİLERİ	ETİK KURULUN ADI	Kütahya Sağlık Bilimleri Üniversitesi Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu
	KURUL ADRESİ	Kütahya Sağlık Bilimleri Üniversitesi Evliya Çelebi Yerleşkesi Tavşanlı Yolu 10. Km KÜTAHYA
	TELEFON	(0 274) 260 00 43 / 1139
	FAKS	(0 274) 265 22 85
	E-POSTA	etik.gir.olmayan@ksbu.edu.tr
BAŞVURU BİLGİLERİ	KOORDİNATÖR/SORUMLU ARAŞTIRMACI UNVANI/ADI/SOYADI	Dr.Öğrt.Üyesi Yakup GÜLEKÇİ
	KOORDİNATÖR/SORUMLU ARAŞTIRMACININ UZMANLIK ALANI	Adli Bilimler Anabilim Dalı
	KOORDİNATÖR/SORUMLU ARAŞTIRMACININ BULUNDUĞU MERKEZ	Kütahya Sağlık Bilimleri Üniversitesi Mühendislik ve Doğa Bilimleri Fakültesi
	YARDIMCI ARAŞTIRMACI VE BÖLÜMÜ	Dr.Öğrt.Üyesi Harun ŞENER- Kütahya Sağlık Bilimleri Üniversitesi Mühendislik ve Doğa Bilimleri Fakültesi Adli Bilimler Anabilim Dalı Öğr.Grv.Dr.Fatma ÇAVUŞ YONAR-Adli Bilimler Uzmanı-Istanbul Üniv. Cerrahpaşa Adli Tıp ve Adli Bilimler Ens.
ARAŞTIRMANIN TÜRÜ	Gözlemsel, Tanımlayıcı	
KARAR BİLGİLERİ	Karar No: 2020/09-05	Tarih: 04.06.2020
	Başvuru dosyası ile ilgili belgeler araştırmanın gerekçe, amaç, yaklaşım ve yöntemleri dikkate alınarak incelenmiş ve uygun bulunmuş olup araştırmanın başvuru dosyasında belirtilen merkezde gerçekleştirilmesinde etik ve bilimsel sakınca bulunmadığına oy birliği ile karar verilmiştir.	

Prof.Dr.Duygu PERÇİN RENDERS
Etik Kurul Başkanı
İmza: 
Tarih:08.06.2020



E-ISSN: 2687-6167

Number 56, March 2024

RESEARCH ARTICLE

Receive Date: 16.01.2024

Accepted Date: 20.02.2024

Supplier selection using the integrated MEREC – CoCoSo methods in a medical device company

Gülnehal Özel Sönmez^a, Pelin Toktaş^{b,*}

^aBaşkent University, Institute of Science, Quality Engineering Program, Ankara, 06790, Türkiye, ORCID: 0000-0001-6622-4646

^bBaşkent University, Faculty of Engineering, Industrial Engineering Department, Ankara, 06790, Türkiye, ORCID: 0000-0002-8503-2321

Abstract

The medical device industry is a rapidly developing industry that includes various dynamics. Developed technologies show continuous improvement depending on diagnosis and treatment applications in health services. To keep up with this change and survive in an increasingly competitive environment, medical device manufacturers must be engaged in continuous improvement activities. This situation, necessary for many companies producing in the industrial field, gains even more importance in the medical device sector when the direct impact of product safety and quality on human life is considered. In companies producing medical devices, the legal requirements of the product being a medical device are followed by the notified bodies and authorized authorities within the framework of standards and regulations within the scope of quality processes. Increasing costs and liabilities with MDR 2017/45 have pushed medical device manufacturers to question their methods. In this study, it was determined that customer requests could not be met in a company producing medical devices, and it was observed that delivery times increased. In evaluating the reasons for the increase in delivery time, it was determined that supplier selection could have been carried out more effectively. For this purpose, six suppliers and six criteria were selected because of the company's sector knowledge and the evaluations of the company managers. The Combined Compromise Solution (CoCoSo) method, one of the new generation multi-criteria decision-making (MCDM) methods, is proposed for ranking the suppliers in the supplier selection problem. Method The removal effects of criteria (MEREC) weighting method was used to weight supplier selection criteria. In this study, a new generation supplier selection method application in medical devices has been carried out. Considering the inadequacy of the studies on supplier selection in medical devices, the relevant research will contribute to the literature.

© 2024 DPU All rights reserved.

Keywords: MEREC; CoCoSo; Supplier Selection; Medical Device; MCDM

* Corresponding author. Tel.: +90 312 246 6666 / 4092.

E-mail address: ptoktas@baskent.edu.tr

<http://dx.doi.org/10.1016/j.cviu.2017.00.000>

1. Introduction

The medical device industry is a rapidly developing sector in the world that includes various dynamics. Developed technologies are improving depending on diagnosis and treatment applications in healthcare services. Medical device manufacturers must engage in continuous improvement activities to keep up with this change and maintain their existence in an increasingly competitive environment. For processes to be carried out efficiently and production activities to provide precise and effective results, they must engage in continuous improvement activities. The company must review its procedures, take the necessary actions, implement them when detecting a problem, and then control the impact and efficiency of its activities. This process, essential for many companies producing in the industrial field, becomes even more critical in the medical device sector, considering the direct impact of product safety and quality on human life.

It is monitored by notified bodies and competent authorities within the framework of legal requirements, quality standards, and regulations imposed by a product being a medical device. When placing their devices on the market or putting them into service, the manufacturer must ensure that their devices are designed and manufactured by the requirements of the relevant regulation. The manufacturer's obligations are specified in the applicable EU legislation and the rules published by the Ministry of Health in compliance with this legislation. After the Medical Device Regulation 2017/45 (MDR 2017/45), these requirements have increased even more, and quality management has gained critical importance. Rising costs and liabilities with MDR 2017/45 have made medical device manufacturers question their processes. Meeting customer demands while fulfilling all legal requirements becomes critical for manufacturers to avoid market loss. The harmony between the manufacturer's activities and customer expectations is crucial for the operation's success.

According to the Medical Device Regulation published in the Legal Gazette dated 02.06.2021 and numbered 31499, products that are used on humans with an indication as part of a diagnosis or treatment and show their effects by physical or mechanical means in the areas where they are used are considered medical devices. Their definition states that medical devices are high-risk products potentially affecting human life and public health. For this reason, these devices are monitored by approved organizations and competent authorities within the legal requirements, standards, and regulations of being a medical device. Manufacturers must fulfill the requirements of the relevant EU legislation to ensure the free movement of their products among European Union (EU) member countries [1].

Medical devices and equipment are essential for quality healthcare. For this purpose, medical device manufacturers must produce according to the relevant legislation and standard requirements. Effective execution of a quality management system has always been considered a good strategy for a manufacturer to maintain and improve product and service quality [2].

The quality management system in medical devices is carried out by the "ISO 13485 - Medical Devices Management System" standard. This standard can be defined as the harmonized version of the "ISO 9001 - Quality Management System" standard for medical devices. It sets forth quality system requirements for organizations operating in the medical sector. Effective execution of the quality management system directly affects product safety and performance.

Manufacturers need to meet customer demands without compromising product safety and performance. In medical device production, product safety is evaluated as a whole, considering patient and user health, and the process from raw material purchase to sales and subsequent follow-up, depending on the product class, is evaluated as traceability. Traceability begins when the order arrives and continues until its sale and destruction. The first critical step is the supply of raw materials/semi-finished products. Supplier selection is a factor that directly affects product quality and safety. Raw material/semi-finished product supply uses a large portion of the company's resources and is very important for the organization's success.

On the other hand, the fact that the product is a medical device brings some legal requirements. These requirements are supervised by competent authorities and notified bodies. After MDR 2017/45, these requirements

increased even more, and supplier selection became critical. Therefore, the organization's success is measured not only by its sales figures and profit rates but also by its product quality and traceability success.

Although there are studies on supply chain management in the healthcare sector in the literature research, it has been seen that studies on supply chain management in the medical device industry need to be more comprehensive. It was concluded that there is an area in need of improvement.

Supplier selection in supply chain management may vary according to the needs, expectations and vision of the sector and the company. Correctly determining the criteria is a very critical step for successful supplier selection. Some supplier selection articles were examined, and some selected studies were evaluated in Table 1 based on field, methods, and criteria.

Table 1. Examples of some supplier selection studies.

Author	Field	Methods	Criteria
Vipul Jain et al., 2016 [3]	Automotive	AHP and TOPSIS	Product quality, price, relationship quality, production capacity, on-time delivery, warranty, environmental performance, supplier brand name.
Fallahpour et al., 2017 [4]	Textile	Questionnaire, FPP, and FTOPSIS	Cost, quality, supply and service, flexibility, environmental management system, green product, green storage, eco-design, green transportation, green technology, employee rights, occupational health and safety, supporting activities.
Song et al., 2017 [5]	Solar Energy and Air Conditioning	DEMATEL	Economic criteria: Quality, delivery, cost price. Environmental criteria: Environmental management system, resource consumption, environmental design, reduce-reuse and recycle. Social Criteria: Occupational health and safety, employee rights and welfare, education, and community development.
Yazdani et al., 2017 [6]	Food Industry	DEMATEL, QFD and COPRAS	Price, quality, delivery performance, branding, reputation, flexibility, cost
Banaeian et al., 2018 [7]	Agriculture-Food Industry	TOPSIS, VIKOR and GRA	On-time delivery, after-sales service, supply capacity, quality, price
Abdel-Basset et al., 2018 [8]	Logistics	DEMATEL	Cost, delivery time, quality, innovation, reputation, location, response to customers
Yazdani et al., 2019 [9]	Construction	DEMATEL, BWM, and CoCoSo-G	Design, greenhouse gas pollution, distribution and flexibility, responsiveness and communication, finances, price offered, environmental management system.
Stević et al., 2020 [10]	Healthcare industry	MARCOS	Price, quality, product range, on-time delivery, innovation, organization and management, reliability, reputation, occupational and worker safety, information supply, employee rights, training, compliance with the law, environmental qualifications, environmental management system, recycling, pollution control, green R&D, green product, number of ISO Standards held.
Yazdani et al., 2020 [11]	Healthcare Industry	DEMATEL and BWM	Price, inventory capacity, batch volume, flexibility, use of technology, quality.
Sumrit, 2020 [12]	Healthcare Industry	VMI, Fuzzy DELPHI, SWARA and COPRAS	Past delivery performance, corporate trust, investment cost, information exchange and sharing, continuous improvement, supply chain process integration, information technology readiness, allocated resources, spatial complexity, prior knowledge and experience, risk/reward sharing
Göncü and Çetin,	Healthcare	DEMATEL and	Price, quality, logistics, and occupational health and safety

2022 [13]	business	ANP	
Nguyen et al., 2022 [14]	Steel manufacturing industry	Data Envelopment Analysis, SF-AHP and SF-WASPAS	Economic criteria: Production facilities, technological and financial competence, delivery time, flexibility, transportation cost, delivery, product price, quality. Environmental criteria: Environmental costs, Environmental Management Systems, Green R&D and innovation. Social Criteria: Health and safety, customer satisfaction, satisfaction of stakeholders, interests, and rights of employees

When it comes to the medical device industry, it is essential to effectively carry out the steps that affect product quality and safety, and this field has become a research subject for researchers.

The medical device industry has been one of the sectors whose importance has been most emphasized after the COVID-19 pandemic. This situation has pushed researchers to work on the problems experienced in the supply chain. The main problem encountered in the supply chain is the supplier selection problem.

In their study in 2014, Ghadimi and Heavey mentioned that supplier selection needs to be addressed in the medical device industry in the literature. They informed us that the criteria in the literature may be less than 100% suitable for supplier selection in the medical device industry. They conducted a study on sustainable supplier selection. Ghadimi and Heavey emphasized in their study that the supplier selection problem is the most crucial research topic for supply chain management. He suggested that the low number of studies on supplier selection in the medical device industry is due to companies' privacy concerns. In their research, the steps of the supplier selection process are applied, such as identifying potential suppliers, determining criteria and sub-factors, collecting company requirements, classifying the gathered requirements, and ranking potential suppliers. They evaluated sustainable supplier selection under the leading environmental, economic, and social headings. In addition, they assessed the requirements of the medical device manufacturer company ISO 13485-Quality Management System Standard under the headings of regulation, company responsibilities on the product, after-sales traceability, storage, and logistics [15].

In the study conducted by Ghadimi et al. in 2019, they developed the Multi-Agent System (MAS) method and demonstrated the application of this method on a medical device manufacturer [16].

Today, companies' perspectives have changed along with consumers' changing attitudes within the framework of the legal obligations to which the medical device industry is bound, and it has been understood that more than studies focused on financial profit are needed.

In his study conducted in 2023, Forouzehnejad concluded that agility and sustainability are the essential criteria for agility and sustainability in the medical device industry and that flexibility, cost, reliability, smart factory, and quality are crucial criteria determined subsequently [17].

All companies manufacturing in the medical device industry must comply with the ISO 13485- Quality Management System Standard. This standard requires manufacturers to select suppliers from a pre-approved list. This restricts manufacturers, meaning companies can only go up to the approved suppliers accepted by authorized organizations. Therefore, supplier selection is very critical in the medical device industry. Decision-making in the supplier selection process is challenging because many criteria are used, and these criteria conflict with each other.

In supplier selection, it takes work to make the right choice among various selection criteria and multiple alternatives. Multi-criteria decision-making (MCDM) methods offer appropriate solutions to problems with more than one criterion and alternative. In this study of the supplier selection problem, it is suggested to use the Method the Removal Effects of Criteria (MERECE) method to weigh the supplier selection criteria and the Combined Compromise Solution (CoCoSo) method to rank the suppliers.

The reasons for choosing MERECE and CoCoSo methods in this study are as follows. In both methods, there is no need for decision-makers. All calculations are effectively made according to the actual values revealed by the initial

decision matrix. In a sensitivity analysis of the MEREC method, Keshavarz-Ghorabae et al. found that the MEREC method provides stable results compared to objective weighting methods such as Entropy and CRITIC [18]. In the CoCoSo method, the final ranking of suppliers is not based on just one evaluation score but rather on a score obtained by combining three different evaluation scores in one calculation. The literature review's conclusions indicate that companies that manufacture medical devices did not use the integrated MEREC-CoCoSo approaches when choosing their suppliers.

The MEREC approach was first introduced as an objective criterion weighting method in 2021 by Keshavarz-Ghorabae et al. Using criterion weights, this method ascertains the impact of each criterion on the overall performance of the alternatives.

The overall and partial performance of the alternatives is also measured using a logarithmic function. MEREC assigns higher priority to criteria that significantly affect how well other options perform. MEREC can be used to determine the weights for each criterion while accounting for variations in each alternative's performance. If there is more variation in the performance of a criterion, that criterion is given a higher weight [18]. The MEREC method has become a research subject in many areas since its debut article. Some of these studies are the following:

While the MEREC method was used for criterion weighting in the coal supplier selection problem for the thermal power plant, the suppliers were ranked using the Multiple Criteria Ranking by Alternative Trace (MCRAT) method [19]. In their study, Trung and Think conducted sixteen experiments on the turning process based on the principle of using a cutting tool on the workpiece rotating around its own axis. In their experiments, they used four different MCDM methods to evaluate the effects of cutting speed, feed rate, and depth of cut parameters on surface smoothness and material removal rate and determined the weights of the criteria with Entropy and MEREC [20].

In the study conducted to decide the locations of distribution centers, Keshavarz-Ghorabae used the Stepwise Weight Assessment Ratio Analysis II (SWARA II), Weighted Aggregated Sum Product Assessment (WASPAS) and MEREC methods together. The study proceeded through a decision matrix containing weights determined in two different ways, subjectively and objectively [21]. Nguyen et al. studied the MCDM problem using the MARCOS, TOPSIS and MAIRA methods, and used the MEREC method for criterion weighting [22].

In a different study carried out that same year, Nicolalde et al. determined the criterion weights using the Entropy and MEREC methods and selected the material to be used on a vehicle's roof using the Višekriterijumsko Kompromisno Rangiranje (VIKOR), Complex Proportional Assessment (COPRAS), and TOPSIS methods [23].

The use of robots is becoming more common in industrial settings, as noted by Shanmugasundar and his colleagues, who also discussed the challenge of selecting the right kinds of robots. The MEREC method was used to weigh the criteria, and the Combination Distance-based Assessment (CODAS), COPRAS, CoCoSo, Multi-attributive Border Approximation Area Comparison (MABAC), and VIKOR methods were used to finish the decision-making process [24]. In the Yu et al. study from 2022, which highlighted sustainability. The effectiveness of the method proposal was evaluated through a case study [25]. In a study by Do and Nguyen, they talked about the importance of the criterion weighting step in decision-making processes and used five different weighting methods, namely MEREC, EQUAL, ROC, RS and FUCOM, along with the CoCoSo, MABAC, MAIRCA, EAMR and TOPSIS methods in the solution ordering problem in the turning process [26]. Saidin et al presented a study on fuzzy MEREC. The fuzzy MEREC method, which they offer as a solution to the uncertainty problem experienced in MCDM problems, has been exemplified by using it in the evaluation of academic personnel [27].

Ayan et al. conducted a study comparing various weighting methods, including the MEREC method. The authors mentioned that the most effective criterion among the criteria in the problem under consideration in the MEREC method has the highest level of importance. In addition, he showed that one of the differences with other methods is that the decision matrix values must be greater than zero. He stated that the biggest advantage is ease of application because the calculations are easy to understand and simple [28]. In a study by Chaurasiya and Jain on information technology service selection in health services, the weights in the best software selection problem were determined by MEREC and SWARA, and the ranking of the alternatives was obtained by using MAR [29].

In this study, after obtaining supplier selection criterion weights with MEREC, it was suggested to use the CoCoSo method, first introduced by Yazdani et al. in 2019, to rank the suppliers. The CoCoSo method measures the distance of each performance level of the alternatives from the ideal. When making the final alternative ranking, three different scores are calculated [30]. Although CoCoSo method is a new-generation MCDM method, CoCoSo method has been the subject of study by many researchers.

Ecer et al. used the CoCoSo method to evaluate the organization of oil exporting countries according to 41 sustainable development criteria in 10 dimensions. They compared the results with existing multi-criteria decision-making methods and demonstrated the effectiveness of CoCoSo [31]. In the study emphasizing the importance of stock management, the CoCoSo method was applied together with ABC analysis and FUCOM (Full Consistency Method). The ranking of supplier alternatives was obtained by the CoCoSo method [32]. Wen et al. evaluated cold chain suppliers to keep the logistics risks of medicines to a minimum. To solve this problem, they conducted a study in which they used the SWARA method to perform criterion weighting and the CoCoSo methods to make the final ranking of suppliers. In this study, which also included method application, they emphasized that the CoCoSo method gave reliable results [33].

Drawing attention to the importance of supplier selection in industrial organizations, Zolfani et al. They implemented an application in the steel industry where they used the Best Worst Method (BWM) and CoCoSo methods together [34]. Bagal et al. examined the effects of marble dust and fly ash on concrete mixtures. They used CODAS and CoCoSo methods to optimize process parameters. The study results were evaluated on the pairwise comparison matrix [35]. Barua et al. They presented a study in which they evaluated the mechanical behavior of hybrid natural fiber reinforced nano-sic particle composite. In this study, it was aimed to contribute to the decision-making process of process parameters with the CoCoSo method [36]. In the study carried out in the field of automobile industry, five alternative vehicles with different specific features were evaluated for the most suitable passenger car selection. While the criterion weighting is obtained by the CRITIC method, the alternatives are ranked by CoCoSo. The results have been validated with other MCDM methods [37].

Emphasizing the importance of the supply chain, Biswas addressed the supply chain problem of a healthcare institution operating in India in his study in 2020. In this study, while determining the criterion weights with PIPRECIA (PIVot Pairwise RELative Criteria Importance Assessment), three different rankings were obtained using MABAC, MARCOS and CoCoSo. These three methods have been shown to give consistent results [38]. In another study on supply chain management, weights were determined with fuzzy BWM, while the most appropriate supplier selection was ensured with CoCoSo and Bonferroni integration [39]. In the study conducted by Peng et al. in 2020, they emphasized the importance of evaluating financial risks and used the CRITIC and CoCoSo methods together to solve the comparison problem they encountered in this field [40]. Addressing the problem of selection of logistics centers, M.Yazdani et al. used the CoCoSo method integrated with data envelopment to choose among logistics service providers [41]. Altıntaş evaluated the knowledge performances of G7 countries according to the 2020 Global Knowledge Index (GKI) components using CoCoSo [42].

The method proposed in this study was applied to the supplier selection problem of a medical consumables manufacturer company that has been operating for more than 25 years and is in preparation for MDR 2017/45. It has been determined that the company that produces according to order cannot deliver its orders on time. Infusion and extension sets are made in this company. Extension sets are used as liquid carrier connection elements for transferring a medical liquid (such as serum, medicine, or vitamin) from a liquid source to a catheter to enter the liquid into the body and/or for collecting body fluids in a collection container from a catheter already attached to the human body. Infusion sets are used to safely transfer intravenous drugs and fluids to patients in all healthcare institutions that provide critical care and where IV drug administration is performed. The hose is the most vital part used in extension and infusion sets. As a result of the current situation analysis, it was determined that production delays occurred because the supplied hoses needed to be delivered on time and arrived dirty. After evaluating all these reasons with the company managers, it was decided to choose a supplier for the hose.

The rest of the study is organized as follows. MEREC and CoCoSo methods used in supplier selection are introduced in the second section of the study. In the third section, the application steps of the proposed method are explained, and the results of the sensitivity analysis are given. The last section mentions results, discussion, and future work. According to the results obtained, suggestions were made to the company regarding supplier selection.

2. Material and method

The steps of the proposed approach are given in Figure 1. Now, how to implement these steps will be explained step by step.

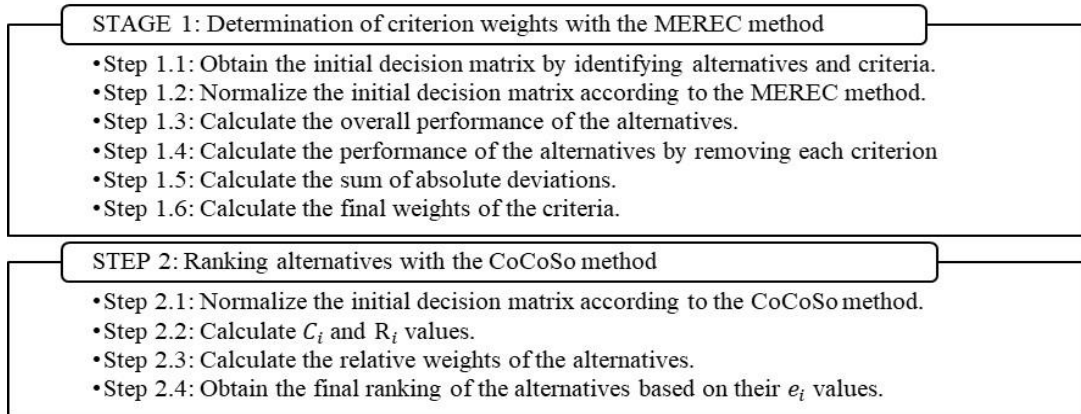


Fig.1. Steps of the proposed approach.

2.1. Stage 1: Determination of criterion weights with the MEREC method

The MEREC method determines the weights and priority of the criteria used in the decision-making process. This method evaluates the effect of removing each criterion used on the result. The steps of the method are given below [18].

Step 1.1: Obtain the initial decision matrix by identifying alternatives and criteria.

In the first step of the proposed method, after the alternatives $(A_i, i = 1, 2, \dots, n)$ and criteria $(k_j, j = 1, 2, \dots, m)$ are determined, the initial decision matrix (G) is obtained according to Equation 1. In Equation 1, $g_{(i,j)}, i = 1, 2, \dots, n; j = 1, 2, \dots, m$ shows the value of the j th criterion for the i th alternative.

$$G = \begin{bmatrix} g_{(1,1)} & g_{(1,2)} & \dots & g_{(1,j)} & \dots & g_{(1,m)} \\ g_{(2,1)} & g_{(2,2)} & \dots & g_{(2,j)} & \dots & g_{(2,m)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ g_{(i,1)} & g_{(i,2)} & \dots & g_{(i,j)} & \dots & g_{(i,m)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ g_{(n,1)} & g_{(n,2)} & \dots & g_{(n,j)} & \dots & g_{(n,m)} \end{bmatrix} \quad (1)$$

Step 1.2: Normalize the initial decision matrix according to the MEREC method.

In this step, the elements of the initial decision matrix given in Equation 1 are normalized using Equation 2.

$$d_{(i,j)} = \begin{cases} \frac{\min_k g(k,j)}{g(i,j)} & , j \in B \\ \frac{g(i,j)}{\max_k g(k,j)} & , j \in C \end{cases} \quad (2)$$

In Equation 2, the normalized initial decision matrix elements are $d_{(i,j)}, i = 1, 2, \dots, n; j = 1, 2, \dots, m$. Set B shows the index set of benefit type criteria, and set C shows the index set of cost type criteria.

Step 1.3: Calculate the overall performance of the alternatives.

In this step, the overall performance of the alternatives is obtained using a logarithmic measure with equal criterion weights. Using the normalized values calculated in Step 1.2, the performance of each alternative is calculated as in Equation 3.

$$Y_i = \ln \left(1 + \left(\frac{1}{m} \sum_{j=1}^m |\ln(d_{(i,j)})| \right) \right), i = 1, 2, \dots, n \quad (3)$$

Y_i given in Equation 3 shows the overall performance of the alternatives.

Step 1.4: Calculate the performance of the alternatives by removing each criterion.

Again, a logarithmic measure like that in Step 1.3 is used. The difference between Step 1.3 and this step is that the performances of the alternatives are calculated based on the separate impact of each criterion.

Therefore, there are m performance sets associated with m number of criteria. Regarding removing the j th criterion, the performance of the i th alternative is denoted by Y'_{ij} and is calculated as in Equation 4.

$$Y'_{ij} = \ln \left(1 + \left(\frac{1}{m} \sum_{k,k \neq j} |\ln(d_{(i,k)})| \right) \right) \quad (4)$$

Step 1.5: Calculate the sum of absolute deviations.

The removal effect of the j th criterion is calculated according to the Y_{ij} and Y'_{ij} values calculated in Step 1.3 and Step 1.4. Equation 5 calculates each criterion's removal effect Z_j . With the removal effect, the weights are determined in proportion to their impact on the criteria.

$$Z_j = \sum_{i=1}^m |Y'_{ij} - Y_i| \quad j = 1, 2, \dots, m \quad (5)$$

Step 1.6: Calculate the final weights of the criteria.

The weight of each criterion is calculated as in Equation 6, using the removal effects Z_j calculated in Step 1.5, where h_j is the weight of the j th criterion.

$$h_j = \frac{z_j}{\sum_k z_k}, \quad j = 1, 2, \dots, m \quad (6)$$

In the rest of the study, the alternatives' final ranking will be obtained using the criterion weights calculated with the MEREC method in the CoCoSo method.

2.2. Stage 2: Ranking alternatives with the CoCoSo method

The CoCoSo method is based on simple and exponential weighting. The following steps are followed for the suppliers and selection criteria previously determined for the MEREC method:

Step 2.1: Normalize the initial decision matrix according to the CoCoSo method.

Depending on whether the criteria are benefit type (B) or cost type (C), the initial decision matrix is normalized using Equation 7.

$$d_{(i,j)}^* = \begin{cases} \frac{g_{(i,j)} - \min_i g_{(i,j)}}{\max_i g_{(i,j)} - \min_i g_{(i,j)}}, & j \in B \\ \frac{\max_i g_{(i,j)} - g_{(i,j)}}{\max_i g_{(i,j)} - \min_i g_{(i,j)}}, & j \in C \end{cases} \quad (7)$$

In Equation 7, $d_{(i,j)}^*$ represents the elements of the initial decision matrix normalized according to the CoCoSo method.

Step 2.2: Calculate the C_i and R_i values.

$C_i, i = 1, \dots, n$ values which express the sums of the weighted comparability sequences, are calculated using Equation 8. The power weight of the comparability sequences for each alternative is defined as $R_i, i = 1, \dots, n$ and is calculated with Equation 9.

$$C_i = \sum_{j=1}^m (h_j d_{(i,j)}^*), \quad i = 1, 2, \dots, n \quad (8)$$

$$R_i = \sum_{j=1}^m (d_{(i,j)}^*)^{h_j}, \quad i = 1, 2, \dots, n \quad (9)$$

Step 2.3: Calculate the relative weights of the alternatives.

Three evaluation scores are first obtained to calculate the relative weights of the alternatives. The first evaluation score (e_{ia}) represents the arithmetic average of the sums of the weighted sum and multiplication methods and is calculated as in Equation 10.

$$e_{ia} = \frac{R_i + C_i}{\sum_{i=1}^n (R_i + C_i)}, \quad i = 1, 2, \dots, n \quad (10)$$

The second evaluation score (e_{ib}) refers to the sum of the relative scores of the weighted sum and multiplication methods, and Equation 11 is used to calculate.

$$e_{ib} = \frac{C_i}{\min_i C_i} + \frac{R_i}{\min_i R_i}, \quad i = 1, 2, \dots, n \quad (11)$$

The third evaluation score (e_{ic}) is calculated as in Equation 12 by balancing the weighted sum and multiplication methods with λ . The weight rating λ is usually taken as 0.5.

$$e_{ic} = \frac{\lambda(C_i) + (1-\lambda)(R_i)}{(\lambda \max_i C_i + (1-\lambda) \max_i R_i)}, \quad 0 \leq \lambda \leq 1, \quad i = 1, 2, \dots, n \quad (12)$$

Step 2.4: Obtain the final ranking of the alternatives based on their e_i values.

The calculations in Step 2.3 are used to obtain the final ranking of the alternatives. Using these calculations in Equation 13, $e_i, i = 1, 2, \dots, n$ values are obtained.

$$e_i = (e_{ia}e_{ib}e_{ic})^{\frac{1}{3}} + \frac{1}{3}(e_{ia} + e_{ib} + e_{ic}), \quad i = 1, 2, \dots, n \quad (13)$$

The final ranking of the alternatives is obtained, with the alternative with the largest e_i value being ranked first.

3. Application of the proposed method

In its Priority Medical Devices Project, the World Health Organization (WHO) suggests two potential causes for problems in healthcare delivery in low- and middle-income countries. The first of these reasons is that it targets high-income country economies, which it sees as having high-profit potential, and the second is the problems related to the inability to supply medical devices reasonably [9]. This makes supplier selection even more critical, especially for medium and small-sized companies operating in the medical device industry.

The method proposed in this study was applied in selecting hose suppliers in a company that produces biomedical devices. These hoses are used in the production of infusion and extension sets. It has been observed that the hoses supplied by the current supplier must have the desired features and be delivered on time.

In the following subsection, the application steps of the proposed method in selecting the hose supplier in the company are explained.

3.1. Application of stage 1: Determination of supplier selection criteria weights with the MEREC method

Step 1.1: Obtain the initial decision matrix by identifying alternatives and criteria.

As a result of the literature research and the evaluations of the company managers, six alternative suppliers ($t_i, i = 1, 2, \dots, 6$) and six criteria ($k_j, j = 1, 2, \dots, 6$) were determined to evaluate these suppliers. The determined criteria and definitions of the criteria are as follows:

- Product quality (k_1): It refers to the semi-finished product's raw material and performance quality. While scoring in the decision matrix, it was evaluated by company managers by giving a performance score out of 100.
- Quality certificate (k_2): It indicates how many years the supplier company has had the quality certificate. The length of time companies has this certificate indicates the year they produced by standard requirements. While scoring in the decision matrix, the year of the first issue of the quality certificate was taken as a basis.
- Lead time (k_3): It is the value of the lead time in weeks. It is the deadline companies give after ordering semi-finished products/raw materials.
- Price (k_4): It refers to the cost per meter of the product supplied in TL. The decision matrix includes the current prices of the products as of January 2023.
- Transportation costs (k_5): It refers to the transportation costs invoiced in the supply of semi-finished products. Transportation agreements made with companies vary. While some companies include this fee in the price, others ship by cargo. The difference in shipping costs varies due to shipping and customs costs (if it is an imported product). Values in Turkish Lira are shown in the decision matrix.
- Customer relations (k_6): It refers to the companies' tendency to cooperate during and after the supply period. The evaluation made with the company managers was scored out of 100, considering the company's past experiences. Six supplier alternatives were determined based on the company's experience and sector knowledge. Due to company confidentiality, suppliers will be indicated with $t_i, i = 1, 2, \dots, 6$ in the next steps of the application. In the first step of the proposed method, after the alternatives ($t_i, i = 1, 2, \dots, 6$) and criteria ($k_j, j = 1, 2, \dots, 6$) are determined, the initial decision matrix (G) is created as in Equation 2. The initial decision matrix is given in Table 2.

Table 2. The initial decision matrix.

	$k_1 \in B$	$k_2 \in B$	$k_3 \in C$	$k_4 \in C$	$k_5 \in C$	$k_6 \in B$
t_1	90	13	2	1,23	1	100
t_2	80	20	2	1,04	200	65
t_3	100	19	4	0,73	10000	65
t_4	65	10	2	0,94	500	80
t_5	85	8	1	1,76	300	70
t_6	50	5	1	0,83	150	75

Step 1.2: Normalize the initial decision matrix according to the MEREC method.

In this step, the elements of the initial decision matrix given in Table 2 are normalized using Equation 2. Normalized values are shown in Table 3.

Table 3. The normalized initial decision matrix for the MEREC method.

	$k_1 \in B$	$k_2 \in B$	$k_3 \in C$	$k_4 \in C$	$k_5 \in C$	$k_6 \in C$
t_1	0.5556	0.3846	0.5000	0.6989	0.0001	0.6500
t_2	0.6250	0.2500	0.5000	0.5909	0.0200	1.0000
t_3	0.5000	0.2632	1.0000	0.4148	1.0000	1.0000
t_4	0.7692	0.5000	0.5000	0.5341	0.0500	0.8125
t_5	0.5882	0.6250	0.2500	1.0000	0.0300	0.9286
t_6	1.0000	1.0000	0.2500	0.4716	0.0150	0.8667

Step 1.3: Calculate the overall performance of the alternatives.

In this step, the overall performance of alternative suppliers ($Y_i, i = 1, 2, \dots, 6$) is obtained using a logarithmic measure with equal criterion weights. This measure is based on the formula given in Equation 3. The results obtained are shown in Table 4.

Table 4. The performance table of suppliers.

	t_1	t_2	t_3	t_4	t_5	t_6
Y_i	1.1116	0.7722	0.3952	0.6488	0.6904	0.7324

Step 1.4: Calculate the performance of the alternatives by removing each criterion.

There are m performance sets associated with the number of m criteria. The performance of the i th alternative regarding the removal of the j th criterion is denoted by Y'_{ij} and calculated with Equation 4. Calculation results are given in Table 5.

Table 5. The performance table of suppliers by removing each criterion.

Y'_{ij}	$k_1 \in B$	$k_2 \in B$	$k_3 \in C$	$k_4 \in C$	$k_5 \in C$	$k_6 \in C$
t_1	1.0789	1.0578	1.0729	1.0918	0.4083	1.0877
t_2	0.7354	0.6594	0.7174	0.7309	0.4138	0.7722
t_3	0.3142	0.2329	0.3952	0.2912	0.3952	0.3952
t_4	0.6257	0.5865	0.5865	0.5926	0.3464	0.6305

t_5	0.6451	0.6504	0.5673	0.6904	0.3437	0.6842
t_6	0.7324	0.7324	0.6147	0.6703	0.3222	0.7209

Step 1.5: Calculate the sum of absolute deviations.

According to the Y_{ij} and Y'_{ij} values calculated in Step 1.3 and Step 1.4, the removal effect Z_j of the j th criterion is computed using Equation 5 and given in Table 6.

Table 6. The sum of absolute deviations for all criteria.

	$k_1 \in B$	$k_2 \in B$	$k_3 \in C$	$k_4 \in C$	$k_5 \in C$	$k_6 \in C$
Z_j	1.1116	0.7722	0.3952	0.6488	0.6904	0.7324

Step 1.6: Calculate the final weights of the criteria.

The weight of each criterion (h_j) is determined using the Z_j removal effects calculated in Step 1.5 using Equation 7. The results are given in Table 7.

Table 7. The weights and the rankings of the criteria.

	$k_1 \in B$	$k_2 \in B$	$k_3 \in C$	$k_4 \in C$	$k_5 \in C$	$k_6 \in C$
h_j	0,0624	0,1228	0,1130	0,0807	0,6040	0,0171
Rank	5	2	3	4	1	6

After the weights were calculated using MEREC, the final ranking was made with the CoCoSo method.

3.2. Application of stage 2: Ranking alternatives with the CoCoSo method

Step 2.1: Normalize the initial decision matrix according to the CoCoSo method.

Depending on whether the criteria are benefit-type (B) or cost-type (C), the initial decision matrix is normalized using Equation 7. The matrix normalized according to the CoCoSo method is given in Table 8.

Table 8. The normalized initial decision matrix for the CoCoSo method.

	$k_1 \in B$	$k_2 \in B$	$k_3 \in C$	$k_4 \in C$	$k_5 \in C$	$k_6 \in C$
t_1	0.8000	0.5333	0.6667	0.5146	1.0000	1.0000
t_2	0.6000	1.0000	0.6667	0.6990	0.9801	0.0000
t_3	1.0000	0.9333	0.0000	1.0000	0.0000	0.0000
t_4	0.3000	0.3333	0.6667	0.7961	0.9501	0.4286
t_5	0.7000	0.2000	1.0000	0.0000	0.9701	0.1429
t_6	0.0000	0.0000	1.0000	0.9029	0.9851	0.2857

Step 2.2. Calculate the C_i and R_i values.

The sum of the weighted comparability sequences is denoted as $C_i, i = 1, \dots, n$ and the power weight of comparability sequences for each alternative is denoted as $R_i, i = 1, \dots, n$. Using the weights (h_j) obtained from the MEREC method, R_i and C_i values were calculated as in Equation 8 and Equation 9, respectively. R_i and C_i values are given in Table 9.

Table 9. The values of C_i and R_i .

	t_1	t_2	t_3	t_4	t_5	t_6
R_i	5.8149	4.8833	2.9916	5.6936	4.7478	3.9616
C_i	0.8533	0.8840	0.2578	0.7804	0.7696	0.7857

Step 2.3: Calculate the relative weights of the alternatives.

Supplier evaluation scores e_{ia} , e_{ib} and e_{ic} were calculated according to Equation 10-12, respectively, and the calculation results are given in Table 10. The λ value in Equation 12 was taken as 0.5.

Table 10. CoCoSo evaluation strategy scores.

	t_1	t_2	t_3	t_4	t_5	t_6
e_{ia}	0.2057	0.1779	0.1002	0.1997	0.1702	0.1464
e_{ib}	5.2542	5.0617	2.0000	4.9307	4.5726	4.3724
e_{ic}	0.9954	0.8609	0.4851	0.9664	0.8236	0.7087

Step 2.4: Obtain the final ranking of the alternatives based on their e_i values.

The calculations in Step 2.3 are used to obtain the final ranking of the alternatives. These calculations are used in Equation 13 to calculate $e_i, i = 1, 2, \dots, 6$ values. The scores of alternative suppliers are given in Table 11. In this table, the rankings are numbered from largest to smallest.

Table 11. The scores and the final rankings of the suppliers.

	t_1	t_2	t_3	t_4	t_5	t_6
e_i	3.1763	2.9521	1.3216	3.0158	2.7176	2.5109
Rank	1	3	6	2	4	5

The evaluation of the suppliers' scores concluded that the best supplier was supplier 1 (t_1).

3.3. Sensitivity analysis

To validate the findings and support the precision and deviation of the decision outcomes, a sensitivity analysis is used. By making minor adjustments to the main model, a sensitivity analysis test could assist decision-makers in demonstrating the results of their methods. In this section, the effects of MEREC, CRITIC, Entropy and Equal Weight (EW) methods used in criterion weighting on alternative supplier rankings were observed. First, the weights of the criteria were calculated using all methods and their rankings were obtained as in Table 12. The steps of the calculations can be found in studies [18], [43]-[44].

As seen in Table 11, although the MEREC and Entropy methods revealed similar criterion rankings, the criterion rankings differed when CRITIC and EW methods were used. The reasons for different criterion weights and rankings can be listed as follows. MEREC, Entropy and CRITIC methods are objective methods that reveal criterion weights according to the actual values of the initial decision matrix. Each method aims to achieve effective criterion weighting in a different way. The MEREC method uses criterion weights to calculate the impact of each criterion on the overall performance of the alternatives. It also uses a logarithmic function to measure the overall and partial performance of alternatives. MEREC gives greater weight to criteria that have a greater impact on the performance of alternatives [18]. In the Entropy method, uncertainty is greater in the data group with higher values [43]. CRITIC is a method in which the standard deviations of the criteria and the correlation between the criteria are used together

[44]. In this study, Spearman's correlation was used because the data did not comply with normal distribution. In the EW method, all criteria are given equal weight.

Table 12. The weights and rankings of the criteria using different objective criteria weighting methods.

	MEREK		CRITIC		Entropy		EW	
	h_j	Rank	h_j	Rank	h_j	Rank	h_j	Rank
C_1	0.062	5	0.160	4	0.014	5	0.167	1
C_2	0.123	2	0.183	1	0.061	3	0.167	1
C_3	0.113	3	0.176	2	0.071	2	0.167	1
C_4	0.081	4	0.174	3	0.028	4	0.167	1
C_5	0.604	1	0.153	6	0.818	1	0.167	1
C_6	0.017	6	0.154	5	0.007	6	0.167	1

The criterion weights given in Table 12 were used in the CoCoSo method and the supplier rankings are given in Table 13. In Table 13, it is seen that the ranks for all criteria weighting methods are the same. These results show that CoCoSo is an effective and robust method for alternative ranking.

Table 13. The e_i values and rankings of the alternatives using different objective criteria weighting methods.

	MEREK		CRITIC		Entropy		EW	
	e_i	Rank	e_i	Rank	e_i	Rank	e_i	Rank
t_1	3.176	1	2.428	1	5.502	1	2.450	1
t_2	2.952	3	2.093	3	5.224	3	2.087	3
t_3	1.322	6	1.394	6	1.287	6	1.381	6
t_4	3.016	2	2.182	2	5.255	2	2.189	2
t_5	2.718	4	1.810	4	5.003	4	1.820	4
t_6	2.511	5	1.660	5	4.783	5	1.660	5

4. Conclusion and discussion

The medical device industry is a rapidly growing sector that includes many parameters. Constantly evolving technologies contribute to advancing diagnostic and treatment practices in healthcare. Medical device manufacturers must continuously carry out improvement activities within the company to survive in an increasingly competitive environment and keep up with this change. The company should review its processes, identify problems, take the necessary precautions, implement them, and check the impact and efficiency of the implemented activities. This process becomes even more critical when considering the direct effect of product safety and quality on human life in the medical device industry because many companies produce in the industrial field.

In companies that produce medical devices, the legal requirements of the product being a medical device are followed by notified organizations and competent authorities within the framework of standards and regulations within the scope of quality processes. When placing its devices on the market or putting them into service, the manufacturer must ensure that its devices are designed and manufactured by the requirements of the relevant regulation. The manufacturer's obligations are specified in the applicable EU legislation and the rules published by the Ministry of Health in compliance with this legislation. Increasing costs and liabilities with MDR 2017/45 have made medical device manufacturers question their processes. Meeting customer demands while fulfilling all legal

requirements becomes critical for manufacturers to avoid market losses. The harmony between the manufacturer's activities and customer expectations is essential for the operation's success.

Manufacturers need to respond to customer demands without compromising product safety and performance. Product safety is evaluated during the medical device production process, considering patient and user health. This process covers the sales and subsequent tracking process, depending on the product class, starting from the purchase of raw materials, and is considered traceability. Availability begins when the order arrives and continues until its sale and destruction. The first critical step is the supply of raw materials/semi-finished products.

This study addressed the supplier selection problem experienced in a company that is a medical device manufacturer and produces consumables. First, it was determined that the company could not meet customer demands such as on-time delivery and shorter lead times, and it was observed that delivery times increased. A group of company managers was determined, and the reasons for these delays were discussed. As a result of this evaluation, it was concluded that the main reason for the delivery times was the disruptions in the supply processes. It is envisaged that if the problems in supply are resolved, the disruptions in planning will be significantly reduced. In addition, it is anticipated that solving this problem will contribute considerably to the subsequent processes for the company, which is in the preparation phase for MDR 2017/45.

In this study, the supplier selection problem was identified at a medical device manufacturer. The hoses are the primary elements of extension and infusion sets. When the company's 2022 production data was referenced, it was observed that there was a problem with the hose supply, and this situation affected the production process directly.

There are many alternatives and criteria to be considered in the supplier selection problem. Due to the structure of supplier selection, MCDM methods are frequently used in the literature. For the hose supplier selection problem, first, alternative suppliers and selection criteria were determined using the company's industry knowledge and focus group evaluations. In the evaluation made with the focus group and considering the literature review, six criteria were determined: Product quality (k_1), year of quality certificates (k_2), delivery time (k_3), price (k_4), transportation costs (k_5), customer relations (k_6). In this study, the MEREC-CoCoSo integration is suggested for supplier selection. While the criteria weights were determined using the MEREC method, the hose suppliers were ranked using the CoCoSo method. The results were evaluated together with the company managers. According to the results of the MEREC method, the obtained criteria weights were calculated as $h_1 = 0.0624$, $h_2 = 0.1228$, $h_3 = 0.1130$, $h_4 = 0.0807$, $h_5 = 0.6040$, $h_6 = 0.0171$. According to these results, transportation cost criterion is the most weighted criterion. It has been determined that this criterion has a more significant impact on the result. The criterion weights are shipping costs, year of quality certificate, lead time, price, quality, and customer relations.

After determining the weights, the rankings of the alternatives were obtained with the CoCoSo method. It was concluded that the best supplier in the final ranking of the alternatives was the number 1 supplier. At the same time, this supplier has had a quality certificate for 13 years, and the supply period is two weeks, which can be considered a good choice. The results were consistent with the evaluations of company managers and engineers.

The contribution of this study to the literature is as follows [45].

- This study addressed the supplier selection problem for medical device manufacturers. In the literature study, although there are studies about supplier selection problems in the healthcare industry, these studies conducted in the medical device sector need to be more comprehensive and specific for this field. The healthcare industry is intertwined with the medical device industry, but it should not be overlooked that they are two different industries. The study anticipates contributing to the literature in this sense.
- Although MEREC-CoCoSo methods were used together in different areas, this was the first time the MEREC-CoCoSo method has been applied for supplier selection in the medical device industry. While determining the criteria, along with MDR 2017/45, the fact that the manufacturers and suppliers have quality certificates was considered.
- Considering the consistency of the results, this study is a guide for medical device manufacturers who are envisaged to review their supplier selection processes with the proposed method while preparing for MDR

2017/45, which will assist companies in directing their resources correctly. The relevant study is thought to help decision-makers in the criteria determination stage of medical device manufacturers with a high medical device class or operating as active device manufacturers, considering the product features, and intended use.

Future studies may include the following: The number of criteria and suppliers used in this study may be increased. The MCDM methods used can be differentiated and sensitivity analysis can be performed by comparing the results. The impact of decision makers on supplier rankings can be discussed by choosing subjective MCDM methods instead of the objective MCDM methods used in this study.

Acknowledgement

We would like to thank distinguished anonymous reviewers for their guidance and support in our work. Their review has helped us to make our study richer and more impactful.

References

- [1] "TIBBİ CİHAZ YÖNETMELİĞİ Sayfa 2 / 143," *Türkiye İlaç ve Tıbbi Cihaz Kurumundan*.
- [2] T. W. Li, P. W. Tu, L. L. Liu, and S. I. Wu, "Assurance of medical device quality with quality management system: An analysis of good manufacturing practice implementation in Taiwan," *Biomed Res. Int.*, vol. 2015, 2015, doi: 10.1155/2015/670420.
- [3] V. Jain, A. K. Sangaiah, S. Sakhuja, N. Thoduka, and R. Aggarwal, "Supplier selection using fuzzy AHP and TOPSIS: a case study in the Indian automotive industry," *Neural Comput. Appl.*, vol. 29, no. 7, pp. 555–564, Apr. 2018, doi: 10.1007/s00521-016-2533-z.
- [4] A. Fallahpour, E. Udoncy Olugu, S. Nurmaya Musa, K. Yew Wong, and S. Noori, "A decision support model for sustainable supplier selection in sustainable supply chain management," *Comput. Ind. Eng.*, vol. 105, pp. 391–410, Mar. 2017, doi: 10.1016/J.CIE.2017.01.005.
- [5] W. Song, Z. Xu, and H.-C. Liu, "Developing sustainable supplier selection criteria for solar air-conditioner manufacturer: An integrated approach," 2017, doi: 10.1016/j.rser.2017.05.081.
- [6] M. Yazdani, P. Chatterjee, E. K. Zavadskas, and S. Hashemkhani Zolfani, "Integrated QFD-MCDM framework for green supplier selection," *J. Clean. Prod.*, vol. 142, pp. 3728–3740, Jan. 2017, doi: 10.1016/J.JCLEPRO.2016.10.095.
- [7] N. Banaeian, H. Mobli, B. Fahimnia, I. E. Nielsen, and M. Omid, "Green supplier selection using fuzzy group decision making methods: A case study from the agri-food industry," *Comput. Oper. Res.*, vol. 89, pp. 337–347, Jan. 2018, doi: 10.1016/J.COR.2016.02.015.
- [8] M. Abdel-Basset, G. Manogaran, A. Gamal, and F. Smarandache, "A hybrid approach of neutrosophic sets and DEMATEL method for developing supplier selection criteria," *Des. Autom. Embed. Syst.*, vol. 22, no. 3, pp. 257–278, Sep. 2018, doi: 10.1007/S10617-018-9203-6/METRICS.
- [9] M. Yazdani, Z. Wen, H. Liao, A. Banaitis, and Z. Turskis, "A grey combined compromise solution (CoCoSo-G) method for supplier selection in construction management," *J. Civ. Eng. Manag.*, vol. 25, no. 8, pp. 858–874, Nov. 2019, doi: 10.3846/JCEM.2019.11309.
- [10] Ž. Stević, D. Pamučar, A. Puška, and P. Chatterjee, "Sustainable supplier selection in healthcare industries using a new MCDM method: Measurement of alternatives and ranking according to COMpromise solution (MARCOS)," *Comput. Ind. Eng.*, vol. 140, p. 106231, Feb. 2020, doi: 10.1016/J.CIE.2019.106231.
- [11] M. Yazdani, A. E. Torkayesh, and P. Chatterjee, "An integrated decision-making model for supplier evaluation in public healthcare system: the case study of a Spanish hospital," *J. Enterp. Inf. Manag.*, vol. 33, no. 5, pp. 965–989, Dec. 2020, doi: 10.1108/JEIM-09-2019-0294/FULL/HTML.
- [12] D. Sumrit, "Supplier selection for vendor-managed inventory in healthcare using fuzzy multi-criteria decision-making approach," *Decis. Sci. Lett.*, vol. 9, no. 2, pp. 233–256, 2020, doi: 10.5267/J.DSL.2019.10.002.
- [13] K. K. Göncü and O. Çetin, "A Decision Model for Supplier Selection Criteria in Healthcare Enterprises with Dematel ANP Method," *Sustain.* 2022, Vol. 14, Page 13912, vol. 14, no. 21, p. 13912, Oct. 2022, doi: 10.3390/SU142113912.
- [14] T. L. Nguyen et al., "A Novel Integrating Data Envelopment Analysis and Spherical Fuzzy MCDM Approach for Sustainable Supplier Selection in Steel Industry," *Math.* 2022, Vol. 10, Page 1897, vol. 10, no. 11, p. 1897, Jun. 2022, doi: 10.3390/MATH10111897.
- [15] P. Ghadimi and C. Heavey, "Sustainable supplier selection in medical device industry: Toward sustainable manufacturing," *Procedia CIRP*, vol. 15, pp. 165–170, 2014, doi: 10.1016/J.PROCIR.2014.06.096.
- [16] P. Ghadimi, C. Wang, M. K. Lim, and C. Heavey, "Intelligent sustainable supplier selection using multi-agent technology: Theory and application for Industry 4.0 supply chains," *Comput. Ind. Eng.*, vol. 127, pp. 588–600, Jan. 2019, doi: 10.1016/J.CIE.2018.10.050.
- [17] A. A. Forouzeshnejad, "Leagile and sustainable supplier selection problem in the Industry 4.0 era: a case study of the medical devices using hybrid multi-criteria decision making tool," *Environ. Sci. Pollut. Res.*, vol. 30, pp. 13418–13437, 2023, doi: 10.1007/s11356-022-22916-x.
- [18] M. Keshavarz-Ghorabae, M. Amiri, E.K. Zavadskas, Z. Turskis, & J. Antucheviciene "Determination of objective weights using a new method based on the removal effects of criteria (MERECE)". *Symmetry*, vol 13, no.4, p.525, 2021.
- [19] M. Gligoric, K. Urošević, D. Halilovic, M. Gligorić, S. Lutovac, and D. Halilović, "Optimal Coal Supplier Selection for Thermal Power Plant Based on MCRAT Method," Accessed: May 07, 2023. [Online]. Available: <https://www.researchgate.net/publication/365471602>.

- [20] D. Trung, & H. T.-A. in P. E., and undefined 2021, "A multi-criteria decision-making in turning process using the MAIRCA, EAMR, MARCOS and TOPSIS methods: A comparative study," *researchgate.net*, vol. 16, no. 4, pp. 443–456, 2021, doi: 10.14743/apem2021.4.412.
- [21] M. K.-G.-S. Reports and undefined 2021, "Assessment of distribution center locations using a multi-expert subjective–objective decision-making approach," *nature.com*, Accessed: May 06, 2023. [Online]. Available: <https://www.nature.com/articles/s41598-021-98698-y>.
- [22] T. M. H. Nguyen, V. P. Nguyen, and D. T. Nguyen, "A new hybrid Pythagorean fuzzy AHP and CoCoSo MCDM based approach by adopting artificial intelligence technologies," <https://doi.org/10.1080/0952813X.2022.2143908>, 2022, doi: 10.1080/0952813X.2022.2143908.
- [23] J. F. Nicolalde, M. Cabrera, J. Martínez-Gómez, R. B. Salazar, and E. Reyes, "Selection of a phase change material for energy storage by multi-criteria decision method regarding the thermal comfort in a vehicle," *J. Energy Storage*, vol. 51, p. 104437, Jul. 2022, doi: 10.1016/J.EST.2022.104437.
- [24] G. Shanmugasundar, G. Sapkota, R. Čep, and K. Kalita, "Application of MEREK in Multi-Criteria Selection of Optimal Spray-Painting Robot," *Process*. 2022, Vol. 10, Page 1172, vol. 10, no. 6, p. 1172, Jun. 2022, doi: 10.3390/PR10061172.
- [25] Y. Yu, S. Wu, J. Yu, Y. Xu, L. Song, and W. Xu, "A hybrid multi-criteria decision-making framework for offshore wind turbine selection: A case study in China," *Appl. Energy*, vol. 328, p. 120173, Dec. 2022, doi: 10.1016/J.APENERGY.2022.120173.
- [26] D. T. Do and N. T. Nguyen, "Applying Cocoso, Mabac, Mairca, Eamr, Topsis and Weight Determination Methods for Multi-Criteria Decision Making in Hole Turning Process," *Sroj. Cas.*, vol. 72, no. 2, pp. 15–40, Nov. 2022, doi: 10.2478/SCJME-2022-0014.
- [27] K. Diaconu, Y. F. Chen, S. Manaseki-Holland, C. Cummins, and R. Lilford, "Medical device procurement in low- and middle-income settings: Protocol for a systematic review," *Syst. Rev.*, vol. 3, no. 1, pp. 1–11, Oct. 2014, doi: 10.1186/2046-4053-3-118/TABLES/4.
- [28] B. Ayan, S. Abacıoğlu, and M. P. Basilio, "A Comprehensive Review of the Novel Weighting Methods for Multi-Criteria Decision-Making," *Inf. 2023*, Vol. 14, Page 285, vol. 14, no. 5, p. 285, May 2023, doi: 10.3390/INFO14050285.
- [29] R. Chaurasiya and D. Jain, "A New Algorithm on Pythagorean Fuzzy-Based Multi-Criteria Decision-Making and Its Application," *Iran. J. Sci. Technol. - Trans. Electr. Eng.*, pp. 1–16, May 2023, doi: 10.1007/S40998-023-00600-1/TABLES/15.
- [30] M. Yazdani, P. Zarate, E. K. Zavadskas, & Z. Turskis. A combined compromise solution (CoCoSo) method for multi-criteria decision-making problems. *Management Decision*, vol.57, no.9, pp. 2501-2519, 2019.
- [31] F. Ecer, D. Pamucar, S. Zolfani, M. E.-J. of Cleaner, and undefined 2019, "Sustainability assessment of OPEC countries: Application of a multiple attribute decision making tool," *Elsevier*, Accessed: May 19, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0959652619331944>.
- [32] Ž. Erceg, V. Starčević, D. Pamučar, G. Mitrovi, Ž. Stevi, and S. Žiki, "A new model for stock management in order to rationalize costs: ABC-FUCOM-interval rough CoCoSo model," *mdpi.com*, 2019, doi: 10.3390/sym11121527.
- [33] Z. Wen, H. Liao, E. Kazimieras Zavadskas, and A. Al-Barakati, "Selection third-party logistics service providers in supply chain finance by a hesitant fuzzy linguistic combined compromise solution method," <http://www.tandfonline.com/action/authorSubmission?journalCode=ro20&page=instructions>, vol. 32, no. 1, pp. 4033–4058, Jan. 2019, doi: 10.1080/1331677X.2019.1678502.
- [34] S. H. Zolfani and M. Yazdani, "A structured framework for sustainable supplier selection using a combined BWM-CoCoSo model CALL FOR BOOK PROPOSALS View project 1st Indo-Serbian International Conference on Computational Intelligence for Engineering and Management Applications (CIEMA)-2022 View project," 2019, doi: 10.3846/cibmee.2019.081.
- [35] D. Bagal, B. Naik, B. Parida, ... A. B.-I. C., and undefined 2020, "Comparative mechanical characterization of M30 concrete grade by fractional replacement of portland pozzolana cement with industrial waste using CoCoSo and," *iopscience.iop.org*, Accessed: May 19, 2023. [Online]. Available: <https://iopscience.iop.org/article/10.1088/1757-899X/970/1/012015/meta>.
- [36] A. Barua, S. Jeet, D. Bagal, ... P. S.-I. J. I. T., and undefined 2019, "Evaluation of mechanical behavior of hybrid natural fiber reinforced nano sic particles composite using hybrid Taguchi-CoCoSo method," *researchgate.net*, no. 10, pp. 2278–3075, 2019, doi: 10.35940/ijitee.J1232.0881019.
- [37] T. Biswas, P. Chatterjee, B. C.-O. research in, and undefined 2020, "Selection of commercially available alternative passenger vehicle in automotive environment," *oresta.rabek.org*, Accessed: May 19, 2023. [Online]. Available: <https://www.oresta.rabek.org/index.php/oresta/article/view/39>.
- [38] S. B.-D. M. A. in M. and and undefined 2020, "Measuring performance of healthcare supply chains in India: A comparative analysis of multi-criteria decision making methods," *dmame.rabek.org*, Accessed: May 19, 2023. [Online]. Available: <https://www.dmame.rabek.org/index.php/dmame/article/view/133>.
- [39] F. Ecer, D. P.-J. of cleaner production, and undefined 2020, "Sustainable supplier selection: A novel integrated fuzzy best worst method (F-BWM) and fuzzy CoCoSo with Bonferroni (CoCoSo'B) multi-criteria model," *Elsevier*, Accessed: May 19, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S095965262032028X>.
- [40] X. Peng, H. H.-T. and E. Development, and undefined 2020, "Fuzzy decision making method based on CoCoSo with critic for financial risk evaluation," [journals.vilniustech.lt](https://journals.vilniustech.lt/index.php/TEDE/article/view/11920), Accessed: May 19, 2023. [Online]. Available: <https://journals.vilniustech.lt/index.php/TEDE/article/view/11920>.
- [41] M. Yazdani, P. Chatterjee, D. Pamucar, and S. Chakraborty, "Development of an integrated decision making model for location selection of logistics centers in the Spanish autonomous communities," *Expert Syst. Appl.*, vol. 148, Jun. 2020, doi: 10.1016/J.ESWA.2020.113208.
- [42] F. Fahri Altuntaş and J. Genel Komutanlığı -Yönetici, "G7 ülkelerinin bilgi performanslarının analizi: CoCoSo yöntemi ile bir uygulama Analysis of knowledge performance of G7 countries: An application with the CoCoSo method," doi: 10.15637/jlecon.8.3.06.
- [43] C.L. Hwang and K. Yoon, , Methods for Multiple Attribute Decision Making. In: Multiple Attribute Decision Making. Lecture Notes in Economics and Mathematical Systems, vol 186. Springer, Berlin, Heidelberg. 1981, doi: 10.1007/978-3-642-48318-9_3
- [44] D. Diakoulaki, G. Mavrotas and L. Papayannakis, Determining objective weights in multiple criteria problems: The critic method. *Computers & Operations Research*, 22(7), 763-770, 1995.

[45] G. Özel, Tıbbi Cihaz Üretimi Yapan Bir Firmada Bütünleşik MEREC – CoCoSo Yönetimi İle Tedarikçi Seçimi (Master tezi, Başkent Üniversitesi Fen Bilimleri Enstitüsü), 2023.



E-ISSN: 2687-6167

Number 56, March 2024

RESEARCH ARTICLE

Receive Date: 09.11.2023

Accepted Date: 27.03.2024

Calculations of $Ba_{(1-x)}Sr_xTiO_3$ structure and band gap properties by using density functional theory

Sinem Aksan^{a*}

Kütahya Dumlupınar University, Faculty of Engineering, Department of Metallurgical and Materials Engineering, Kütahya, ORCID: 0000-0001-5771-2111

Abstract

The aim of this study is to simulate features using molecular modeling methods. The point is to show that it will accelerate research in material development studies by directing us, researchers, in terms of gaining time, material and workforce. In this study, the structural and electronic properties of undoped $BaTiO_3$ and Sr-doped $BaTiO_3$ were calculated by molecular modeling. In the study, energy calculations were made with the PBE and GGA (Generalized Gradient Approximation approach) developed by Perdew, Burke and Ernzerhof (PBE) using the density functional theory (DFT) calculation method, the CASTEP module of the Materials Studio program. First, the structural and electronic properties of the $BaTiO_3$ crystal phase were calculated. Then, the lattice constants, band gap values and electron state densities of the Sr doped structure to $BaTiO_3$ structure were calculated. The values in the literature were compared with the calculations made using the (DFT) density functional theory and it was determined that the calculations were in agreement with the values in the literature. It has been revealed that it will accelerate research in material development studies by giving direction to us researchers in terms of gaining from materials and workforce. As a result of geometric optimization of the non-stoichiometric $Ba_{(1-x)}Sr_xTiO_3$ structure and DFT calculations, it was determined that the electronic band gap shifted after %1 and %3Sr addition towards the conduction band and the band gap respectively decreased to 1,911 eV and 1.989eV.

© 2023 DPU All rights reserved.

Keywords: $BaTiO_3$; DFT; Molecular Modeling

1. Introduction

Barium titanate ($BaTiO_3$) has been of great interest for decades as a perovskite ferroelectric material and due to its superior physical properties, such as high dielectric constant, ferroelectricity, high voltage tunability, positive temperature resistivity coefficient, pyroelectricity and piezoelectricity, it is used in multilayer ceramic capacitors, piezoelectric ceramic transducers, It is widely used in memories, infrared sensors and temperature controllers [1][2].

* Corresponding author. Tel.: +905334712530.

E-mail address: sinem.aksan@dpu.edu.tr

BaTiO₃ capacitors used in many applications such as piezoelectric sensors, transducers, thermistors, energy carriers, etc. BaTiO₃ is used as dielectric material in multilayer capacitors, because of its high capacitance [3]. Many studies have been carried out with the addition of dopant to further improve the dielectric and ferroelectric properties of BaTiO₃ [4].

Today, many disciplines contain many complex problems that need to be solved. The fact that these complex problems cannot be solved analytically in many areas or that their solution is very difficult has led to the simulation technique with new searches in parallel with the development of technology. Molecular modeling is generally used to evaluate the current performance of a business process or to predict the future. It is designed to help practitioners discover new ways to achieve optimal results using mathematical, statistical and other analytical methods.

Molecular modeling methods have developed rapidly with the developments in computer technology. As a result of the developments in Molecular modeling methods, physical events that are very difficult to perform experimentally have been created in the computer environment and resolved within acceptable error limits [5].

BaTiO₃ has a high dielectric constant of 5000 at room temperature, with a Curie temperature of 120 °C, and an optical band gap of 3.3 eV, indicating its insulating nature. Its optical absorption in the UV range occurs below 400 nm. Additionally, BaTiO₃ possesses a perovskite structure of ABO₃ type and experiences three structural phase transitions from cubic to tetragonal, then orthorhombic, and finally rhombic phases as temperature decreases. [13]

The aim of this study is to simulate features using molecular modeling methods. The study is based on the effect of dopants on BaTiO₃ structure. The aim of this study is to show that it will accelerate research in material development studies by directing us, researchers, in terms of gaining time, material and workforce.

Nomenclature

DFT	Density Functional Theory
DOS	Density of States
pDOS	Partial Density of States

2. DFT studies

2.1. Molecular modeling with density functional theory

This study calculations were made using the simulation package CASTEP (Materials Studio CASTEP Package ID:11795). Materials Studio is a simulation program that models materials at the atomic scale with electronic structure calculations or quantum mechanical molecular dynamics (QMD) using DFT-based ab-initio computation techniques[15]. The Materials Studio Castep program aims to obtain approximate solutions of Schrodinger equations representing multibody systems by solving DFT and Kohn-Sham equations or HF and Roothaan equations. Hybrid functionals in the program are a mixture of HF approach and DFT. Moreover, Green's function methods and many-body perturbation theory are also embedded in the program [6]. The interactions of plane wave basis sets with electronic charge density, local potential and single electron orbital are considered. Interactions between electrons and ions are described using the norm-concerning or ultrasoft pseudopotential or projective extended wave method. By using the density functional theory (DFT) calculation method CASTEP module, energy and band gap calculations were made with the PBE and GGA (Generalized Gradient Approximation approach) developed by Perdew, Burke and Ernzerhof (PBE) [6].

Perdew-Burke-Ernzerhof (GGA-PBE) function and the soft pseudopotential proposed by used . Also the generalized gradient approximation (GGA) have been performed and electron displacement coaction, as well as electron-ion interaction[16]. The cut off energy of the calculation used as 500 eV. Electronic Configuration for Ba is 5s² 5p⁶ 6s², for O 2s² 2p⁴, Ti electronic configuration is Ti 3s² 3p⁶ 3d² 4s² where Sr configuration is Sr 4s² 4p⁶ 5s².

By establishing the convergence criteria at 1.0×10^{-5} electronvolts per atom for energy, 0.03 eV per angstrom for the highest force exerted on atoms, 0.05 gigapascals for the maximum stress, and 1.0×10^{-3} angstroms for the maximum atomic displacement as detailed in the publication authored by M. Hassan and collaborators [12].

3. Results

3.1. Geometry optimization of stoichiometric $BaTiO_3$ and non-stoichiometric $Ba_{(1-x)}Sr_xO_3$ structures

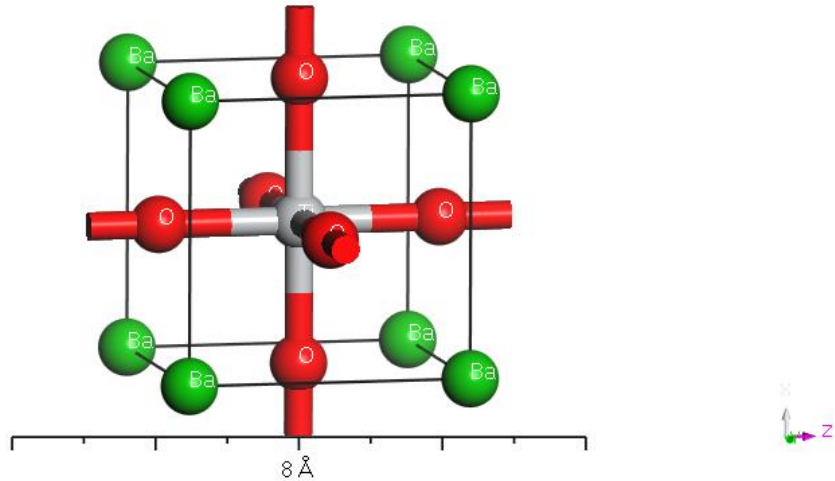


Fig. 1. Stoichiometric $BaTiO_3$ Structure

Calculations are completed using DFT, which includes the plane wave pseudopotentials. This approach allows us to make quick calculations with better efficiency. The soft pseudopotential proposed by Perdew-Burke-Ernzerhof (GGA-PBE) used to compare (GGA) generalized gradient approximation and electron displacement coaction, as well as electron-ion interaction. The cut off energy of the calculation used as 600 eV. Configuration of electronic for Ba is $5s^2 5p^6 6s^2$, for O $2s^2 2p^4$, Ti electronic configuration is $Ti 3s^2 3p^6 3d^2 4s^2$ where Sr configuration is $Sr 4s^2 4p^6 5s^2$. After the geometric optimization of $Ba 5s^2 5p^6 6s^2$ was completed, a single point energy calculation was made. After Sr doping, the lattice constants of the structure were calculated as $a=b=c= 8.001085 \text{ \AA}$ and the unit cell volume was calculated as 512.208 \AA^3 . The unit cell is slightly reduced by the Sr dopant. The unit cell volume of the pure material was calculated as $64,481 \text{ \AA}^3$. (Figure 1)

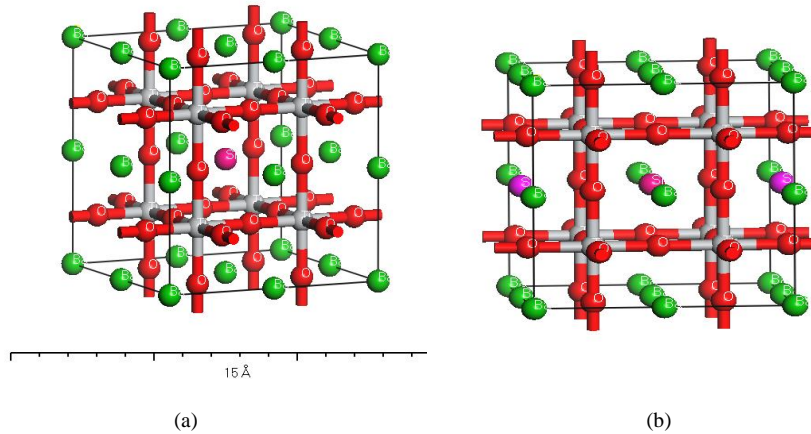


Fig. 2. (a) Non-Stoichiometric $Ba_{0.99}Sr_{0.01}TiO_3$ Structure (b) Non-Stoichiometric $Ba_{0.97}Sr_{0.03}TiO_3$ Structure

Table 1. Optimized Lattice Constants of unit cell volume and Sr-doped $BaTiO_3$

	Lattice Parameters (Å)			Volume (Å ³)
	A	b	C	
Literature ($BaTiO_3$)	4.034	4.034	4.034	64.481 [6]
Experimental Value	4.000	4.000	4.000	64.000 [6]
Literature ($BaTiO_3$)	4.03	4.03	4.03	65.67 [11]
BSTO ($Ba_{0.875}Sr_{0.125}TiO_3$)	4.03	4.03	4.03	65.28 [11]
Stoichiometric $BaTiO_3$ -This study	4.010	4.010	4.010	65.6459
$Ba(1-x)Sr_xTiO_3$ This study	4.005	4.005	4.005	64.240

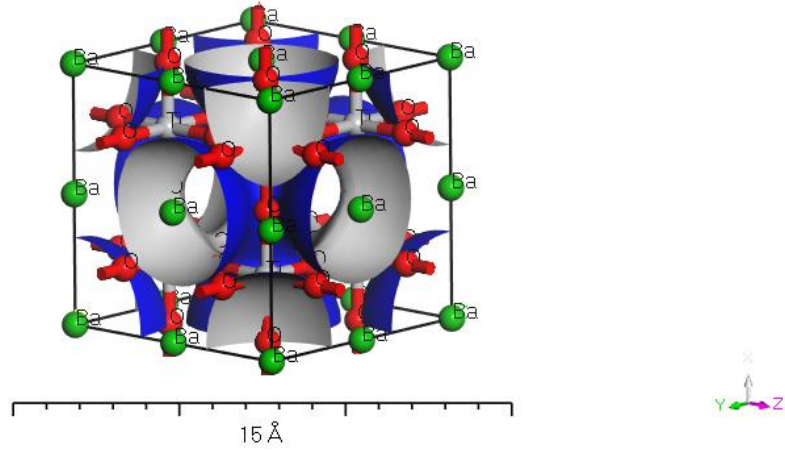
$BaTiO_3$ exists in cubic form with PM3M space group at 120 °C [7]. Quantum mechanics has been used to calculate the structural properties and electronic properties of $Ba(1-x)Sr_xTiO_3$. In this structure, Ba is located at the corners. The lattice parameters of $Ba_{(1-x)}Sr_xTiO_3$ are optimized. The lattice constants have calculated as $a = 4.010 \text{ \AA}$; $b = 4.010 \text{ \AA}$; $c = 4.010 \text{ \AA}$ in our study. This lattice constant value has calculated approximately close in value with the experimentally reported 4.000 \AA [8]. The differences between the value we have calculated and the experimental value is only 0.010 \AA and is close to 99.75%. This shows the closeness of this study to reality. A supercell was formed by shifting the stoichiometric structure $2 \times 2 \times 2$. For the supercell $a=b=c=8.020 \text{ \AA}$ and the volume is calculated as 515.849 \AA^3 .

3.2. Electronic properties of stoichiometric $BaTiO_3$ and non- stoichiometric $BaTiO_3$

The density of states (DOS), band structure, electron density . such electronic properties provide important information about the bonding nature of the material and its physical properties. The specific study of the binding properties of stoichiometric $BaTiO_3$ and Non-Stoichiometric $Ba_{(0.97)}Sr_{(0.03)}TiO_3$ is very important. [12] However, the impact of Sr doping on the electronic properties remains significant, as the band structure is altered by Sr being added to the Ba-site, which in turn affects the Ti-O binding.[14] We used GGA to calculate the electronic properties and k-points are used for these calculations. In order to understand the binding nature of these compounds

well, for (TDOS) the total density of states and for (PDOS) partial density of states were calculated. The first Brillouin region and doped electron density and Brillouin zone regions has shown in Figure3 (a) and Fig 3(b).

a



b

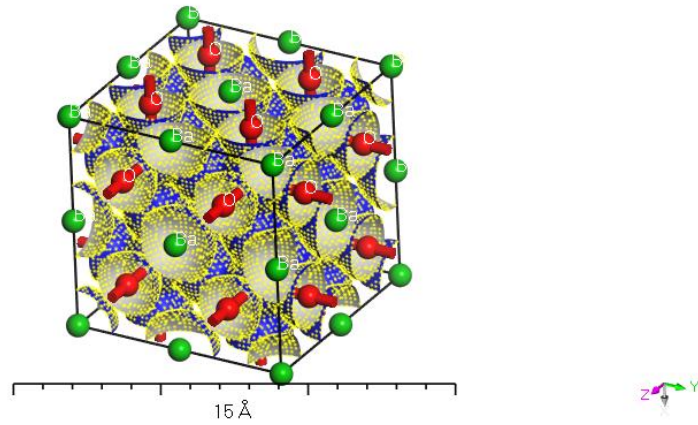
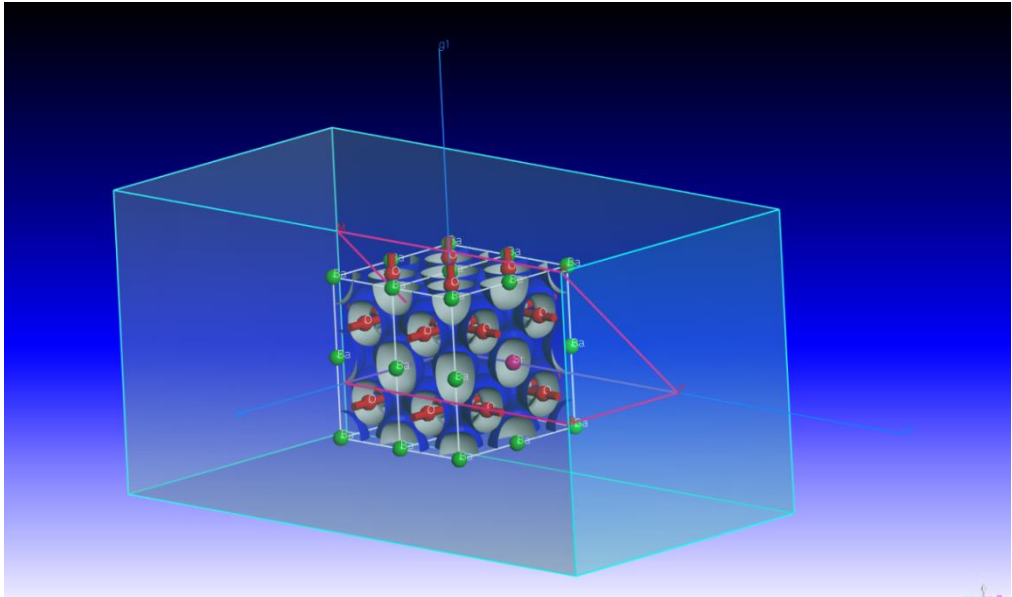


Fig. 3. (a) electron density of BaTiO_3 (b) electron density of $\text{BaTi}_{(1-x)}\text{Sr}_x\text{O}_3$

a



b

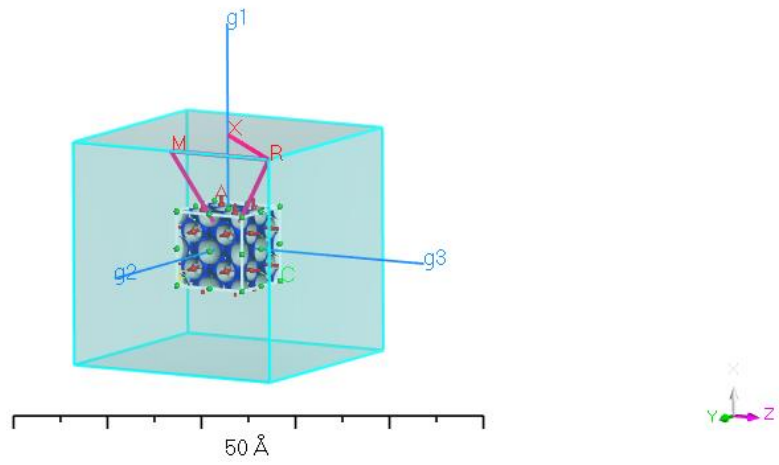
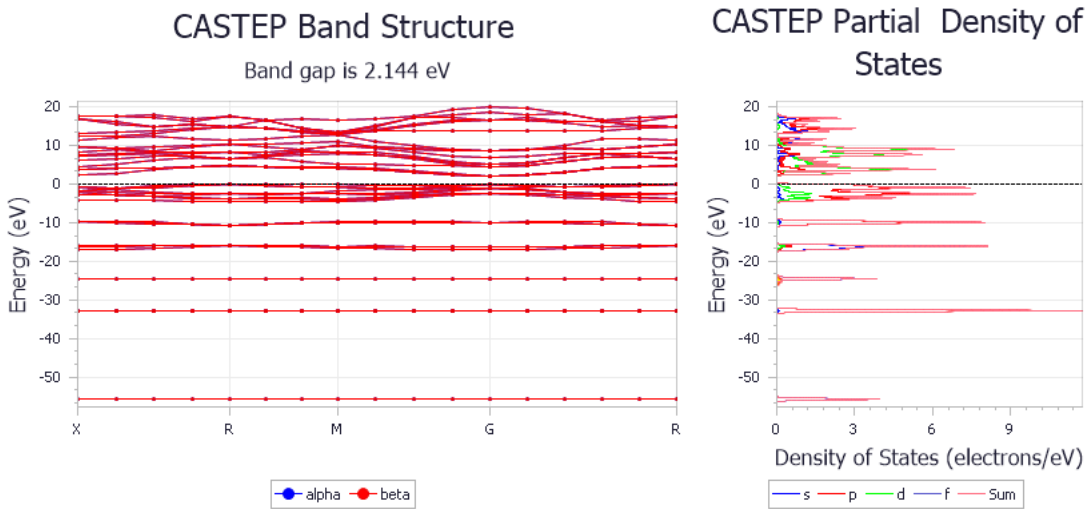


Fig. 4. (a) Brillouin zone path of BaTiO_3 ; (b) Brillouin zone path of $\text{BaTi}_{(1-x)}\text{SrO}_3\text{C}$

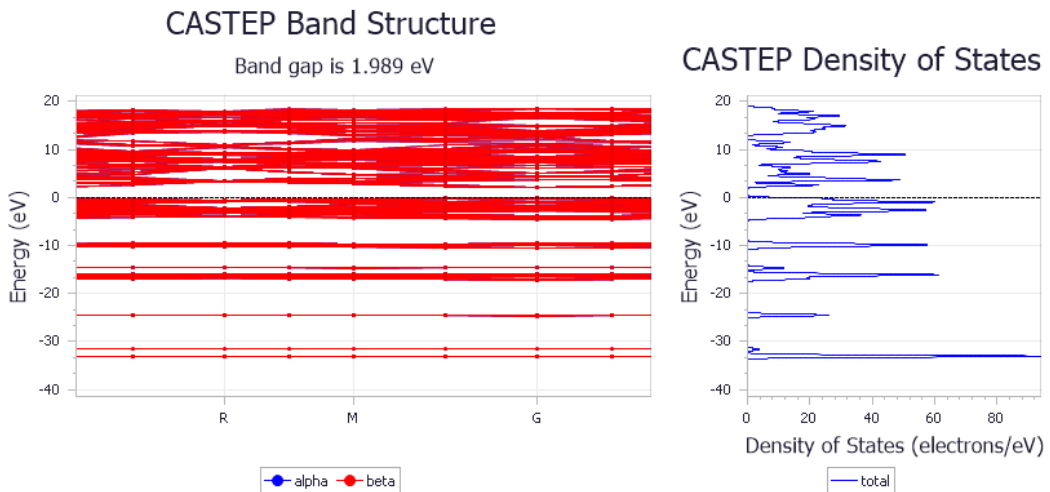
3.3. Band gap results

Materials band structure is examined as it provides useful information for understanding the nature of pyroelectricity and ferroelectricity.[6] The band structure of BaTiO₃ and doped Ba_(1-x)Sr_xTiO₃ is shown in Figures 3a and b. The electronic band gap diagram calculations in the direction of the highly symmetric Brillouin zone of cubic BaTiO₃ are shown in Figure 4 . The band gap of pure BaTiO₃ is found 2,144 eV, 1,989 eV for Ba_{0,99}Sr_{0,01}TiO₃ semiconducting and 1,911 eV for Ba_{0,97}Sr_{0,03}TiO₃ nature that had been shown in figure. The band gap of Ba_(1-x)Sr_xTiO₃ is lower than the band gap of pure BaTiO₃ They told that band gap reduces by doping. [12]

a



b



c

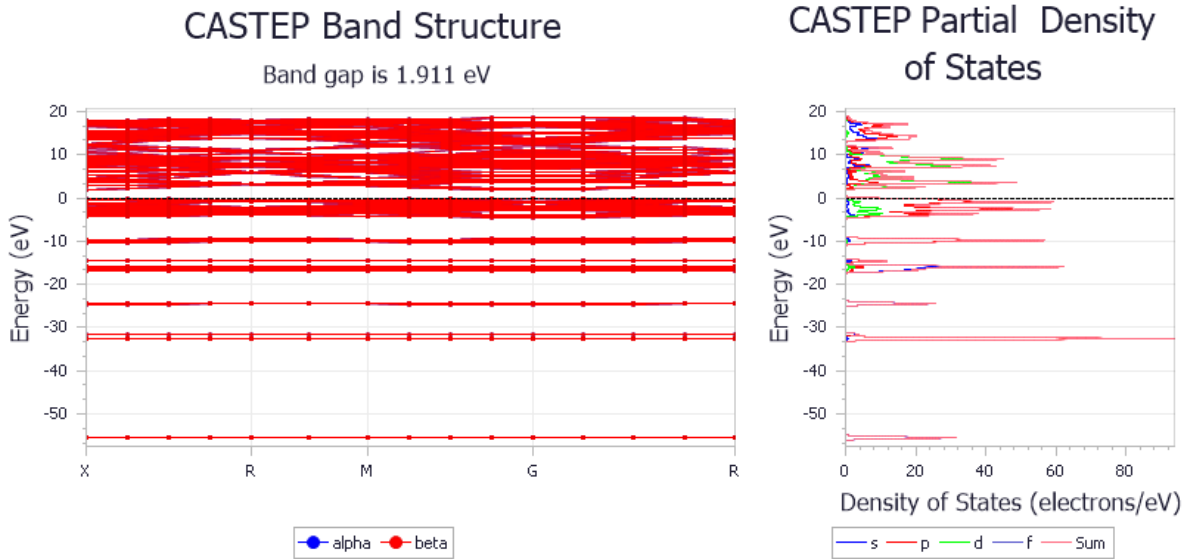


Fig. 5.(a) band gap of BaTiO_3 (b) band gap of %1 Sr doped $\text{Ba}_{(1-x)}\text{Sr}_x\text{TiO}_3$ (c) band gap of %3 Sr doped $\text{Ba}_{(1-x)}\text{Sr}_x\text{TiO}_3$

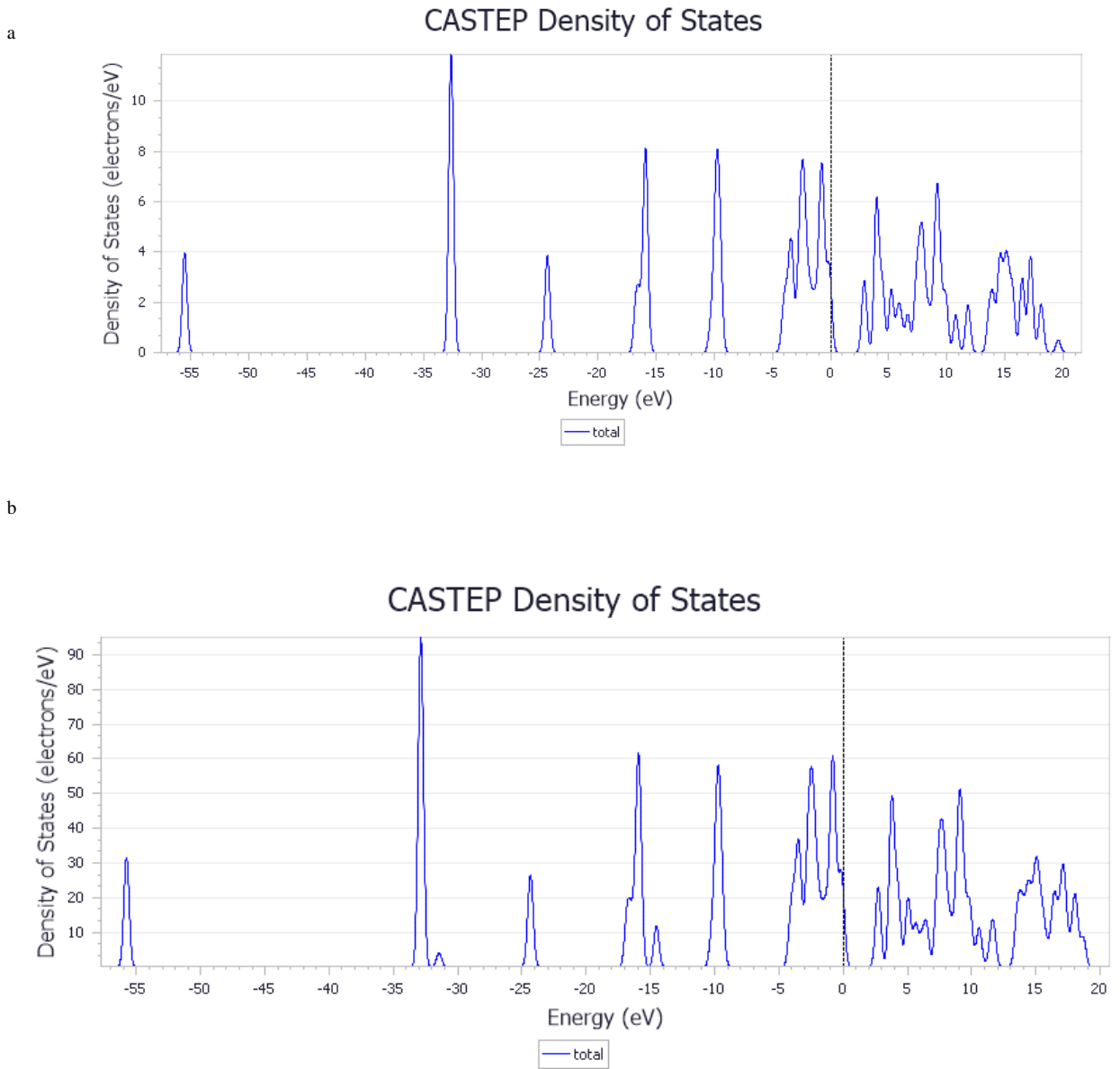
The minimums of the Conduction Band which have been controlled and pushed by the Titanium (Ti) 3d-states are located at the G point. The Valence Band maximum occurring is driven by the O-2p states of Oxygen. Figure 3 (a) shows the electronic band for stoichiometric $\text{Ba}_{(1-x)}\text{Sr}_x\text{TiO}_3$ and (b) non-stoichiometric $\text{Ba}_{(1-x)}\text{Sr}_x\text{TiO}_3$. In this study, the band gap at the R-G points is 2.144 eV and is indirect. This value is closer to the theoretical ~ 3.2 eV [10] Experimental band gap value compared to the band gap value of 1.723 eV reported in the previous study [8]. In DFT calculations, the value that we have calculated is smaller due to the pd pushing of the cation and the anion of the Generalized Gradient Approach application.

Calculations show that after doping BaTiO_3 with Sr in the Barium zones in Figure 4b, the electronic band structure shifts to the conduction band and the band gap value is 1,911 eV for %1 Sr doped eV as seen in Figure 5a and Figure 5b we can see from figure 5(c) the band gap value is 1,989 eV %3 Sr doped. As a result of our calculations, it was determined that the nature of the Sr-doped BaTiO_3 band structure changed directly from the indirect structure. The minimum energy required to excite the electron is determined by the band gap in semiconductors and insulators. However, it cannot fully explain whether the phonon will be absorbed by the material. It has been noticed that due to finite momentum the indirect band gap shows weak optical transmission. [9] The shift of the band gap after Sr doping in BaTiO_3 defines that the conductivity of the material is increased directly due to the easy recombination of electron hole in the band gap.

3.4. DOS-pDOS results

Electronic configuration for Ba is $5s^2 5p^6 6s^2$, for O $2s^2 2p^4$, Ti electronic configuration is $3s^2 3p^6 3d^2 4s^2$ where Sr configuration is $4s^2 4p^6 5s^2$. Figure 6(a) shows the total density of states for stoichiometric BaTiO_3 ,

figure 6(b) shows the total density of states for $\text{Ba}_{(1-x)}\text{Sr}_x\text{TiO}_3$ and figure 6(c) shows the partial density of states for stoichiometric and Sr doped BaTiO_3 systems.



c

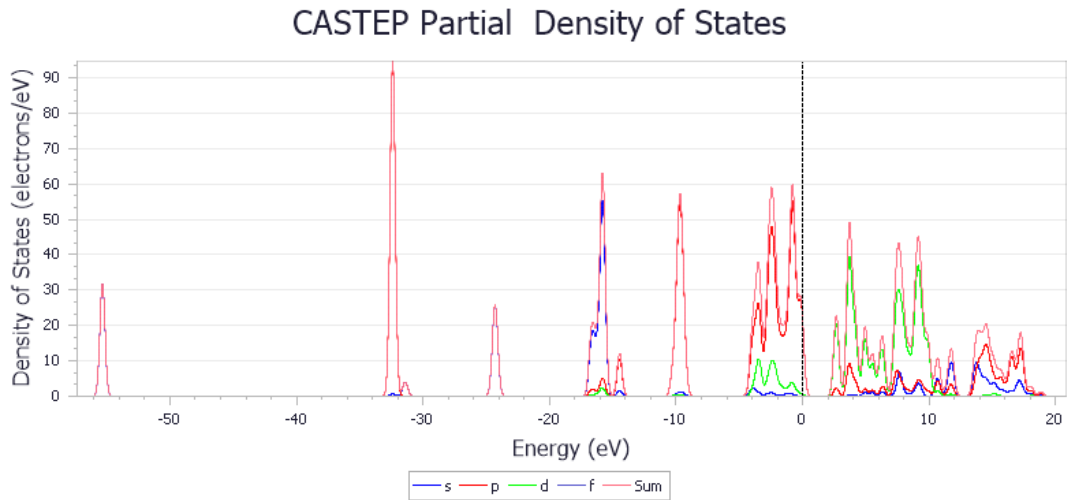


Fig. 6. (a) DOS Graph of BaTiO₃; (b) DOS Graph of %1 Sr doped Ba_(1-x)Sr_xTiO₃; (c) DOS Graph of %3 Sr doped Ba_(1-x)Sr_xTiO₃

In the examination, the primary importance of BaTiO₃ resides in the 6p states present within the valence band (VB), undergoing modifications upon the incorporation of Sr doping into the structure. We can say that the top of VB indicates the main contribution by the p-density of the states and the d-density of the states. In Figure 6, the density of states was increased after the addition of Sr. In Figure 6, p-DOS and d-DOS were considered to be the dominant contributors.

4. Discussion

In the study, our DFT calculation results in experimental and consistent with other studies. After Sr doping, the lattice constants of the structure were calculated as $a=b=c= 8.001085 \text{ \AA}$ and the unit cell volume was calculated as 512,208 Å³. The unit cell is slightly reduced by the Sr dopant. In this study, the band gap at stoichiometric BaTiO₃ R-G points is 2.144 eV and is indirect. As a result of geometric optimization of the non-stoichiometric Ba_(1-x)Sr_xTiO₃ structure and DFT calculations, it was determined that the band structure switch towards the conduction band. The band gap decreased to 1,911 eV for %1 Sr dopant to the BaTiO₃ system and for %3 dopant to the BaTiO₃ system the band gap decreased to 1.989 eV while the band gap is 2,144 eV for stoichiometric BaTiO₃

First, the band gap decreased with BaTiO₃ Sr doping. Secondly, the band gap has also changed directly to the band gap. In that case, BaTiO₃ produced by Sr doping is suitable for electronic devices.

We think that this study will contribute to, at the beginning of experimental studies to be carried out for future optoelectronic device applications, thus benefiting from saving time and materials.

Acknowledgements

I would like to thank Kütahya Dumlupınar University for the use of the Materials Studio program in this study.

References

[1] M. Vijatović, J. Bobić, & B. Stojanović,, History and challenges of barium titanate: part I, *Science of Sintering*, 40(2), 155-165,

doi:10.2298/SOS0802155V, 2008

- [2] S. Dahbi, N. Tahiri, O. El Bounagui, H. & Ez-Zahraouy, Electronic, optical, and thermoelectric properties of perovskite BaTiO₃ compound under the effect of compressive strain, *Chem. Phys.*, doi:10.1016/j.chemphys.2021.111105, 2021
- [3] F.D. Morrison, D.C. Sinclair, & A.R. West., Doping mechanisms and electrical properties of La-doped BaTiO₃ ceramics, *Int. J. Inorg. Mater.*, 3(8),1205-1210, 2001
- [4] F., Yang., First-principles investigation of metal-doped cubic BaTiO₃, *Mater. Res. Bull.*, 96(4), 372-378, doi.org/10.1016/j.materresbull.2017.03.023
- [5] C. Soykan, *Yeni manyetik şekil hafızalı alaşımların yoğunluk fonksiyonel teorisine dayalı ab-initio metodu ile tasarlanması: ni-fe-ga.*, Doktora Tezi, Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, Denizli, 130s., 2014
- [6] L. Hedin, New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem. *Physical Review*, 139, 796-823, doi.org/10.1103/PhysRev.139.A796, 1965
- [7] C. Soykan, H.Gocmez, The physical properties of bismuth replacement in lead halogen perovskite solar cells: CH₃NH₃Pb_{1-x}Bi_xI₃ compounds by ab-initio calculations, *Results in Physics*, 13,102278, doi.org/10.1016/j.rinp.2019.102278, 2019
- [8] R. King-Smith, D. Vanderbilt, A first-principles pseudopotential investigation of ferroelectricity in barium titanate, *Ferroelectrics*, 136 (1) 85–94., 1992
- [9] M.Rizwana., H. Zebab, M.Shakila, S.S.A. Gillanic, & Z.Usmand, Electronic, structural and optical properties of BaTiO₃ doped with lanthanum (La): Insight from DFT calculation, *Optic - International Journal for Light and Electron Optics*, 211,164611
- [10] H.Shen, K.Xia, P.Wang, & R.Tan., "The electronic, structural, ferroelectric and optical properties of strontium and zirconium co-doped BaTiO₃": First-principles calculations, *Solid State Communications*, 355, 114930, 2022
- [11] A.Boubaia, A. Assali , S. Berrah , H. Bennacer, I. Zerifi ,& A. Boukourt,Band gap and emission wavelength tuning of Sr-doped BaTiO₃ (BST) perovskites for high-efficiency visible-light emitters and solar cells, *Materials Science in Semiconductor Processing*, 130 ,105837, 2021
- [12] M.Hasan, A.K.M.,& A.Hossain, Structural, electronic and optical properties of strontium and nickel co-doped BaTiO₃: A DFT based study, *Computational Condensed Matter*, 28, 578, 2021
- [13] A. Rehman , A. Muhammad Iqbal , A. I. Channa , S. U. Awan , & M. T. Khan, Structural, electronic and optical properties of BaTiO₃-CoFe₂O₄ nanocomposites for optoelectronic devices, *Materials Science and Engineering: B*, 296, 116626, 2023
- [14] C. Sidar, M.N. Tripathi, & P.K. Bajpai, Effect of Sr-doping on the band structure of BaTiO₃ through density functional theoretical calculations, *Computational Condensed Matter*, 11, 27-32, 2017
- [15] S. Aksan, Determination of Structural and Electronic Properties of Sr Doped BaTiO₃ Using Density Functional Theory, *XI. Ceramics Congress with International Participation*, the proceedings summary book, 2022
- [16] Aksan, S.,Hussaini, A.A., Erdal, M.O. et al. Investigation of photosensitive properties of novel TiO₂:Cu₂O mixed complex interlayered heterojunction: showcasing experimental and DFT calculations. *Opt Quant Electron*, 56, 578, doi.org/10.1007/s11082-023-06266-7, 2024



E-ISSN: 2687-6167

Number 56, March 2024

REVIEW ARTICLE

Receive Date: 08.03.2024

Accepted Date: 25.03.2024

Urban planning and development in harmony with the geosciences

Ümit Yıldız^{a*}

*Lead Instructor, Black Hills Natural Science Field Station, South Dakota School of Mines, Rapid City, SD, U.S.A.
ORCID: 0000-0002-3843-7203*

Abstract

Urban geology is the study of the different geological elements that impact and restrict human activities in engineering and economics inside urban areas. Over half of the world's population, or 4.2 billion people, lived in urban areas in 2018. Projections suggest that by 2030, this figure will rise to nearly 5 billion. Notably, the majority of this urban growth is anticipated to occur in developing countries, with towns and cities in these regions accommodating around 80% of the urban population by 2030, as stated by the United Nations Population Fund in 2007. A quick and comprehensive review of the literature reveals the growing importance of urban geology as an emerging area of study as well as the vitality of geosciences for natural disaster mitigation, resource management, sustainability, and understanding geological processes and natural hazards. Interdisciplinary research and collaboration between geologists, engineers, architects, urban planners, and policymakers at the national and local levels is inevitable given the current acceleration of urbanization and rapid environmental degradation as a backdrop. Geology is also an essential part of site selection, infrastructure design, and construction, water resources management, land use planning, and environmental protection during the urban planning and development phases. Incorporating geoscience insights into planning processes and raising public awareness allows cities to be planned and managed in ways that promote sustainable development, and resilience to natural disasters, and safeguard residents' health and well-being. In order to construct safer, more resilient, and sustainable cities for our communities and future generations, geoscience education and research must be acknowledged within the scientific research agendas related to urban planning, development, and transformation endeavors. Given the above, this study aims to examine the close relationship of geosciences with urban planning and development activities and to investigate and analyze the impact of geological factors on the urbanization process. This research also aims to raise public awareness about the importance of geology among the people who live in urban areas.

© 2023 DPU All rights reserved.

Keywords: Urban Planning; Urban Geology; Urban Development; Urban Transformation

* Corresponding author. Tel.: 001-605-394-2494; fax: +0-000-000-0000 .

E-mail address: umit.yildiz@sdsmt.edu

1. Introduction

Urbanization and the special challenges it presents to geologists led to a steady evolution of the idea of urban geology, which deals with understanding geological processes, geohazards, and resources within urban environments [1]. Urban geology is not the same as environmental geology; rather, it is the study of land resources and geologic hazards as they relate to the growth, rehabilitation, and expansion of urban areas [3]. Urban geology is the study of the different geological elements that impact and restrict human activities in engineering and economics inside urban areas [2]. Urban geology holds significant importance within the field of environmental geology, serving as a crucial component in human engineering and construction endeavors [3]. Its primary focus is on urban construction planning, making it a practical geological discipline. Although geology has always had an impact on cities, the phrase "urban geology" and the field's focused investigation only began to take shape in the middle of the 20th century, particularly in California. The U.S. Geological Survey (USGS) started mapping cities in the early 1900s after realizing the value of comprehending the geology of urban regions, even though the term had not yet been coined. The phrase "urban geology" was assigned to John C. McGill in a 1964 USGS publication [4]. Urban geology originated in North America in the late 1960s and early 1970s. The research and strategies to reduce the urban geological hazards in Asia have gained momentum after the integration of the field of geology with social and economic aspects in the 1990s. The role of urbanization and urban geology in this case have been brought to light and therefore numerous research in this field have been carried out. These studies are of significance as they bring the subject of urban planning and geology to the table. For instance, Zengin [5] conducted a research on the inundation effects of the Antalya Gulf, an over-developed tourist area located on the Mediterranean coasts of Türkiye. The findings of this study show that the sea level rise in the following 60 to 80 years is likely to lead to the loss of more than 250 square kilometers of land and the radical change of the coastline [5]. Another research [6] has proposed the Tomorrow's Cities Decision Support Environment (TCDSE), which is a method for advanced physics-based optimum determination of hazard and engineering modeling. This interdisciplinary approach is comprised of a combination of the physical and social effects and a flexible method for quantifying impacts which makes the concept of risk more understandable to the general public. Also, the study was meant to be flexible and could be applied to different scales of urban areas, involving steps of stakeholder engagement and technical tasks. This comprehensive study of the issue indicates that future urban planning may be done considering multi-hazard risks by means of a new approach for decision-making [6]. Furthermore, in Denmark, a 3D model of the geotechnical characteristics and geology of the urban regions was built in 2020 for urban planning and development efforts [7]. During this study, multiple-scale geophysical and geological maps were conducted in conjunction with the well data and geotechnical investigations to assess the functions of the geoscientific survey techniques utilized in mapping the geology, hydrology, and topsoil characteristics of the urban regions in Denmark [7]. After the devastating earthquakes in Türkiye in 2023, many municipalities swiftly implemented urban transformation plans to ensure buildings are resilient against earthquakes and began to embrace the importance of urban geology more significantly.

Over half of the world's population, or 4.2 billion people, lived in urban areas in 2018 (Figure 1). Projections suggest that by 2030, this figure will rise to nearly 5 billion. Notably, the majority of this urban growth is anticipated to occur in developing countries, with towns and cities in these regions accommodating around 80% of the urban population by 2030 [8]. Sustainable cities are one of the 17 Sustainable Development Goals of the United Nations set forth to be accomplished by 2030 [9]. The organization has identified major objectives for the upcoming ten years, including resistance to geohazards, sustainable development, responsible urban design, and preservation of natural heritage [10]. Unplanned urbanization, a significant consequence of population growth, is frequently intensified by the influx of people migrating from rural to urban regions [11]. Urban regions are commonly marked by high levels of industrial activity, often without proper regulation, as well as rapid and inadequately planned expansion. Additionally, natural habitats are frequently fragmented, and both surface and groundwater quality suffer from degradation caused by various chemical pollutants [12]. Urbanization concentrates human ingenuity, propels

the world economy, and produces major social benefits. However, it is also an important factor in most environmental problems [12].

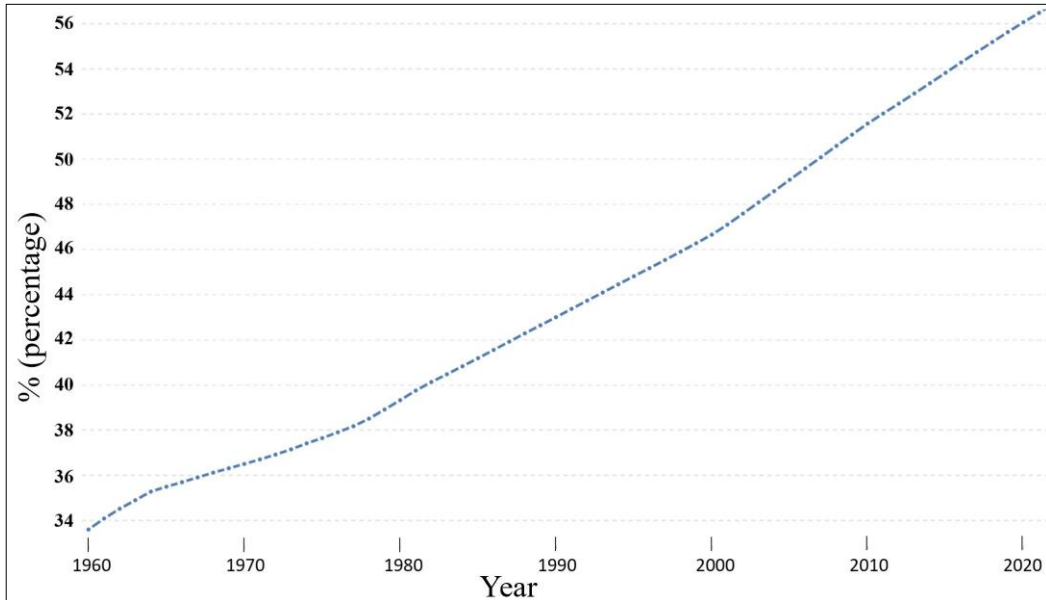


Fig.1. The chart illustrates the fluctuation in the percentage of global urban populations over the years [9].

Economic and social variables are typically the main emphasis of urban planning, development, and management processes [8]. Many people participating in the planning process have little interest in geoscience knowledge unless it is driven by a specific development project or hazard. A portion of this apathy stems from the fact that geoscience topics are rarely adequately covered in fundamental education; topics like minerals, water supply, and environmental hazards are usually saved until later grade levels. Convincing individuals to give geoscience issues top priority might therefore be difficult. With the exception of natural phenomena like earthquakes, volcanic eruptions, and tsunamis, the general public's perspective of geology is that it deals with events that happen on a timescale that seems remote and unrelated to present issues [13]. Rather, economic and social concerns including housing, work, health, and crime rates are more often given priority [13]. An overview of the main factors considered in urban geology, urbanization, and urban planning and development was given in Table 1, highlighting their interconnections and significance in sustainable city development. According to Marker [13], the significance of integrating foundational data into the creation of efficient planning regulations, sustainability analyses, and environmental impact assessments is recognized by geoscientists. In addition, this information offers crucial background for carrying out site assessments and determining the requirements for planning and environmental permits, which was provided by Marker [13] as follows: Quarry locations, mineral protection zones, water resource protection, agricultural preservation, ground condition assessment, conservation sites, hazard zones, pollution control, waste management zones. Furthermore, the main geological problems that need to be focused on before starting urban transformation or urbanization plans are provided in the European Union's Urbanization Plan as follows [14]:

- Landslides
- Earthquakes

- Sinkholes
- Land subsidence
- Groundwater pollution
- Volcanic activity
- Floods/Tsunamis
- Erosion and deposition
- Saline soils

To address these fundamental geological challenges in urban areas, it is essential to ensure the availability of the data outlined below. If this data is not already accessible, efforts should be made to obtain it promptly [14]:

- Geological maps.
- Groundwater depth/flow/quality maps.
- Maps that display catchment area locations.
- Maps illustrating the locations of rock quarries, pits, and mining activities.
- Locations maps for ground subsidence.
- Resource maps illustrating the distribution of possible landslides.
- Soil and water pollution and geochemical distribution maps.
- Maps for foundation characteristics and levels displaying the types and depths of foundation levels.
- Soil maps indicating the soils prone to erosion.
- Geologic map of Quaternary (alluvium, colluvium, terrace, etc.) units.
- Distribution map for urban and agricultural lands.
- Literature and publications

Table 1. The table presents a comprehensive outline of the key elements examined in urban geology, urbanization, and urban planning and development, emphasizing their interrelations and importance in fostering sustainable city growth.

Factors	Urban Geology	Urbanization	Urban Planning and Development
Geological Hazards	Identification of seismic zones, fault lines	Impact of geological hazards on urbanization	Incorporation of hazard mitigation plans
Subsurface Conditions	Soil types, groundwater levels, bedrock geology	Influence of subsurface geology on infrastructure	Consideration of subsurface conditions in construction
Land Use Patterns	Geological suitability for different land uses	Impact of urbanization on land use patterns	Zoning regulations based on geological factors
Natural Resources	Identification and management of mineral deposits	Urbanization's impact on resource depletion	Sustainable resource management in urban areas
Environmental Impact	Assessment of geogenic impact on ecosystems	Anthropogenic effect on environmental degradation	Implementation of measures to mitigate environmental impact
Infrastructure	Geological factors affecting infrastructure stability	Urbanization's demands on infrastructure	Planning and development of resilient infrastructure
Geological Mapping	Mapping of geological features and hazards	Incorporation of geological data in urban planning	Utilization of geological maps for city planning

Land Stability	Evaluation of slope stability, landslide susceptibility	Urbanization's impact on land stability	Implementation of measures to ensure land stability
Geological Surveys	Conducting surveys for land development projects	Addressing geological concerns in urban growth	Utilization of surveys for informed decision-making
Climate Change Impacts	Understanding geological responses to climate change	Impact of climate change on urban areas	Integration of climate change adaptation strategies

Sustainable urban growth, planning, and transformation strategies, as well as environmental preservation and comprehensive land use planning, depend on these steps. As cities continue to evolve, it is critical to understand the role that geology plays in urban development and planning. For example, the catastrophic effects of Türkiye's 2023 earthquakes on urban areas have once again underlined the importance of urban geology in the city and regional development, and transformation efforts in cities.

The main goal of this study is to examine the close relationship of geosciences with urban planning and development activities and to investigate and analyze the impact of geological factors on the urbanization process. This research also aims to raise public awareness about the importance of geology among the people who live in urban areas.

2. Collaborative approach

An interdisciplinary working group of engineers, architects, urban planners, geologists, and city authorities is needed to offer a solution-oriented approach to adopt all the kinds of challenges faced by local people during urban planning, development, and transformation activities [15]. With the rise in the population of cities in recent years which has made the socio-economic issues more complicated, it has become inevitable to set up the social-economic working group within the local governments. Interdisciplinary work is a cooperation that is done through holistic and integrative processes, which are based on the individual analysis of data by each discipline and the integration of the knowledge, as the creation or expansion of safe and sustainable urban areas [16]. In particular, geoscientists can share critical information with this working group by conducting comprehensive studies to identify and, if possible, prevent potential geological natural disasters in the environments where urban areas are planned to be established. On the other hand, architects and civil engineers play an important role in creating safer and more robust building stocks by adapting the critical data provided by geologists to their designs to ensure infrastructure stability and the safety of human life and property. Additionally, urban planners, in the light of the data they receive from geologists, make urban transformation or expansion plans in places where environmental factors will have the least impact. This interdisciplinary approach will not only minimize environmental destruction but also minimize the effects of any natural disasters that could potentially harm people. For all these reasons, it is very important for a sustainable and safe urban life that geoscientists conduct comprehensive studies before urban planning, development, and transformation activities and share this data with other disciplines.

3. Geological factors in site selection

In countries like Türkiye, Japan, New Zealand, and Indonesia, where tectonic movements occur frequently, it is a well-known fact that every devastating earthquake and natural disaster reminds people that geology is particularly critical for urban development and urban planning. However, despite the pain and loss of human lives, people quickly forget the importance of geology in urbanization. In fact, the selection of sustainable and safe settlements, prior to urban planning and transformation, can be optimized if enough attention is paid to interdisciplinary scientific studies.

The stability and integrity of soil mechanics, slope morphology, and vulnerability to geological hazards are critical considerations when identifying secure and safe sites for urban development. Although the exact timing of many natural disasters cannot be known, it is a fact that comprehensive geological studies can determine where they may occur. It is therefore essential for geoscientists to play one of the most important roles in guiding studies for urban planning and development initiatives and taking measures to minimize the loss of human life and property. Geoscientists use remote sensing and integrated geographic information systems (GIS), as well as Multi-Criteria Decision Analysis techniques that are integrated with artificial intelligence to investigate and identify potential natural disasters in urban areas that may impact the environment as well as the residents of that urban area [17-18-19-20-21]. Based on the findings of those studies, geologists can take measures accordingly for the well-being of the residents. With these techniques, geologists conduct qualified research in many areas such as proximity to natural disasters, adaptation to climatic conditions, environmental factors that may affect potential urban development, and mitigation of their effects. The number of these studies has increased considerably in recent years, however, more studies need to be conducted. These studies and assessments serve to reduce and minimize the risks associated with geological hazards by protecting the infrastructures in urban areas through the systematic study of potential natural disasters associated with geological factors, creating a building stock resistant to these disasters, and protecting the safety of people's lives and property. Thus, geoscience and its experts should be included in urban planning and development processes, and geology should play a key role in making informed decisions about the urban ecosystem and natural disasters. In this way, before urban planning and development, decision-makers and city authorities make sure that their decisions are based on the science that is more constructive and prioritizes human health and safety.

4. Infrastructure design and construction

In a broader sense, urban infrastructure refers to the built environment, which includes structures and equipment for energy, transportation, water supply, sewage treatment, and solid waste disposal [22]. The quality of life and comfort of urban people is, in reality, determined by the state of the urban infrastructure [23]. Environmentally friendly urban infrastructure design and construction is increasingly being seen as a preferable means of providing state-of-the-art infrastructure such as transportation and utilities to densely populated urban areas. Additionally, geological factors heavily influence the design and construction of urban infrastructure, such as buildings, bridges, tunnels, highways, etc. The local geology has a major influence on infrastructure, particularly subsurface infrastructure. Therefore, understanding subsurface geology is essential to ensuring the reliability and sustainability of these structures, particularly in terms of minimizing the risks associated with geological hazards such as earthquakes, land subsidence, and volcanic eruptions. For instance, a volcanic eruption devastated a tiny village in Iceland in 2023 and 2024. Urban areas were evacuated in time thanks to geoscientists' diligent efforts in providing warnings in advance. As a result, while the volcanic activity caused substantial property damage, no lives were lost. Additionally, engineering solutions that are specially adapted to geological conditions are necessary to improve the strength and lifespan of urban infrastructure systems. Another example of a natural catastrophe is inundation. Geoscientists might conveniently create risk maps to assess urban areas near coasts and rivers. A recent study [24] examined 464 historic places along Türkiye's and Greece's Eastern Mediterranean coastlines. According to the findings of this study, a large number of archeological and historical sites will be inundated even under the most minimal sea level rise projections [24]. Moreover, soil liquefaction is a primary cause of structural damage in earthquakes. Liquefaction studies are critical for disaster mitigation planning, particularly in urban settings. Another study [25] investigated the socioeconomic vulnerability of Greater Chennai, India in terms of seismic hazard risk, namely soil liquefaction. According to this local study, 19.4% of the study area falls into the high-risk group [25].

The resilience and durability of the buildings and infrastructures are critical because they improve the ability to deal with the uncertain and extreme natural disasters that may occur during the extended life cycle of urban infrastructure. There is an increasing recognition that sustainability should be incorporated and implemented in

geological engineering, but little emphasis has been placed on adding resilience and durability, even though sustainability and resilience share common objectives and goals [26].

5. Water resources management

While water makes up 71% of the earth's surface, only less than one percent can be used for human consumption [27]. Furthermore, less than one percent of freshwater that is available for human use is not only being consumed unconsciously but is also being contaminated recklessly. Anthropogenic activities such as agriculture, industry, and urban development can easily disrupt the hydrogeological and geochemical conditions of urban groundwater [28-29-30-31-32-33-34-35-36-37-27]. On the other hand, natural processes and parameters like lithology, groundwater velocity, recharged water quality and quantity, water's interactions with bedrock and soil components, and groundwater networks with different aquifer types influence a region's groundwater quality [38-39-27]. Hydrogeology is a branch of geosciences that studies all of these key factors and processes which is essential for urban water management efforts. Furthermore, effective floodwater management is also required for urban areas. Because urban environments are more vulnerable to the effects of heavy rainfall; thus, designing reliable stormwater infrastructure requires an interdisciplinary approach incorporating meteorological engineers and geoscientists. Moreover, it is yet unclear how climate change may affect water supplies, thus research on this topic is also necessary. The sustainable management of rainfall and stormwater runoff in urban contexts might be an opportunity for a yield of renewable water when effective flood management is on the table. Recently, local and national authorities of developed countries have been implementing small- to medium-scale sustainable urban development policies, which encourage private homes and businesses to collect rainwater and use it for greywater applications [40]. While this application has demonstrated success in Europe, Australia, North America, and some developing countries (e.g. Türkiye), it has not yet been tested in underdeveloped nations. Urbanization and urban redevelopment plans must take geological considerations into account more than ever before, given the increasing impact of climate change and the growing importance of water supplies.

6. Land use planning and environmental protection

Urban areas are the centers of human civilization, but the foundation of these civilizations rests on the science of geology which is often overlooked. Thanks to well-trained geoscientists and engineers, it is no longer a luxury to learn the geology of an urban area and to model and plan buildings accordingly. For example, in urban areas where earthquakes, sinkholes, and landslides commonly occur, buildings, which are one of the most important symbols of civilization, should be built on solid ground instead of soft alluvial soil. Prohibiting construction in a region with known valuable resources (e.g. minerals, groundwater, hydrocarbons), but mining with environmental sensitivity and responsibility, and offering the extracted natural resources to the service of humanity or taking precautions to prevent pollution by human and environmental factors while determining the availability and accessibility of those resources are all area of interests of geosciences. In order to ensure that society will have access to natural resources, land-use planning must protect this access. Multiple studies have been carried out on land use and land cover (LULC) assessments to promote sustainable urban planning through the use of remote sensing and GIS technologies. A study examined LULC analyses for sustainable urban development in Kütahya, Türkiye [41]. This study resulted in the production of LULC maps through the use of unsupervised learning techniques, GIS software, and the Landsat and Sentinel-2 remote sensing data from 2017 and 2021. The research looked at the effectiveness of the automated classifications in developing accurate LULC maps for 2017 and 2021 [41]. Another research [42] held in Tehran, Iran, aimed at investigating the urbanization process. This study used Landsat time series imagery from Google Earth dating from 1991 to 2021 and constructed LULC maps using these images. The study showed that there was a substantial rise in urban land cover, which is an indication of the fast urbanization of Tehran over the last three decades [42]. In another study, LULC analysis was carried out around the Jatibarang Reservoir in

Semarang City, Indonesia [43]. This research has employed the Nearest Neighbor Analysis as the spatial analysis tool. The fact is the results of this study have given a special and all-embracing view of the changes, patterns, and possible consequences of LULC in this specific research area [43].

Protecting access to important natural resources—if not all of them—or at the very least bringing attention to their presence has been demonstrated to be possible with the help of recognized scientific approaches [44]. Cities all across the world require these kinds of solutions now more than ever because of growing urban populations and increasing rates of land consumption [45]. As the population of cities grows and their size increases, responsible and effective environmental practices have become one of the most critical issues to avoid. For instance, the construction industry can minimize environmental impact by using more resource-friendly and low-carbon building materials based on existing geological resources. In addition to buildings and infrastructure, green spaces and parks are also needed for a sustainable life in cities. Local geology plays a critical role in determining suitable locations for parks and green spaces in urban areas that are not suitable for development, so local geology should be well studied and understood.

Soil and water pollution is likely to occur with urbanization. Geological surveys and studies play an important role in identifying pollution hotspots and the sources of pollution. Geology can also provide adequate consultancy services through the sub-disciplines of geology, such as calculating the damage caused by urbanization to the environment and taking necessary measures to protect the environment. A good understanding of geology is essential for successful and sustainable urban planning, development, and transformation. By understanding the geological landscape and prioritizing environmental protection, cities can be established that are not only functional and aesthetically pleasing but also safe, sustainable, and more inhabitable for future generations.

7. Discussions

Incorporating geosciences into processes of urban planning, development, and transformation has gained greater attention in recent years. Urbanization activities clearly demonstrate the necessity for interdisciplinary working groups and collaboration among diverse specialists, including geologists, engineers, architects, urban planners, and city officials, given that some geological factors in urban areas are susceptible to natural disasters. Urban places can be transformed into safer, more inhabitable, and more sustainable by bringing together all relevant stakeholders and utilizing all of their qualifications.

It is impossible to overestimate the importance of thorough geological surveys and studies in determining the best sites for urban development initiatives and plans for urban transformation. Precautions against natural disasters require comprehensive geological studies such as soil mechanics, slope morphology, and fault mapping, particularly for areas with known seismicity. The safety of the inhabitants is one of the most crucial factors in choosing the right locations for urbanization. In this context, geoscientists can assess and interpret potential risks of natural disasters using cutting-edge technologies like artificial intelligence, geographic information systems, and remote sensing. They can also assist people by taking the necessary precautions to mitigate the devastating effects of natural disasters.

Additionally, in regions where development is anticipated, the sustainability of infrastructure will be determined by the local geology. As a result, before planning and executing that plan, comprehensive geological studies are crucial. Furthermore, to increase the resilience of urban infrastructure systems and so guarantee the safety of life, property, and the well-being of urban residents, engineering studies that are robust as well as tailored to local geological settings are necessary.

In urban settings, anthropogenic and geogenic activities can have a substantial impact on both the quantity and quality of groundwater. To minimize dependency on groundwater alone, sustainable techniques like rainwater harvesting ought to be made mandatory in urban areas. Additionally, measures like preventing illegal well drilling and managing groundwater and rainfall runoff effectively can assist reduce the issue of water shortages in urban areas. Moreover, the scarcity, fluctuation, and contamination of water are increasing due to climate change. Extreme

weather events have an adverse influence on biodiversity, sustainable development, and people's access to clean water and sanitation throughout the water cycle. Urban communities hence require a well-thought-out plan for managing their water resources.

In addition to advising where urbanization and urban transformation initiatives are most appropriate, geology is also crucial in locating contaminated areas of water and soil and identifying their potential sources. Finally, viable and sustainable urban living depends on recognizing geology's crucial role in urbanization. Considering all of these aspects, geology's often underappreciated significance must be acknowledged.

8. Conclusion

In summary, this research emphasizes that geosciences are now an inherent part of urban planning, development, and transformation, in that fundamental geological data obtained from comprehensive research may assist decision-making in reducing the potential hazards of natural disasters, improving infrastructure resilience, and serving as insurance by justifying the sustainable use of water resources. This paper also highlights the importance of an interdisciplinary approach prior to urban planning and development. Interdisciplinary workshops can address and manage urban and environmental concerns and potential hazards. Finally, geoscience education and research should take precedence in urban planning measures to develop or reconstruct safer, more resilient, and environmentally friendly future cities for future generations.

Acknowledgments

This review study received no specific grants from any funding agency in public, commercial, or non-profit organizations.

References

- [1] Ş. Tüdeş, K. B. Kumlu, and S. Ceryan, "Integration between urban planning and natural hazards for resilient city." In *Handbook of Research on Trends and Digital Advances in Engineering Geology*, 591-630. Igi Global, 2018. doi:10.4018/978-1-5225-5646-6.ch055
- [2] D. Fuchu, L. Yuhai, and W. Sijing, "Urban geology: a case study of Tongchuan City, Shaanxi Province, China," *Engineering Geology*, vol. 38, no. 1-2, pp. 165-175, 1994.
- [3] M. G. Culshaw and S. J. Price, "The 2010 Hans Cloos lecture: the contribution of urban geology to the development, regeneration and conservation of cities," *Bulletin of Engineering Geology and the Environment*, vol. 70, pp. 333-376, 2011.
- [4] J. T. McGill, "Growing importance of urban geology," United States Department of the Interior, U.S. Geological Survey, No. 487, 1964.
- [5] E. Zengin, "A Combined Assessment of Sea Level Rise (SLR) Effect on Antalya Gulf (Türkiye) and Future Predictions on Land Loss," *J Indian Soc Remote Sens*, vol. 51, pp. 1121-1133, 2023, doi:10.1007/s12524-023-01694-0
- [6] G. Cremen et al., "A state-of-the-art decision-support environment for risk-sensitive and pro-poor urban planning and design in Tomorrow's cities," *International Journal of Disaster Risk Reduction*, vol. 85, p. 103400, 2023.
- [7] T. R. Andersen, S. E. Poulsen, M. A. Pagola, and A. B. Medhus, "Geophysical mapping and 3D geological modeling to support urban planning: A case study from Vejle, Denmark," *Journal of Applied Geophysics*, vol. 180, p. 104130, 2020.
- [8] G. D. Bathrellos, K. Gaki-Papanastassiou, H. D. Skilodimou, D. Papanastassiou, and K. G. Chousianitis, "Potential suitability for urban planning and industry development using natural hazard maps and geological-geomorphological parameters," *Environmental Earth Sciences*, vol. 66, pp. 537-548, 2012.
- [9] United Nations, "Sustainable development knowledge platform," Retrieved February 18, 2024, from <https://sustainabledevelopment.un.org/sdgs>.
- [10] A. A. Kutty, G. M. Abdella, M. Kucukvar, N. C. Onat, and M. Bulu, "A system thinking approach for harmonizing smart and sustainable city initiatives with United Nations sustainable development goals," *Sustainable Development*, vol. 28, no. 5, pp. 1347-1365, 2020.
- [11] C. Kilicoglu, M. Cetin, B. Aricak, and H. Sevik, "Site selection by using the multi-criteria technique—a case study of Bafra, Turkey," *Environmental Monitoring And Assessment*, vol. 192, pp. 1-12, 2020.
- [12] N. Eyles, "Environmental geology of urban areas," in *Environmental geology of urban areas*, Geological Association of Canada, Ontario, *Geotext* 3:1-5, 1997.
- [13] B. R. Marker, "Urban planning: the geoscience input," *Geological Society, London, Engineering Geology Special Publications*, vol. 27, no. 1, pp. 35-43, 2016.
- [14] E. F. de Mulder, "Urban geology in Europe: an overview," *Quaternary International*, vol. 20, pp. 5-11, 1993.

- [15] T. McPhearson et al., "Advancing urban ecology toward a science of cities," *BioScience*, vol. 66, no. 3, pp. 198-212, 2016.
- [16] M. A. Thompson, S. Owen, J. M. Lindsay, G. S. Leonard, and S. J. Cronin, "Scientist and stakeholder perspectives of transdisciplinary research: Early attitudes, expectations, and tensions," *Environmental Science and Policy*, vol. 74, pp. 30-39, 2017.
- [17] A. S. Mather, G. Hill, and M. Nijnik, "Post-productivism and rural land use: cul de sac or challenge for theorization?," in *The Rural*, Routledge, pp. 185-200, 2017.
- [18] D. O. Appiah, D. Schröder, E. K. Forkuo, and J. T. Bugri, "Application of geo-information techniques in land use and land cover change analysis in a peri-urban district of Ghana," *ISPRS International Journal of Geo-Information*, vol. 4, no. 3, pp. 1265-1289, 2015.
- [19] M. Willkomm, A. Follmann, and P. Dannenberg, "Rule-based, hierarchical land use and land cover classification of urban and peri-urban agriculture in data-poor regions with RapidEye satellite imagery: A case study of Nakuru, Kenya," *Journal of Applied Remote Sensing*, vol. 13, no. 1, p. 016517, 2019.
- [20] M. Mathan and M. Krishnaveni, "Monitoring spatio-temporal dynamics of urban and peri-urban land transitions using ensemble of remote sensing spectral indices—A case study of Chennai Metropolitan Area, India," *Environmental Monitoring and Assessment*, vol. 192, no. 1, 2020.
- [21] E. Ustaoglu, S. Sisman, and A. C. Aydinoglu, "Determining agricultural suitable land in peri-urban geography using GIS and Multi-Criteria Decision Analysis (MCDA) techniques," *Ecological Modelling*, vol. 455, p. 109610, 2021.
- [22] H. Bazazzadeh, B. Pourahmadi, S. S. H. Safaei, and U. Berardi, "Urban scale climate change adaptation through smart technologies," *Urban Climate Adaptation and Mitigation*, Elsevier: Amsterdam, The Netherlands, 2022.
- [23] F. Creutzig et al., "Urban infrastructure choices structure climate solutions," *Nature Climate Change*, vol. 6, no. 12, pp. 1054-1056, 2016.
- [24] E. Zengin, "Inundation risk assessment of Eastern Mediterranean Coastal archaeological and historical sites of Türkiye and Greece," *Environ Monit Assess*, vol. 195, p. 968, 2023, doi:10.1007/s10661-023-11549-3
- [25] S. G. Manoharan and G. P. Ganapathy, "GIS-based urban social vulnerability assessment for liquefaction susceptible areas: a case study for greater Chennai, India," *Geoenviron Disasters*, vol. 10, no. 1, 2023, doi:10.1186/s40677-022-00230-5
- [26] M. Lee and D. Basu, "An Integrated Approach for Resilience and Sustainability in Geotechnical Engineering," *Indian Geotechnical Journal*, vol. 48, pp. 207-234, 2018, doi:10.1007/s40098-018-0297-3
- [27] N. Khatri and S. Tyagi, "Influences of natural and anthropogenic factors on surface and groundwater quality in rural and urban areas," *Frontiers in Life Science*, vol. 8, no. 1, pp. 23-39, 2015.
- [28] A. R. Lawrence, D. C. Gooddy, P. Kanatharana, W. Meesilp, and V. Ramnarong, "Groundwater evolution beneath Hat Yai, a rapidly developing city in Thailand," *Hydrogeol J*, vol. 8, pp. 564-575, 2000, doi:10.1007/s100400000098
- [29] C. H. Jeong, "Effect of land use and urbanization on hydrochemistry and contamination of groundwater from Taejon area, Korea," *J Hydrol*, vol. 253, pp. 194-210, 2001, doi:10.1016/S0022-1694(01)00481-4
- [30] M. Zilberbrand, E. Rosenthal, and E. Shachnai, "Impact of urbanization on hydrochemical evolution of groundwater and on unsaturated-zone gas composition in the coastal city of Tel Aviv, Israel," *J Contam Hydrol*, vol. 50, pp. 175-208, 2001, doi:10.1016/S0169-7722(01)00118-8
- [31] J. W. A. Poppen, "Impact of high-strength wastewater infiltration on groundwater quality and drinking water supply: the case of Sana'a, Yemen," *Journal of Hydrology*, vol. 263, no. 1-4, pp. 198-216, 2002.
- [32] D. N. Lerner, "Identifying and quantifying urban recharge: a review," *Hydrogeol J*, vol. 10, pp. 143-152, 2002, doi:10.1007/s10040-001-0177-1
- [33] A. A. Cronin et al., "Temporal variations in the depth-specific hydrochemistry and sewage-related microbiology of an urban sandstone aquifer, Nottingham, United Kingdom," *Hydrogeology Journal*, vol. 11, pp. 205-216, 2003.
- [34] K. L. Powell et al., "Microbial contamination of two urban sandstone aquifers in the UK," *Water Research*, vol. 37, no. 2, pp. 339-352, 2003.
- [35] M. Eiswirth, L. Wolf, and H. Hötzel, "Balancing the contaminant input into urban water resources," *Environmental Geology*, vol. 46, pp. 246-256, 2004.
- [36] E. Vázquez-Suñé, X. Sánchez-Vila, and J. Carrera, "Introductory review of specific factors influencing urban groundwater, an emerging branch of hydrogeology, with reference to Barcelona, Spain," *Hydrogeology Journal*, vol. 13, pp. 522-533, 2005.
- [37] P. A. Ellis and M. O. Rivett, "Assessing the impact of VOC-contaminated groundwater on surface water at the city scale," *Journal of Contaminant Hydrology*, vol. 91, no. 1-2, pp. 107-127, 2007.
- [38] B. Helena et al., "Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis," *Water Research*, vol. 34, no. 3, pp. 807-816, 2000.
- [39] C. H. Jeong, "Effect of land use and urbanization on hydrochemistry and contamination of groundwater from Taejon area, Korea," *Journal of Hydrology*, vol. 253, no. 1-4, pp. 194-210, 2001.
- [40] S. J. McGrane, "Impacts of urbanisation on hydrological and water quality dynamics, and urban water management: a review," *Hydrological Sciences Journal*, vol. 61, no. 13, pp. 2295-2311, 2016.
- [41] U. R. Acar and E. Zengin, "Performance Assessment of Landsat 8 and Sentinel-2 Satellite Images for the Production of Time Series Land Use/Land Cover (Lulc) Maps," *JSR-A*, no. 053, pp. 1-15, 2023, doi:10.59313/jsr-a.1213548
- [42] M. Ahmadi and M. Ghamary Asl, "Monitoring urban growth in Google Earth Engine from 1991 to 2021 and predicting in 2041 using CA-MARKOV and geometry: case study-Tehran," *Arab J Geosci*, vol. 16, p. 107, 2023, doi:10.1007/s12517-022-11089-z
- [43] A. Handayani and A. Wibowo, "Utilizing Google Earth Data to Assess Spatial-Temporal Land Use Changes around Jatibarang Reservoir, Semarang City," *Indonesian Journal of Earth Sciences*, vol. 3, no. 1, pp. A612-A612, 2023.
- [44] J. Carvalho et al., "A look at European practices for identifying mineral resources that deserve to be safeguarded in land-use planning," *Resources Policy*, vol. 74, p. 102248, 2021.
- [45] M. M. Bryant, "Urban landscape conservation and the role of ecological greenways at local and metropolitan scales," *Landscape and Urban Planning*, vol. 76, no. 1-4, pp. 23-44, 2006.