ISSN 2587-1366

TUJE

TURKISH JOURNAL OF ENGINEERING

VOLUME 8 ISSUE 3 JULY 2024

JOURNAL SECRETARY

Aydin Alptekin
aydinalptekin@mersin.edu.tr

TURKISH JOURNAL OF ENGINEERING (TUJE)

Turkish Journal of Engineering (TUJE) is a multi-disciplinary journal. The Turkish Journal of Engineering (TUJE) publishes the articles in English and is being published 4 times (January, April, July and October) a year. The Journal is a multidisciplinary journal and covers all fields of basic science and engineering. It is the main purpose of the Journal that to convey the latest development on the science and technology towards the related scientists and to the readers. The Journal is also involved in both experimental and theoretical studies on the subject area of basic science and engineering. Submission of an article implies that the work described has not been published previously and it is not under consideration for publication elsewhere. The copyright release form must be signed by the corresponding author on behalf of all authors. All the responsibilities for the article belong to the authors. The publications of papers are selected through double peer reviewed to ensure originality, relevance and readability.

AIM AND SCOPE

The Journal publishes both experimental and theoretical studies which are reviewed by at least two scientists and researchers for the subject area of basic science and engineering in the fields listed below:

Aerospace Engineering
Environmental Engineering
Civil Engineering
Geomatic Engineering
Mechanical Engineering
Geology Science and Engineering
Mining Engineering
Chemical Engineering
Metallurgical and Materials Engineering
Electrical and Electronics Engineering
Mathematical Applications in Engineering
Computer Engineering
Food Engineering

PEER REVIEW PROCESS

All submissions will be scanned by iThenticate® to prevent plagiarism. Author(s) of the present study and the article about the ethical responsibilities that fit PUBLICATION ETHICS agree. Each author is responsible for the content of the article. Articles submitted for publication are priorly controlled via iThenticate ® (Professional Plagiarism Prevention) program. If articles that are controlled by iThenticate® program identified as plagiarism or self-plagiarism with more than 25% manuscript will return to the author for appropriate citation and correction. All submitted manuscripts are read by the editorial staff. To save time for authors and peer-reviewers, only those papers that seem most likely to meet our editorial criteria are sent for formal review. Reviewer selection is critical to the publication process, and we base our choice on many factors, including expertise, reputation, specific recommendations and our own previous experience of a reviewer's characteristics. For instance, we avoid using people who are slow, careless or do not provide reasoning for their views, whether harsh or lenient. All submissions will be double blind peer reviewed. All papers are expected to have original content. They should not have been previously published and it should not be under review. Prior to the sending out to referees, editors check that the paper aim and scope of the journal. The journal seeks minimum three independent referees. All submissions are subject to a double blind peer review; if two of referees gives a negative feedback on a paper, the paper is being rejected. If two of referees gives a positive feedback on a paper and one referee negative, the editor can decide whether accept or reject. All submitted papers and referee reports are archived by journal Submissions whether they are published or not are not returned. Authors who want to give up publishing their paper in TUJE after the submission have to apply to the editorial board in written. Authors are responsible from the writing quality of their papers. TUJE journal will not pay any copyright fee to authors. A signed Copyright Assignment Form has to be submitted together with the paper.

PUBLICATION ETHICS

Our publication ethics and publication malpractice statement is mainly based on the Code of Conduct and Best-Practice Guidelines for Journal Editors. Committee on Publication Ethics (COPE). (2011, March 7). Code of Conduct and Best-Practice Guidelines for Journal Editors. Retrieved from http://publicationethics.org/files/Code%20of%20Conduct_2.pdf

PUBLICATION FREQUENCY

The TUJE accepts the articles in English and is being published 4 times (January, April, July and October) a year.

CORRESPONDENCE ADDRESS

Journal Contact: tuje@mersin.edu.tr

CONTENTS

Volume 8 – Issue 3

ARTICLES

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Minimizing satellite residence time in the GEO region through elevated eccentricity method

İbrahim Öz *[1]

[1] *Ankara Yildirim Beyazit University, Technology Transfer Office, Türkiye, ibrahim11oz@gmail.com*

Cite this study: Öz, İ. (2024). Minimizing satellite residence time in the GEO region through elevated eccentricity method. Turkish Journal of Engineering, 8 (3), 416-426

https://doi.org/10.31127/tuje.1395250

Keywords
Geostationary orbit
Space debris
Graveyard orbits
Remove retired satellites
Geostationary protected region

Research Article

Received: 23.11.2023
Revised: 29.02.2024
Accepted: 01.03.2024
Published: 05.07.2024

Abstract

This research focuses on a critical aspect of the space environment, addressing the escalating issue of space debris and congestion in the geostationary orbit. The geostationary orbit is facing many satellites, leading to hazardous congestion levels and jeopardizing the limited resources available. Although organizations have established regulations for retiring satellites to graveyard orbits, a complete removal is not always achievable for numerous reasons. In response to this challenge, our study proposes a practical and cost-effective solution to mitigate debris accumulation in the region. In addition to the above, our research focuses on protecting the geostationary space environment, especially in unforeseen events involving inclined-operated satellites. We explore the implementation of an elevated eccentricity method, increasing the eccentricity of aging satellites and assessing its impact on their time in the geostationary and geostationary-protected regions. Our analysis encompasses short-term, medium-term, and long-term periods, enabling us to evaluate the effectiveness of this approach over different time frames. The study reveals a significant reduction in the time satellites spend in these regions as their eccentricity increases. Moderate eccentricity levels can reduce satellite residence time in these regions from 100.00% to 3.81%. This compelling evidence demonstrates the feasibility and effectiveness of adopting elevated eccentricity as a viable strategy to mitigate space debris in the regions. This proposed approach offers satellite operators a reliable and cost-effective solution, ensuring safe operations and protecting critical regions for aging GEO satellites. Accordingly, we contribute to space environment protection, securing the sustainability of the geostationary orbit.

1. Introduction

Space is expansive and infinite, and now, it is utilized to host satellites that are essential for modern communication, navigation, earth observation, and scientific research. The geostationary region is a critical zone within this domain, accommodating a cluster of satellites in geosynchronous orbits. However, this coveted geostationary orbit is encountering unprecedented challenges. The rapid spread of satellites, coupled with aging spacecraft and debris fragments, has led to an alarming rise in overcrowding, posing a substantial threat to the sustainability and safety of this orbital zone. Protecting the geostationary region ensures continued access to this valuable resource. Space debris, a collective term for defunct satellites, spent rocket stages, and fragments resulting from collisions or breakup, disseminates a hazardous environment. The escalating accumulation of debris amplifies the risk of collisions, creating a cascading effect that generates more fragments and endangers operational satellites. In response to these challenges, the scientific community and space industry have embarked on a quest to safeguard the geostationary region and mitigate the space debris hazard. Various innovative approaches, including advanced tracking and monitoring systems, collision avoidance maneuvers, and novel strategies, have emerged as potential solutions to ease congestion and protect satellites operating within this region.

Geostationary orbit (GEO, also known as geosynchronous orbit) is a region of space approximately 35786 km above the Earth's equator used by many artificial satellites, including communications satellites, weather satellites, and military satellites. However, the geostationary orbit can become congested with debris over time, posing a risk to operational satellites and creating additional challenges for satellite operators.

GEO-orbital debris mitigation reduces the amount of debris in geostationary orbit and minimizes the risk of collisions with operational satellites. Several strategies can be used to mitigate geo-orbital debris, including removing debris from orbit: several organizations, such

as the European Space Agency (ESA) and the Japan Aerospace Exploration Agency (JAXA), are working on technologies and techniques to remove debris from orbit.

The Inter-Agency Space Debris Coordination Committee (IADC) is an international organization that promotes cooperation and coordination among space agencies and organizations to address the issue of space debris. The IADC has developed several recommendations for satellite operators to follow when planning the end of life of satellites to minimize the generation of additional space debris. This includes conducting maneuvers to deorbit the satellite at the end of its mission or placing it into a graveyard orbit where it will not interfere with other objects in space [1]. These recommendations aim to reduce the amount of space debris generated and to ensure the long-term sustainability of geostationary orbit.

The IADC defines the GEO-protected region as the volume of space between 200 km below the GEO altitude (35786 km) to 200 km above the GEO altitude, extending 15° on either side of the equator as shown in Figure 1a and Figure 1b. This region was designed to include a nominal operational altitude excursion of GEO ± 75km and an additional 125 km on either side for a satellite translational corridor [2].

This scientific exploration aims to investigate the complexities of space environment protection, focusing on preserving the integrity and functionality of the geostationary orbit. This study aims to assess the efficacy of the elevated eccentricity method, analyze its impact on debris mitigation, and propose sustainable strategies to ensure the long-term viability of the geostationary region. Through meticulous research and data-driven analysis, this investigation aims to contribute to the collective effort to safeguard the geostationary orbit, fostering a secure and sustainable environment for current and future space missions. Ultimately, the goal is to fortify our capacity to explore, communicate, and innovate while upholding the integrity of the space environment.

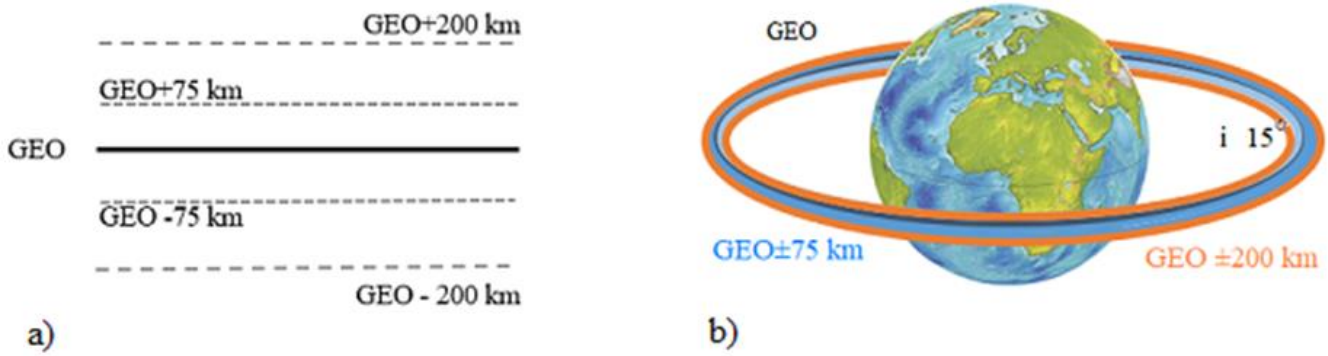


Figure 1. Illustration of geosynchronous orbit and geo-protected region a) two-dimensional b) 3D view (not to scale).

1.1. GEO orbit population

Determining the exact number of satellites in geostationary orbit at any given time is challenging, as the number can fluctuate due to factors such as launches, decommissioning, and others. However, according to data from Celestrak, as of December 2022, there were around 837 satellites in geostationary orbit, out of which 352 were active, as shown in Figure 2a. Figure 2b illustrates the satellite distribution histogram based on inclination and the orbital altitudes, that is, inclination less than 15° and in the geostationary protected region (GEO± 200km).

Figure 2c presents the histogram of 446 inclined satellites in the geostationary-protected region (1°<inclination<15° and altitude GEO±75 km). Active satellites are used for various purposes, including communications, weather forecasting, and military operations.

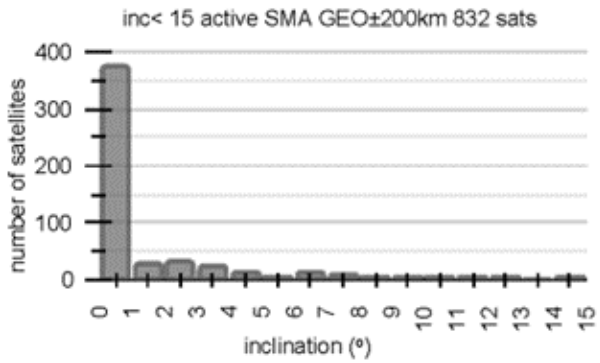
It is found that a significant number of GEO satellites (485 out of 837) have inclinations greater than 1 ° and that this trend has been increasing in recent years, with new classes of GEO networks being deployed with high inclinations and low to moderate eccentricities. This increase in the use of higher inclinations for GEO spacecraft leads to an increase in the number of objects in the GEO region. The number of increased objects in the GEO region highlights the importance of properly retiring satellites to minimize the risk of space debris [3].

1.2. GEO inclined orbit and satellite failure

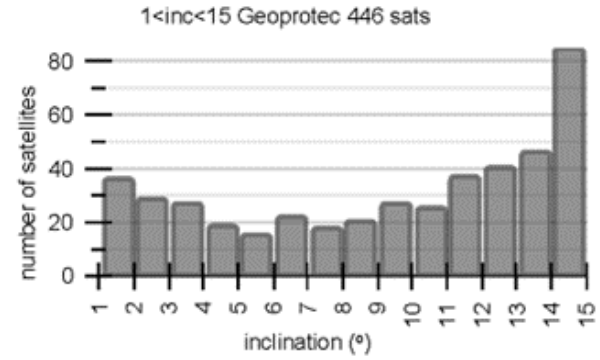
Satellite operators operate the aging satellite, which is close to running out of fuel to extend its maneuver life. The maneuvers consume different amounts of propellant, with the north-south station-keeping consuming more. In order to extend the lifetime of a communication satellite, operators may choose to stop performing the more fuel-intensive north-south station-keeping maneuvers and operate the satellite in an inclined orbit, known as an inclined geosynchronous orbit. All ground stations do not track these orbits but can provide special services to those with tracking capability [4,5]. A satellite's working life refers to the time a satellite can function effectively in orbit before it needs to be replaced or decommissioned. Operators generally prefer satellites with longer working lives, but this cannot be achieved easily due to various factors, such as the complexity of the satellite's subsystems.

Furthermore, it's crucial to note that launcher performance also plays a significant role in determining the maneuver life of a satellite. Extending the design life typically results in heightened mass and higher development costs, demanding additional subsystem

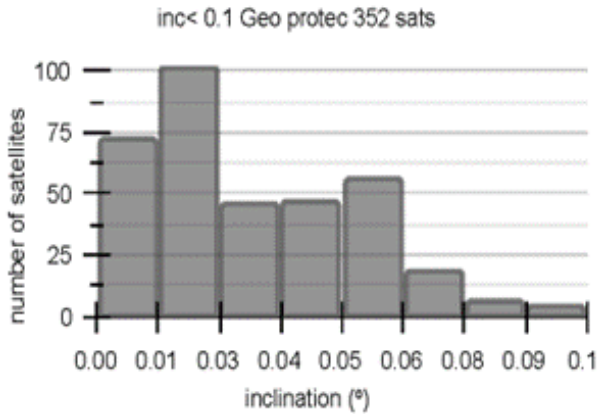
requirements. This delicate balance between design life, cost, and mass remains a pivotal consideration in satellite design and development.



(a)



(b)



(c)

Figure 2. The number of satellites in geo and geo-protected region a) i<15°, b) °1<i<15°, c) i<0.1° (i.e., fully controlled and operational).

The bathtub curve is a common way to represent the failure rate of a satellite's component over time. The curve is shaped like a bathtub, with a high failure rate at the beginning (early life failures), a low failure rate in the middle (random failures), and a high failure rate again at the end (wear-out failures). Early life failures often occur just after the satellite is launched, but these can be eliminated through design and production improvements. The wear-out failures occur at the end of the equipment's life and can be delayed through proper design and production methods. The random failure period is where the failure rate is low and stable and is generally considered the equipment's practical life. These failures are primarily due to factors such as working conditions and external environmental conditions causing equipment stress to exceed a certain level [6].

The space environment can cause various types of damage to the satellite, such as radiation damage, micrometeoroid impact, and thermal stress. In addition, the initial design, manufacturing, and assembly quality can affect the equipment's performance.

Geostationary satellites, like any complex system, can fail for various reasons and may not be able to move them to the graveyard region. Some common causes of geostationary satellite failure include power systems, propulsion systems, structural electronic equipment, and software failures. Geostationary satellites are exposed to a harsh space environment, which can cause damage over time. Solar radiation, cosmic radiation, and temperature extremes can all contribute to satellite failure [7, 8].

The malfunction of a geostationary satellite can result in significant consequences, depending on the purpose and services of the satellite. The malfunctioning satellites can not maneuver and revolve under natural forces. The impact of natural forces in space can alter the satellite's orbit, and the satellite may remain in the geostationary region, the protected geostationary region, or the graveyard region.

2. Geo satellite natural motion

The orbital behavior of a geostationary satellite is influenced by multiple factors, such as the gravitational attraction of the Earth, the solar wind, and the gravitational forces exerted by other celestial bodies.

Table 1 provides the perturbations acting on a geosynchronous satellite that can cause the satellite's orbital elements to vary over time. These variations are categorized into three main groups: short-periodic (SP), long-periodic (LP), and secular (SC) terms. Short-periodic perturbations affect parameters like the semi-major axis, eccentricity, inclination, ascending node, and argument of perigee, attributable to factors such as Earth's non-spherical mass distribution, gravitational forces from the Moon and the Sun, and solar radiation pressure. These changes occur at shorter intervals, spanning days or months. Long-periodic variations also impact the same orbital elements but over longer durations, such as yearly or multi-year spans, due to influences like gravitational forces from the Moon and the Sun, as well as solar radiation pressure. These distinctions delineate the diverse timescales and influences behind alterations in satellite orbits [9, 10].

Secular variations affect the semi-major axis, eccentricity, inclination, ascending node, and argument of perigee. Perturbations cause these variations with no periodic component, resulting in a continuous change in the satellite's orbital elements over an extended period. These variations can be caused by factors such as solar radiation pressure, the moon's and sun's attraction, and the non-uniform mass distribution of the Earth, as shown in Table 1 [11].

The inclination and the rate of eccentricity growth play a role in computing the time spent in the GEO region.

One important factor affecting the long-term orbital behavior of a geostationary satellite is the Earth's gravitational attraction. The satellite's orbit is an ellipse (very close to a circle), with the Earth at one focus. Over time, the Earth's gravitational attraction causes the satellite's orbit to precess or slowly rotate around the Earth. This precession is caused by the fact that the Earth is not a perfect sphere and has a slightly oblate shape.

The solar wind, a stream of charged particles emanating from the Sun, can also affect the long-term orbital behavior of a geostationary satellite. The solar wind can exert a small but measurable force on the satellite, causing its orbit to drift over time.

The gravitational attraction of other celestial bodies, such as the Sun and Moon, can also affect the long-term orbital behavior of a geostationary satellite. The gravitational attraction of these bodies can cause the satellite's orbit to become more or less elliptical over time [12 -16].

Table 1. The short-periodical term (SP), the long-periodical term (LP), and the secular term (SC) effect on a GEO satellite orbital elements.

Orbital elements	Solar radiation Pressure	Non-uniform mass distribution of Earth	Moon and Sun Attraction
Semi major axis	SP+LP	SP	SP+LP
Eccentricity	SP+LP	SP	SP+LP+SC ↑
Inclination (°)	SP+LP+SC ↑	SP	SP+LP+SC ↑
Right ascending node (°)	SP+LP+SC ↓	SP+SC ↓	SP+LP+SC ↓
Argument of perigee (°)	SP+LP	SP+SC ↑	SP+LP+SC ↓

2.1 Satellite dynamic model

Numerical methods for modeling the motion of satellites in orbit are snowballing as computer technology improves. This is because numerical methods have the advantage of being able to incorporate any perturbing force at any point during the simulation. This is achieved by adding the perturbing forces to the two-body equation, also known as Cowell's formulation. As a result, total acceleration of the satellite can be calculated as (Equation 1):

$$a = -\frac{\mu}{r_e} r + a_{perturbed} \tag{1}$$

where: μ: Earth gravitational constant, r_e: Earth Radius $\boldsymbol{r}$: satellite position vector

As the most significant force affecting a satellite in orbit, gravity is not uniform due to the Earth's irregular shape. Therefore, the gravitational force varies based on the satellite's position. A decomposition of spherical harmonics represents the gravitational potential (U) to account for the asymmetric distribution of the Earth, as outlined in the references [17, 18] (Equation 2-3).

where: λ: longitude of satellite, φ: geocentric satellite latitude, R_e: Earth mean radius, $P_{l(,n)}$: Legendre polynomial, $C_{l(,n)}$ and $S(_{l,n)}$ gravitational coefficient from EGM2008 model, L and n: degree and order of decomposition.

$$U = \frac{\mu}{r}\left[1 + \sum_{l=2}^{\infty}\sum_{n=2}^{l}\left(\frac{R_e}{r}\right)^{l} P_{l,n}\, sin\emptyset_{sat}\{C_{l,n} cos\,(n\lambda_{sat}) SC_{l,n}\, sin(n\lambda_{sat})\}\right] \tag{2}$$

$$a_{gravity} = -\nabla U \tag{3}$$

The third body gravitational effect refers to the gravitational attraction exerted on a satellite by celestial bodies such as the Earth and the moon. Third-body gravitational effects can cause perturbations in a satellite's orbit, leading to changes in the satellite's orbital elements. The third body gravitational effect can be modeled using numerical methods and included in the calculation of a satellite's orbit. The GM 2008 and other recent gravity models take into account the third-body gravitational effect and use them to improve the accuracy of satellite orbit prediction. The acceleration caused by these bodies can be represented by Equation 4.

$$a_{3rd\ body} = \mu_{3rd}\left(\frac{r_{sat,3}}{r_{sat,3}^3} - \frac{r}{r^3}\right) \tag{4}$$

where, μ_{3rd}: gravitational parameters of moon and Sun, $r_{sat,3}$: vector from sat to 3rd body.

Solar radiation pressure is a force exerted on a satellite by the Sun's electromagnetic radiation. Solar radiation pressure can affect a satellite's orbit by causing perturbations. The effect of solar radiation pressure on a satellite's orbit can be modeled using numerical methods and included in the calculation of the satellite's orbit. In addition to other forces, the model also considers the effect of solar radiation pressure described in Equation 5.

$$a_{srp} = -\upsilon \frac{C_r A_s}{m} \frac{p_{sr}}{r_{sat,sun}^3} r_{sat,sun} \tag{5}$$

where; υ: the shadow function 0 if sat in shadow; 1 otherwise, C_r solar radiation coefficient between 1 and 2, a_{srp}: cross-section area seen by Sun, P_{sr}: solar radiation pressure [19,20].

3. Elevated eccentricity method

The developed elevated eccentricity method refers to a strategy or technique used to modify the orbit of a satellite by increasing its eccentricity. In the context of

the GEO (geostationary orbit) region, the elevated eccentricity method involves intentionally raising the eccentricity of satellites in order to reduce their time spent in the geostationary and geostationary-protected regions. By increasing eccentricity, satellites spend less time in the regions of interest.

The proposed methodology emphasizes the controlled management of longitude (semi-major axis) and eccentricity during nominal east/west maneuvers. The scheduling of maneuver times is optimized to maximize eccentricity, wherein one maneuver cycle occurs at apogee, followed by the next cycle at perigee, strategically preserving longitude. Moreover, this approach allows for natural eccentricity increases, provided other parameters permit such adjustments.

The equation to calculate the delta-v required to transition a satellite from GEO to a graveyard orbit is represented by Equation 6 [21].

$$\Delta v = 0.0336 * h_o \quad (6)$$

where, h_0 denotes the raised orbit height in km, and ΔV signifies the increased delta-V in m/s.

When considering scenarios involving increases in perigee and apogee while maintaining the same semi-major axis, the simplified approximation remains valid. However, this assumption assumes no additional delta-v is allocated to modify apogee and perigee independently. Instead, the delta-v from east-west maneuvers is utilized for both purposes, including natural eccentricity increases.

In this study, we consider the scenario where a satellite is rendered non-functional due to subsystem failure, fuel depletion, or other reasons, and it lacks the capability to maneuver to the graveyard orbit using its propulsion system.

The disposal of satellites with high eccentricity will still cross the GEO region. However, they may not cross the GEO population ring, which decreases the chance of collision compared to a scenario where the GEO population is evenly spread across the GEO region. The effectiveness of this method can be verified by calculating the amount of time the discarded satellite spends within the GEO-protected region (GEO ± 200 km and inclination 15°).

This study employed six geostationary satellites with specific initial parameters (Sat0 to Sat5) for testing purposes, as shown in Table 2. The satellites were chosen with increasing eccentricity values, with Sat0 having the lowest and Sat5 having the highest. All the satellites were chosen with an inclination of 3°, an area-to-mass ratio of 0.02 m^2/kg, and an epoch date of January 1st, 2021, at 00:00:00 in the J2000 reference frame.

Table 2. Geostationary test satellites with different eccentricities.

Satellite	SMA (km)	Symbol	Eccentricity	Apogee (km)	Perigee (km)	ΔV (m/S)
Sat0	42166.300	e0	5.00×10^{-5}	35790.264	35786.048	0.00
Sat1	42166.300	e1	1.78×10^{-3}	35863.157	35713.155	0.15
Sat2	42166.300	e2	4.74×10^{-3}	35988.157	35588.155	1.83
Sat3	42166.300	e3	8.30×10^{-3}	36138.158	35438.154	5.46
Sat4	42166.300	e4	1.19×10^{-2}	36288.160	35288.152	10.89
Sat5	42166.300	e5	3.00×10^{-2}	37053.165	34523.147	35.82

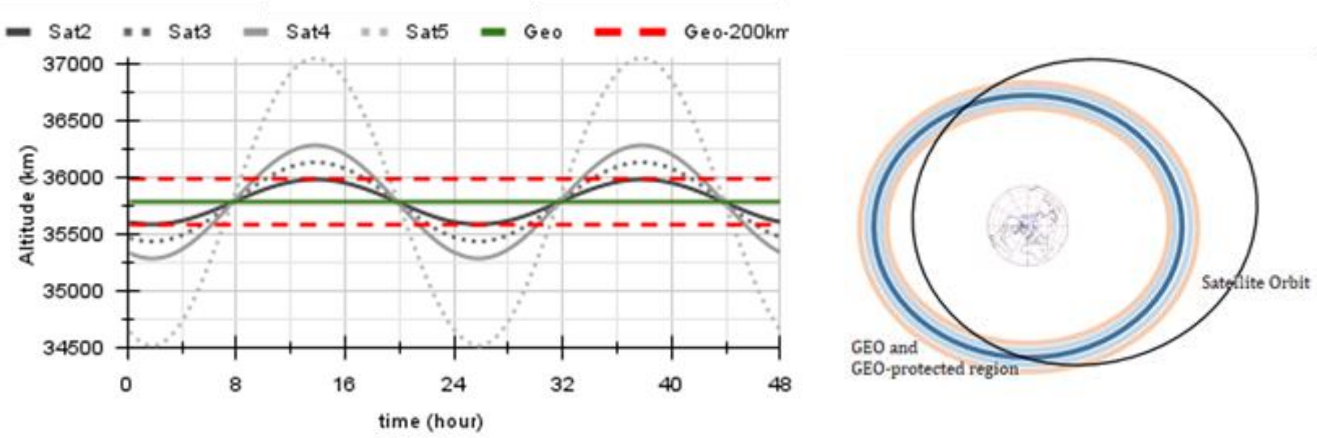
This data provides information about the orbital parameters of six satellites (Sat0-Sat5), including their semi-major axis (SMA) in kilometers, eccentricity symbol, eccentricity, apogee, and perigee altitude in kilometers. The semi-major axis for all the satellites is 42166.3 km, the average distance from the Earth's center to the satellite. The eccentricity symbol (e0-e5) denotes the different levels of eccentricity for each satellite. Eccentricity measures how much the orbit deviates from a perfect circle. The apogee and perigee are the highest and lowest points of the orbit, respectively. The data shows that as the eccentricity increases, the apogee and perigee altitudes increase, and the difference between apogee and perigee altitudes increases. The final column in Table 2 furnishes the requisite delta-v (dV) needed to attain the specified eccentricity value from the initial e0 scenario. Notably, while the table indicates the dV required, executing maneuvers solely to alter eccentricity values is not advised. The expected approach involves augmenting eccentricity during longitude-keeping maneuvers.

Determining a satellite's orbit relies on observational data from ground-based or space-based systems. These systems collect observed and measured datasets, which serve as crucial inputs for orbit estimation. Global Navigation Satellite Systems (GNSS), designed initially to furnish precise positioning and timing information for both military and civilian users [22-25], have emerged as a primary tool for Precise Point Positioning (PPP) and diverse civil applications [25]. Integrating spaceborne GNSS receivers onto geosynchronous communication satellites presents an efficacious alternative for orbit determination compared to conventional methods [26]. Utilizing observation data obtained from GNSS or ground tracking systems, we implement the High-Performance Orbit Propagator (HPOP). This numerical propagation method employs a specific type of numerical integration known as the Runge-Kutta-Fehlberg method of order 7-8. The HPOP serves as a robust technique for satellite orbit determination, facilitating precise and efficient calculations essential for space missions [27].

The theoretical altitude of the satellites over a period of two days is shown in Figure 3a. The figure illustrates the boundary of the geostationary-protected region with a dashed red line and the boundary of the geostationary region with a green line. It can be observed from the figure that the altitude of the satellite's changes with time and with the eccentricity of the satellite. Figure 3b illustrates the intersection of the geostationary region and the actual satellite orbit in the X-Y plane. Minimizing this intersection line will decrease the time the satellite spends in the geostationary region. The satellite's orbit is

analyzed over various time periods, including 1 day, 8 years, and 140 years. The short-term and long-term behavior of the satellite in this orbit is evaluated by monitoring its time spent in the geostationary and geostationary-protected regions.

In this study, the satellites with orbital parameters provided in Table 2 were used to calculate the time spent in the geostationary region. The orbital data was propagated using the high orbit propagating (HPOP) method described in section 2.1. The HPOP method is known for its high accuracy in determining the orbital behavior of a satellite, and thus the time spent in the geostationary region was obtained with high precision.



(a) (b)

Figure 3. a) Sat2 to Sat5 altitude variation for two days. **b)** GEO, GEO-protected, and moderate eccentricity satellite orbits viewed from the north (not to scale).

A satellite operator typically requires approximately 12 m/s of ΔV to reposition a GEO satellite 350 km above the designated GEO region. This maneuver is theoretically executed as two equal maneuvers, separated by 12 hours, when the satellite is at its apogee. In practical applications, this operation is often divided into 4 or even 8 separate maneuvers as per industry practices.

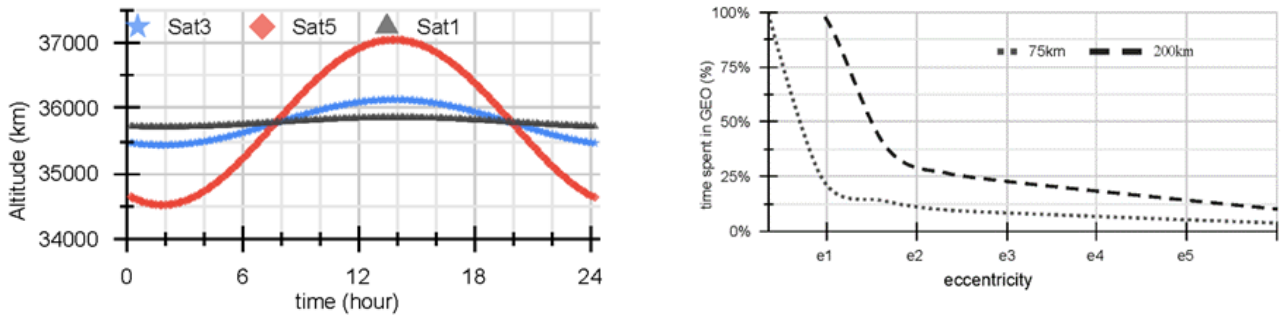
In scenarios where no failures occur, these repositioning operations are conducted in accordance with established procedures. However, when a failure occurs and the satellite cannot be appropriately repositioned, a satellite with zero eccentricity would spend a significant portion of its time within the GEO region. In contrast, in an elevated eccentricity scenario, should the same issue arise, the residence time of the aging satellite in the GEO region would be significantly reduced.

When a geostationary satellite malfunctions for any reason, it is subject to natural forces that cause it to change in its orbit. Analyzing the satellite's orbital parameters, such as eccentricity, longitude, and semi-major axis in natural cases, makes it possible to estimate how long the satellite will remain in the geostationary or geostationary-protected region. Minimizing the time spent in the geostationary region can help reduce the accumulation of debris in that area.

4. Results and Discussion

This paper introduces a novel strategy to safeguard both the geostationary and geostationary-protected regions by implementing elevated eccentric operations for aging satellites in case of orbit raising failure.

The altitude of satellites varies over time due to their orbital parameters. Eccentricity plays a significant role in the variation of altitude. Figure 4a illustrates the distance of three selected sample satellites, Sat1, Sat3, and Sat5, from the GEO zero orbit (35786 km). Figure 4b displays the time spent in the GEO region for Sat0 to Sat5. The graph shows that satellites with high eccentricity have a greater distance to the GEO zero orbit and spend less time in the GEO region.



(a) (b)

Figure 4. a) Sat1, Sat3, and Sat5 orbit distance to GEO for 24 hours **b)** satellite spent time in GEO and GEO protected region for eccentricity e0 to e5.

Table 3 provides information about the time spent in the regions of interest (from 75 km to 500 km) for each satellite without orbit perturbation. Sat4 and Sat5 spend 73.81% and 89.89% of one revolution's time in the graveyard region, while Sat1, Sat2, and Sat3 do not reach the graveyard region. As the eccentricity of a satellite increases, the time spent in the geostationary region decreases, which helps to mitigate the amount of time spent in the geostationary and geostationary protected regions.

With this method, by increasing the eccentricity to a value of e=0.03, as shown in Table 3, it is estimated that a satellite would spend only 3.77% of its orbital revolution in the geostationary region, 6.33% in the geostationary protected region and the majority in the super-geostationary (graveyard) region.

Table 3. Sat0 to Sat 5 time spent in GEO, GEO protected and graveyard region.

Sat	0-75km	75-200km	200-350km	350-500km	500+ km	ecc	ecc
Sat0	100.00%	0.00%	0.00%	0.00%	0.00%	5.00×10^{-5}	e0
Sat1	99.76%	0.24%	0.00%	0.00%	0.00%	1.78×10^{-3}	e1
Sat2	24.47%	75.29%	0.24%	0.00%	0.00%	4.74×10^{-3}	e2
Sat3	13.75%	24.97%	61.04%	0.24%	0.00%	8.30×10^{-3}	e3
Sat4	9.58%	16.61%	23.17%	50.40%	0.24%	1.19×10^{-2}	e4
Sat5	3.77%	6.33%	7.74%	8.02%	74.13%	3.00×10^{-2}	e5

Figure 5 provides the results of Sat0 to Sat5's time spent in the GEO, GEO protected, and graveyard regions and shows the time spent in each region. An increase in eccentricity results in the satellite spending more time at higher altitudes, which reduces the time spent in the GEO region.

Figure 5, x-axis provides information about the time spent by the satellites in different altitude regions, and the y-axis shows the eccentricity values of Sat0 to Sat5.

The impact of natural forces on the orbital behavior of satellites over the medium and long term was analyzed in terms of the time spent in the GEO and GEO protected region. Figure 6a and 6b depict the effect of operating an aging satellite with moderate eccentricity on the changes in eccentricity and the altitude of apogee and perigee over an 8 and 15-year period in the GEO region.

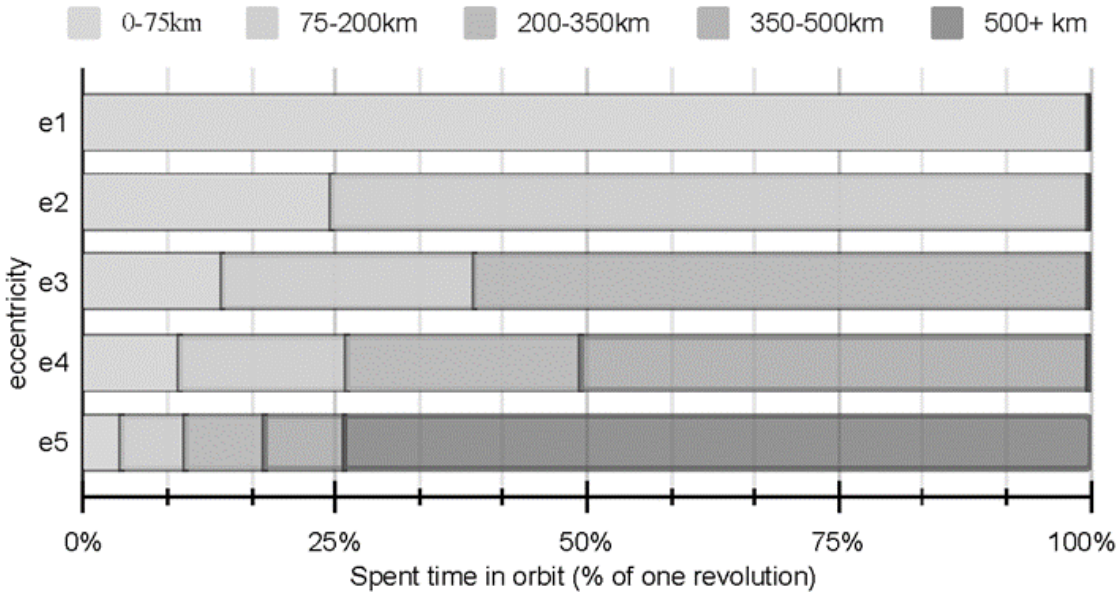


Figure 5. Percentage of time spent in a different region for Sat0 to Sat5.

Figure 6a shows the fluctuation of eccentricity between 1.1×10^{-5} and 8.05×10^{-4}. Figure 6b) displays the long-term changes in the apogee and perigee altitude, ranging between 35947.088 km and 35833.585 km for apogee and between 35737.402 km, and 35790 km for perigee. In Figure 6b, the x-axis represents the time in years, and the y-axis displays the varying orbital parameters of the satellite.

Figure 6c, the changes in longitude over an 8-year period, with the maximum value reaching 122.565° and the minimum value being 24.789°. Figure 6d demonstrates the right ascension of the ascending node (RAAN), which has a period of around 52 years, and RAAN varies between 0 and 360 °s for 140 years.

Figure 7a and 7b demonstrate the impact of operating the aging satellite Sat1, with a moderate eccentricity, on the variation of the satellite's eccentricity and altitude of its apogee and perigee over 140 years in the geostationary region. Figure 7a shows the eccentricity variation, ranging from 7.1×10^{-5} to 1.465×10^{-3}. Figure 7b illustrates the long-term changes in the altitudes of the apogee and perigee, with a range of variation in apogee altitude from 35873.843 km to 35768.695 km and perigee altitude from 35807.058 km to 35707.759 km.

The x-axis represents time in years, the left y-axis in blue shows the altitude of apogee variation, and the right y-axis in red shows the altitude of perigee variation in km.

The semi-major axis, the argument of perigee, the right ascension of the ascending node, and the inclination orbital elements do not significantly affect the amount of time spent in the GEO region. These elements have periods of 2.57, 1.07, and 52.5 years, respectively. Eccentricity, however, significantly impacts the time spent in GEO, as an increase in eccentricity results in a decrease in the amount of time spent in the region for all satellites.

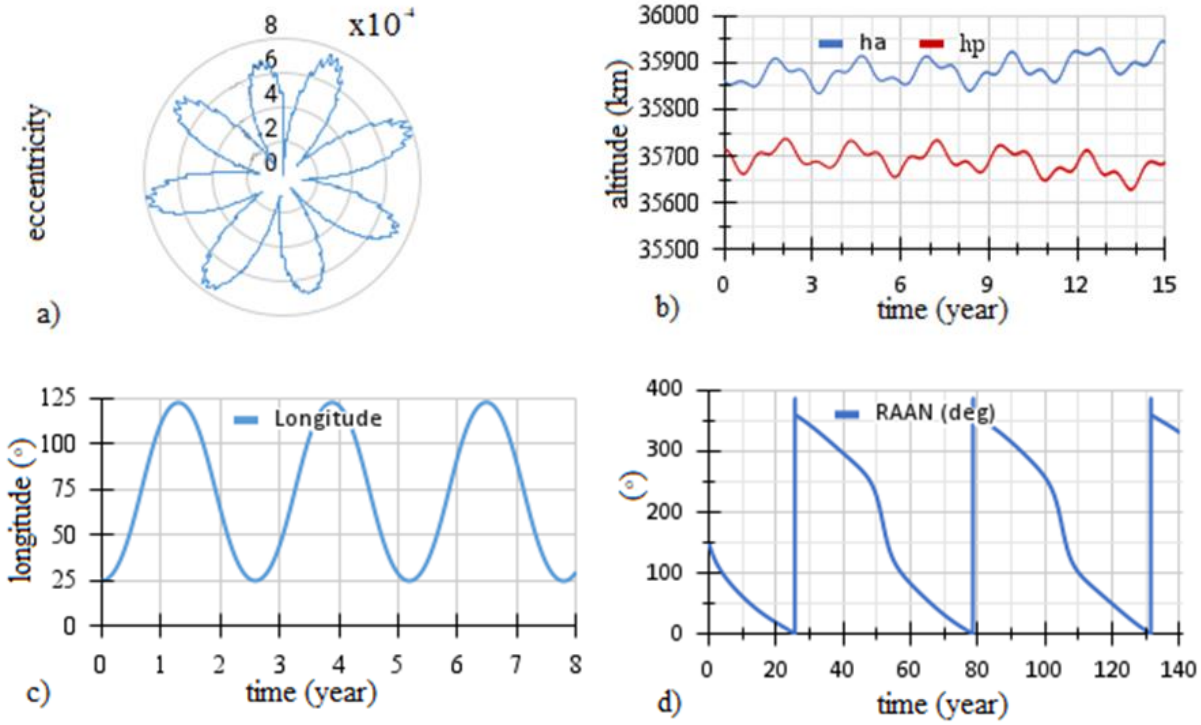


Figure 6. a) Eccentricity variation of Sa0 for 8 years, **b)** Altitude of apogee and perigee variation over 15 years, **c)** Longitude change for 8 years, **d)** Right ascension of ascending node over 140 years.

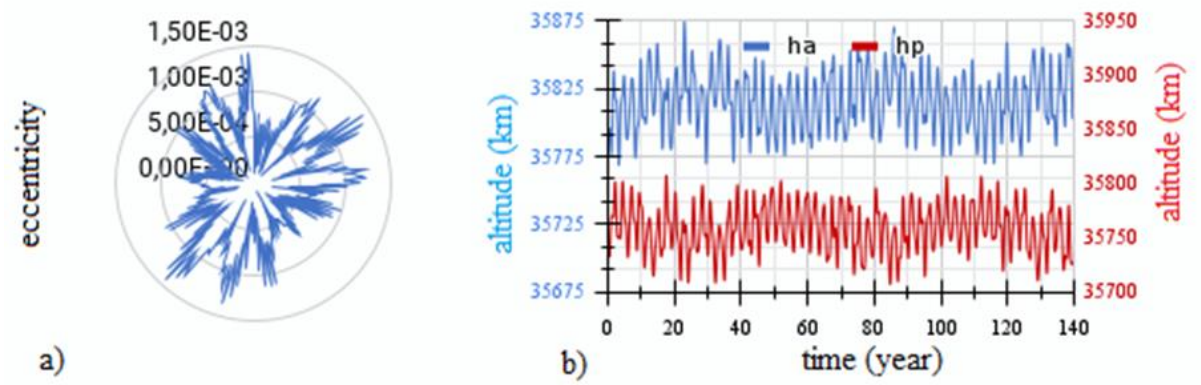


Figure 7. a) Eccentricity variation and **b)** Altitude of apogee and perigee variation over 140 years for Sat1.

Sat0 and Sat5 have been selected as representative samples of all satellites to give a comprehensive overview of the variations in orbital elements. Sat0 has a low eccentricity, while Sat5 has a high eccentricity. Table 4 displays the changes in the orbital parameters of Sat0 and Sat5 over time, with each row representing a different orbital element. The columns depict the statistical values for each orbital parameter. The table highlights the effect of natural forces on the orbital parameters over time, such as Earth's gravity, solar radiation pressure, and other third-body gravitational effects.

Table 4 compares the orbital elements of two satellites, Sat0 and Sat5. The semi-major axis (SMA) of both satellites is very similar. The inclination of Sat0 has a range of 15.0° to 0.0°, with a standard deviation of 5° and a difference of 15.0°. Sat5 has similar values, ranging from 15.0° to 0.0°, with a standard deviation of 5° and a difference of 15.0°. The right ascension of the ascending node and the argument of perigee of both satellites are similar. The longitude of both satellites has changed slightly, but the difference is not significant.

The eccentricity of Sat0 is 5×10^{-5}, which is constant, while that of Sat5 is 0.03° with a standard deviation of 0 and a delta of 0.01°. This indicates that Sat5 has a small elliptical orbit. Regarding altitude, the apogee for Sat0 has a maximum value of 35878.89 km, a minimum of 35761.64 km, a standard deviation of 22.48 km, and a

delta of 117.25 km. The same parameters for Sat5 are 37245.98 km, 36983.65 km, 48.82 km, and 262.34 km. The perigee altitude for Sat0 has a maximum value of 35814.28 km, a minimum of 35701.77 km, a standard deviation of 22.64 km, and a delta of 112.51 km. The same parameters for Sat5 are 34585.16 km, 34328.68 km, 48.43 km, and 256.48 km. Finally, the difference in apogee and perigee altitude for Sat0 is 129.63 km with a standard deviation of 26.81 km and a delta of 128.02 km. For Sat5, these values are 2870.61 km, 2444.98 km, 90.25 km, and 425.63 km, respectively.

Table 4. Orbital elements evolution under the natural forces (perturbed orbit) over 140 years for Sat0 and Sat5.

Orbital Elements	Sat0 max	Sat0 min	Sat0-stdev	Sat0 delta - max-min	Sat5 max	Sat5 min	Sat5 stdev	Sat5 delta - max-min
SMA (km)	42193.35	42138.04	18.15	55.31	42193.30	42138.10	18.11	55.20
ecc	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.01
Inc (°)	15.00	0.00	5.00	15.00	15.00	0.00	5.00	15.00
RAAN (°)	359.99	0.03	128.91	359.96	359.97	0.01	129.00	359.96
Aop (°)	359.91	0.03	102.52	359.88	359.97	0.03	101.60	359.94
Ha (km)	35878.89	35761.64	22.48	117.25	37245.98	36983.65	48.82	262.34
Hp (km)	35814.28	35701.77	22.64	112.51	34585.16	34328.68	48.43	256.48
deltaAP (km)	129.63	1.60	26.81	128.02	2.870.61	2.444.98	90.25	425.63
Longitude (°)	123.39	24.53	35.43	98.86	123.40	24.43	35.42	98.97

The study also found that changes in the semi-major axis and eccentricity can affect a satellite's apogee and perigee, influencing the time spent in the geostationary region. By analyzing the orbital behavior of the satellite under natural forces, the time spent in the geostationary region was calculated, providing insights into strategies for mitigating space debris. Overall, the comparison shows that Sat0 and Sat5 have similar orbital elements but with some differences, especially in terms of eccentricity and altitude. The differences in eccentricity change the spent time in GEO and GEO-protected regions.

Table 5 presents information about six satellites, Sat0 to Sat5, with varying eccentricity values. Sat0 has an eccentricity of 5×10^{-5}, while Sat1, Sat2, Sat3, Sat4, and Sat5 have eccentricity values of 1.77×10^{-3}, 4.74×10^{-3}, 1.18×10^{-2}, and 3×10^{-2}, respectively. Sat1 and Sat2 spend all of their time in the geostationary region for one revolution, while Sat2, Sat3, Sat4, and Sat5 spend 25.0%, 13.85%, 9.72%, and 3.47% of 1 revolution's time in the geostationary region. Similarly, Sat0, Sat1, and Sat2 spend all of their time in the geostationary protected region, but Sat3, Sat4, and Sat5 spend 39.58%, 26.39%, and 9.73% of one revolution's time in the geostationary protected region.

Table 5 also displays the medium and long-term fluctuations in the amount of time satellites spend in GEO and its protected region. The data reveals that the variations are around 0.4% to 0.5%, which are relatively small and suggest that medium and long-term changes in orbital parameters have a limited effect on the time spent in the GEO region.

Table 5. Geo satellites spent time around the GEO region for short, medium, and long term.

Time	Region Boundary	Sat0 (%)	Sat1 (%)	Sat2 (%)	Sat3 (%)	Sat4 (%)	Sat5 (%)
8 years	GEO±75 km	100.0	75.79	24.20	13.70	9.61	3.81
	GEO±200 km	100.00	100.00	85.97	38.63	26.24	10.18
1 year	GEO±75 km	100.00	75.68	24.16	13.64	9.61	3.82
	GEO±200 km	100.00	100.00	85.42	38.54	26.15	10.14
1 day	GEO±75 km	100.00	100.00	25.00	13.89	9.72	3.47
	GEO±200 km	100.00	100.00	100.00	39.58	26.39	9.72

Figure 8 shows Sat0 to Sat5 spent time in GEO and GEO-protected regions to analyze the effect of short, medium, and long duration in the GEO region. Figure 8 shows the percentage of time that given satellites, identified by the "e" values (e0, e1, e2, etc.) time spent in different regions over a specified time period. The first column, "8Y:G" indicates 8 years in the "GEO" region, while "8Y:GP" indicates 8 years in the "GEO protected" region. The other columns, "1Y:G", "1Y:GP", "1DG", "1D:GP" indicate 1 year and 1 day in the GEO and GEO-protected regions, respectively. The percentage values in each cell of the table indicate the percentage of time that the respective satellite spent in the corresponding region over the specified time period. The trend line shown in black describes the path that the data points in Figure 8 converge to zero as e increases. It implies that high eccentricity decreases time spent in the GEO region.

With this method, by increasing the eccentricity to a value of e=0.03, it is obtained that a satellite would spend less time in the geostationary and the geostationary protected region; it can be shown clearly from the graph that as eccentricity increases, the time spent in GEO and GEO protected regions decreases. Therefore, operating aging satellites with moderate eccentricity can be an effective method to empty the GEO region. By increasing the eccentricity of a satellite, it is possible to reduce the amount of time it spends in the geostationary region and thus reduce the risk of collisions and the amount of debris in that region. Additionally, by analyzing the orbital behavior of the satellite, it is possible to ensure

the long-term sustainability of the GEO region and mitigate space debris.

The utilization of the elevated eccentricity method by an operator is suggested to potentially reduce the duration in the GEO and GEO-protected regions in the event of an orbit-raising failure. However, a successful orbit-raising operation may render the use of the elevated eccentricity method unnecessary.

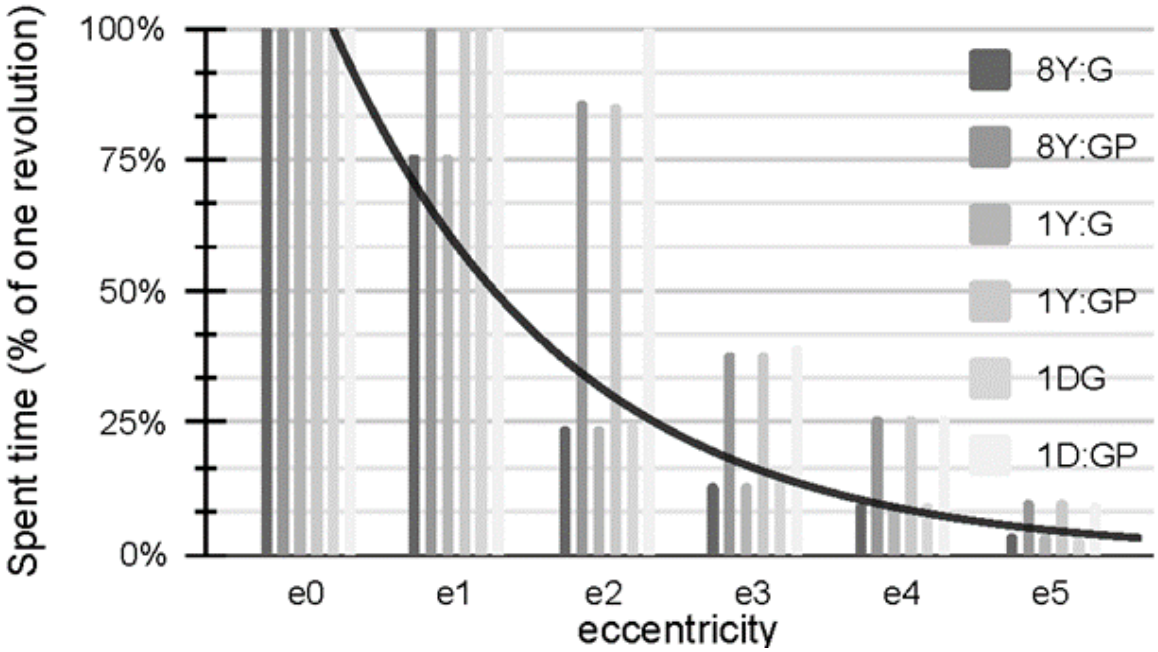


Figure 8. Short, medium and long-term satellites that is under the natural forces (perturbed orbit) spent-time in GEO and GEO protected region evolution and eccentricity trend line.

Many efforts can be made to decrease the time the satellite spends, thereby carefully controlling its initial orbital parameters, such as eccentricity, to mitigate the buildup of space debris. Moreover, employing innovative methods as a backup solution for operating aging satellites can significantly reduce the risk of space debris accumulation. This strategic approach is critical in the sustainability of space operations for all satellite operators.

5. Conclusion

This study shows that operating aging satellites with high eccentricity can effectively mitigate the risk of space debris. As eccentricity increases, the time a satellite spends in the geostationary region decreases, reducing the risk of collisions and the amount of debris in that region. The study found that the time the satellite spent in the geostationary region stayed consistently low over 140 years, confirming this method's effectiveness. This method may not be needed if the satellite's orbit-rising operation is successful. It is advantageous in the case of partially successful orbit raising.

In conclusion, operating aging satellites with elevated eccentricity effectively mitigate space debris by reducing time spent in the geostationary region. This method is relatively simple and inexpensive. While the elevated eccentricity method effectively protects GEO from debris, it should be used with other long-term sustainability techniques.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Delong, N., & Frémeaux, C. (2005). Eccentricity management for geostationary satellites during end of life operations. In 4th European Conference on Space Debris, 587, 297. Darmstadt: ESA Special Publication.
2. Xu, W., Liang, B., Li, B., & Xu, Y. (2011). A universal on-orbit servicing system used in the geostationary orbit. Advances in Space Research, 48(1), 95-119. https://doi.org/10.1016/j.asr.2011.02.012
3. Johnson, N. L. (2012). A new look at the GEO and near-GEO regimes: Operations, disposals, and debris. Acta Astronautica, 80, 82-88. https://doi.org/10.1016/j.actaastro.2012.05.024
4. Öz, İ., & Yılmaz, Ü. C. (2020). Determination of coverage oscillation for inclined communication satellite. Sakarya University Journal of Science, 24(5), 973-983. https://doi.org/10.16984/saufenbilder.702190
5. Oz, I. (2022). Salınımlı yörünge haberleşme uydularında 2 eksen düzeltmeli kapsama alanı stabilizasyonu. Journal of The Faculty of Engineering and Architecture of Gazi University, 38(1), 219-229. https://doi.org/10.17341/gazimmfd.960480
6. Fu, S. Y., Wang, Z. R., Shi, H. L., & Ma, L. H. (2018, June). The application of decommissioned GEO satellites to CAPS. In IOP Conference Series: Materials Science and Engineering, 372(1), 012033. https://doi.org/10.1088/1757-899X/372/1/012033
7. Tafazoli, M. (2009). A study of on-orbit spacecraft failures. Acta Astronautica, 64(2-3), 195-205. https://doi.org/10.1016/j.actaastro.2008.07.019

8. Cougnet, C., Gerber, B., Heemskerk, C., Kapellos, K., & Visentin, G. (2006). On-orbit servicing system of a GEO satellite fleet. In 9th ESA Workshop on Advanced Space Technologies for Robotics and Automation 'ASTRA.
9. Dong, X., Hu, C., Long, T., & Li, Y. (2016). Numerical analysis of orbital perturbation effects on inclined geosynchronous SAR. Sensors, 16(9), 1420. https://doi.org/10.3390/s16091420
10. Anselmo, L. (2004). The long-term evolution of the Italian satellites in the GEO region and their possible interaction with the orbital debris environment. In 54th International Astronautical Congress of the International Astronautical Federation, the International Academy of Astronautics, and the International Institute of Space Law, IAC-03-IAA.5.2.05. https://doi.org/10.2514/6.IAC-03-IAA.5.2.05
11. Rosengren, A. J., Scheeres, D. J., & McMahon, J. W. (2013). Long-term dynamics and stability of GEO orbits: the primacy of the Laplace plane. In Proceedings of the AAS/AIAA Astrodynamics Specialist Conference, Hilton Head, South Carolina, AAS 13-865.
12. Jenkin, A. B., McVey, J. P., & Sorge, M. E. (2022). Assessment of time spent in the LEO, GEO, and semi-synchronous zones by spacecraft on long-term reentering disposal orbits. Acta Astronautica, 193, 579-594. https://doi.org/10.1016/j.actaastro.2021.07.048
13. Mei, H., Damaren, C. J., & Zhan, X. (2021). End-of-life geostationary satellite removal using realistic flat solar sails. Aerospace Systems, 4, 227-238. https://doi.org/10.1007/s42401-021-00089-8
14. Cabrières, B., Alby, F., & Cazaux, C. (2012). Satellite end of life constraints: Technical and organizational solutions. Acta Astronautica, 73, 212-220. https://doi.org/10.1016/j.actaastro.2011.10.014
15. Yilmaz, N. (2023). Assessment of latest global gravity field models by GNSS/Levelling Geoid. International Journal of Engineering and Geosciences, 8(2), 111-118. https://doi.org/10.26833/ijeg.1070042
16. Yilmaz, M., Turgut, B., Gullu, M., & Yilmaz, I. (2016). Evaluation of recent global geopotential models by GNSS/Levelling data: internal Aegean region. International Journal of Engineering and Geosciences, 1(1), 18-23. https://doi.org/10.26833/ijeg.285221
17. Vallado, D. A. (2001). Fundamentals of astrodynamics and applications, 12. Springer Science & Business Media.
18. Montenbruck, O., Gill, E., & Lutze, F. (2002). Satellite orbits: models, methods, and applications. Applied Mechanics Reviews, 55(2), B27-B28. https://doi.org/10.1115/1.1451162
19. Öz, İ. (2024). Eş konumlu uyduların yakınlaşma izlenmesine gerçek zamanlı mesafe ölçümü tabanlı yaklaşım. Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 39(2), 825-834. https://doi.org/10.17341/gazimmfd.1181262
20. Atiz, Ö. F., Konukseven, C., Öğütcü, S., & Alcay, S. (2022). Comparative analysis of the performance of Multi-GNSS RTK: A case study in Turkey. International Journal of Engineering and Geosciences, 7(1), 67-80. https://doi.org/10.26833/ijeg.878236
21. Anselmo, L., & Pardini, C. (2017). On the end-of-life disposal of spacecraft and orbital stages operating in inclined geosynchronous orbits. In Proceedings of the 9th IAASS Conference, Session 05: Space Debris-I, 87-94.
22. Uçarli, A. C., Demir, F., Erol, S., & Alkan, R. M. (2021). Farklı GNSS uydu sistemlerinin hassas nokta konumlama (PPP) tekniğinin performansına etkisinin incelenmesi. Geomatik, 6(3), 247-258. https://doi.org/10.29128/geomatik.779420
23. Pirti, A., Gündoğan, Z. Ö., & Şimşek, M. (2022). QZSS uyduları ve sinyal yapıları. Geomatik, 7(3), 243-252. https://doi.org/10.29128/geomatik.979823
24. Pirti, A., Hoşbaş, R. G., Şenel, B., Köroğlu, M., & Bilim, S. (2021). Galileo uydu sistemi ve sinyal yapısı. Geomatik, 6(3), 207-216. https://doi.org/10.29128/geomatik.750469
25. Altuntaş, C., & Tunalıoğlu, N. (2022). Retrieving the SNR metrics with different antenna configurations for GNSS-IR. Turkish Journal of Engineering, 6(1), 87-94. https://doi.org/10.31127/tuje.870620
26. Koca, B., & Ceylan, A. (2018). Uydu konum belirleme sistemlerindeki (GNSS) güncel durum ve son gelişmeler. Geomatik, 3(1), 63-73. https://doi.org/10.29128/geomatik.348331
27. Refaat, A., Badawy, A., Ashry, M., & Omar, A. (2018). High accuracy spacecraft orbit propagator validation. In The International Conference on Applied Mechanics and Mechanical Engineering, 18th International Conference on Applied Mechanics and Mechanical Engineering, 1-9. Military Technical College.

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

A critical study on the treatability of metal plating industry wastewater and real scale adaptation

Inci Karakas [*1], **Soner Kizil** [2]

[1] Uskudar University, Vocational School of Health Services, Department of Environmental Health, Türkiye, inci.karakas@uskudar.edu.tr
[2] Uskudar University, Faculty of Engineering and Natural Sciences, Türkiye, soner.kizil@uskudar.edu.tr

Cite this study: Karakas, İ., & Kizil, S. (2024). A critical study on the treatability of metal plating industry wastewater and real scale adaptation. Turkish Journal of Engineering, 8 (3), 427-435

https://doi.org/10.31127/tuje.1406437

Keywords
Real wastewater
Metal plating
Coagulation-flocculation
Operating cost
Sludge quantity

Research Article

Received: 18.12.2023
Revised: 17.01.2024
Accepted: 18.01.2024
Published: 05.07.2024

Abstract

Water pollution is one of the major problems for humankind. Various pollutants could be detected in wastewater because of human activities such as industrialization, agriculture, domestic waste and etc. Removal of pollutants such as heavy metals, dyes, oils and pesticides are of great importance which affects human life negatively. Many methods have been extensively used to provide "clean water" for environment and human. Heavy metals are important industrial pollutants that need to be quickly removed from wastewater due to their high toxicity and non-biodegradable structure. In this paper, the heavy metals including copper, nickel and zinc have been examined in real wastewater from metal plating industry in Bursa, Türkiye. Concentrations of pollutants (Cu, Ni, Zn, chemical oxygen demand (COD) and SO_4) in the effluent as a result of coagulation-flocculation process were determined and their compliance with sewage discharge standards was investigated. Moreover, the removal efficiencies of the pollutants were examined (Cu and Ni: 97-100%; Zn: 82-98%; COD: 32-54%; SO_4: 16-23%) and the effect of the coagulant doses used on the operating cost, sludge quantity were also discussed.

1. Introduction

Water is crucial for human life and ecosystem and only 1% of the available water can be used for drinking and agricultural activities. Various activities such as natural events, wars, accidents, industrial activities, farming and agriculture have led to the discharge of many pollutions which represents a public concern [1]. Wastewater has become the most important environmental problem for human and living organisms due to its hazardous effects on our health and habitat [2,3]. Water can be polluted by dye, heavy metals, oil, pesticides, poly- and perfluoroalkyl substances, personal care products, and so on [4,5].

Global water resources are severely polluted by heavy metal ions and organic compounds from industrial activities [6,7]. Heavy metal-based water pollution is one of the most dangerous pollutions for human life because metals tend to accumulate in the body, not biodegradable and heavy metals with a density exceeding 5 g/m^3 are known to be toxic and carcinogenic [8]. Zinc (Zn), Copper (Cu), Nickel (Ni), Mercury (Hg), Cadmium (Cd), Lead (Pb) and Chromium (Cr) are heavy metals found in industrial wastewater and require special attention in the cleaning of industrial wastewater [9,10]. Heavy metals in wastewater from various industrial activities such as metal plating-smelting, electrolysis and electroplating could pollute soil, water, and air [11]. In metal plating industries, platings are often used to protect metals against corrosion and to improve basic properties such as mechanical or chemical stability which is used in platings beautify objects such as car and aircraft parts, bathroom fixtures, chairs, and tables [12,13]. The metal plating industry produces large volumes of strongly acidic wastewater containing high concentrations of organic and inorganic pollutants, cyanide, heavy metals (chromium, nickel, zinc, and copper), degreasing solvents and particulate matter [14–17]. In metal plating industries, objects are plated in plating baths and their surfaces are cleaned in rinsing baths. Heavy metals in the plating solution migrate to the rinsing baths and the water in the rinsing baths is constantly changed. Rinse waters contain toxic heavy metals (Zn, Ni, Cu, Hg, Cd, Pb, Cr etc.), cyanide and various carcinogenic substances that cause environmental pollution. toxic compounds

may be formed by the contact of cyanide with air and metals [18,19].

Zinc is an element necessary for human health, but excessive amounts cause health problems such as abdominal pain, skin problems and nausea [20]. Lead damages the central nervous system which is the second most toxic metal after arsenic [21]. Mercury is neurotoxic and damages the central nervous system. High concentration of mercury could cause deterioration in kidney function, chest pain and shortness of breath. It is well known that removing such pollutants from wastewater is of great importance for human health and environment that we live in [22,23]. Untreated discharge of rinse water into sewage systems causes serious problems due to high toxicity and can negatively affect the microbial activity of activated sludge in the wastewater treatment plant [24]. The content of the wastewater varies depending on the process applied, the type of plating and the amount. Wastewater from metal plating plants can be both alkaline and acidic. Cyanide-containing wastewater or water used in rinsing baths has a very high pH, while wastewater from chromium, copper and nickel baths is acidic [16,25].

A variety of techniques have been devoted in developing quick, effective, and cheap way to remove heavy metals from heavy metals. These methods include coagulation, flocculation adsorption, membrane filtration, reverse osmosis, ion exchange, and precipitation [26,27]. Even all methods have been utilized for water treatment, each has its own limitations such as time consuming, require complicated procedure, low efficiency, high cost etc. Among them, Coagulation-Flocculation process is the most effective, easy-to-implement and economical method for the treatment of heavy metals and inorganic compounds [28,29]. Since the pH of metal plating industry wastewater is highly acidic, it is necessary to bring the pH to the neutral range before discharge. Furthermore, the increase in wastewater pH requires high costs for the use of chemical reagents and treatment of the metal oxide precipitates [30,31]. The basis of coagulation-flocculation processes is to adjust the pH in wastewater. Chemical coagulants such as aluminum or iron salts are added to the wastewater to remove pollutants in colloidal form [32,33]. Polymerized forms of metal coagulants such as aluminum sulfate (alum), ferric chloride, and PAC (poly-aluminum chloride) are compounds that provide high pollutant and organic matter removal and can promote lower volume sludge formation [34]. Inam et al. [35] have studied the removal of As (III) and As (V) from aqueous solution by coagulation-flocculation process using $FeCl_3$ as coagulant, organic ligands (humic acid and salicylic acid) and the removal efficiencies were determined as 99% and 92%, respectively. Ag removal from industrial wastewater using coagulation-flocculation process was investigated. 46.36 mg/L PAC was used as coagulant and 0.318 mg/L anionic polyelectrolyte (Praestol 2640) was used as flocculant and Ag removal efficiency was 99% [36]. Cd(II) removal from aqueous solution by coagulation-flocculation process was investigated and CaO/PAC was used as coagulant, anionic polyacrylamide was used as flocculant. Cd(II) removal efficiency was founded as 94% [37]. Cr and Ni removal were investigated using flocculation process in real chromic acid lotion electroplating wastewater. Functionalized carboxylated chitosan flocculants with xanthate and sulfonic acid groups have been used. Cr and Ni removal efficiencies were 95% and 99%, respectively [38]. Cu(II) removal was investigated in flocculation process from aqueous solution and xanthated chitosan was used as flocculant. In the study, Cu(II) removal efficiency was founded as 97% [39].

In this study, COD and (Cu, Ni, Zn) heavy metals were investigated with the coagulation-flocculation process in the effluent concentrations and removal yield and it was aimed to supply the discharge standard to the sewage in real wastewater from the metal plating industry in Bursa. High concentrations of SO_4 are formed due to the use of H_2SO_4 in rinsing baths in the metal plating industry. Besides COD and Heavy Metal (Cu, Ni, Zn) removal, SO_4 removal has also been investigated. Coagulants (Caustic (NaOH), lime ($Ca(OH)_2$), iron III chloride ($FeCl_3$), barium chloride ($BaCl_2$) and anionic polyelectrolyte ($(NH_4)_2S_2O_8$)) operating costs used were calculated. The outcomes have demonstrated that the wastewater is integrated into the wastewater treatment plant and the system is operated.

2. Method

2.1. Sample collection

The wastewater used in the experiments were obtained from the metal plating industry in Bursa Demirtaş Organised Industrial Zone (Türkiye). The facility has a copper bath with a capacity of 5 m^3, a nickel bath with a capacity of 8 m^3 and a rinse bath with a volume of 20 m^3. Wastewater from plating tanks such as nickel and copper and rinsing bath, enters the balancing tank for homogenously mixed. Figure 1 shows the sampling point taken from the balancing tank in the metal plating industry. Since the plant has a standard production capacity, wastewater is formed at standard flow and characterization. Samples were taken twice a week and a total of 4 liters of composite samples were created, half a liter sample per hour. The obtained wastewater from the balancing tank in metal plating and stored at 4 °C in refrigerator before the experiments.

2.2. Coagulation-flocculation jar test procedure

Coagulation-flocculation study was carried out under laboratory conditions (24 °C) using a jar test device. During the research, 250 mL wastewater volume was used in each trial set. Each coagulation experiment was started by adding coagulant to the reactor containing wastewater, and then the appropriate initial pH was adjusted using HCl and NaOH. In the coagulation step, the reaction time was limited to 3 minutes and the process was continued at a stirring speed of 120 rpm [40]. The flocculation process was carried out at a mixing speed of 30 rpm with a contact time of 10 minutes [41]. The precipitation step was determined as 45 minutes and at the end of the process, samples were taken from the upper phase and analyzed for COD, Cu, Ni, Zn, SO_4.

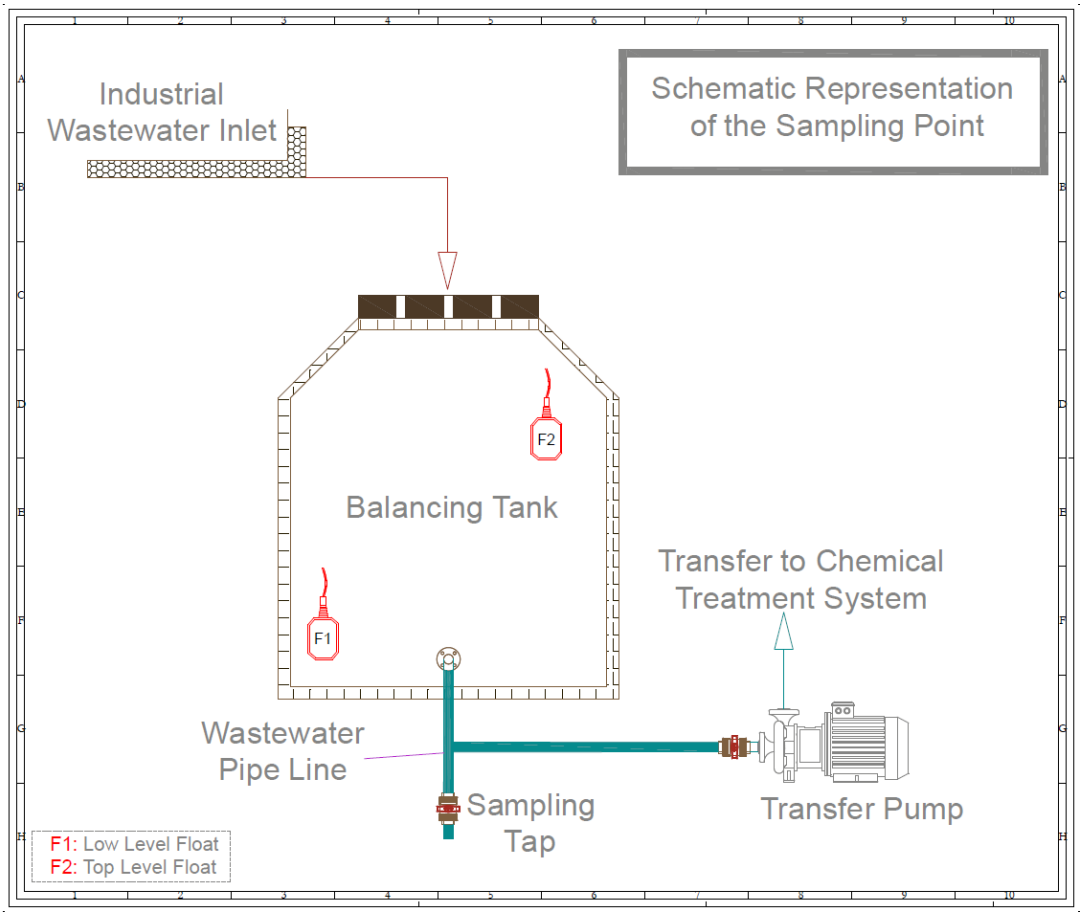


Figure 1. Metal plating industry sampling point.

The concentrations of coagulant to be used in the influent wastewater and the optimum pH values were determined at four different sets by Jar test in. Among all jar tests, doses and pH values of Set 4, which can provide limit values related to the discharge standard, were selected. In the metal plating industry wastewater, 45% NaOH, 5% $Ca(OH)_2$, 40% $FeCl_3$ and 0.1% $(NH_4)_2S_2O_8$ was used for coagulation-flocculation.

The aim of the first stage was to find the optimum concentrations in the coagulant doses to be used and the optimum pH value to which this dose was applied. By applying the chemicals and their doses used in Set 4 at the pH values specified in Table 1 and discharge standards for the compounds Cu, Ni, Zn, COD and SO_4 in sewages were provided. Chemicals used: NaOH: 2.75 mL; $Ca(OH)_2$: 11 g; it was used at the specified concentration, $FeCl_3$: 1.45 mL and $(NH_4)_2S_2O_8$: 6.5 g, and at a pH value between 6.4-12.3.

In the second phase of present study, removal efficiencies and effluent concentrations were investigated in order to reach the limit values specified for Cu, Ni, Zn, COD and SO_4 in the sewage and pre-treatment discharge criteria of Demirtaş Organized Industrial Zone [42]. The characterization of real metal plating wastewater is summarized in Table 2.

Table 1. Coagulants used in the study, doses and pH values

Coagulants	Doses				pH values			
	Set 1	Set 2	Set 3	Set 4	Set 1	Set 2	Set 3	Set 4
NaOH (mL)	3.3	3.1	3.1	2.75	11.8	7.6	8.0	6.4
$Ca(OH)_2$ (g)	-	9.5	12.5	11	-	11.7	12.8	12.3
$FeCl_3$ (mL)	1.3	1.2	1.5	1.45	9.5	9.4	8.8	8.8
$(NH_4)_2S_2O_8$ (g)	11	8	6	6.5	9.5	9.4	8.8	8.0

Table 2. Real metal plating industry wastewater characteristics.

Parameter	Present Study Value	Literature	Reference
pH	1.2	2.57	[44]
Conductivity (µS/cm)	240	-	-
COD (mg/L)	1230	159, 600, 4000	[12,44,45]
MLSS (mg/L)	1098	-	-
Cu (mg/L)	88.67	301	[12]
Ni (mg/L)	438.10	16, 51	[12,41]
Oil-grease (mg/L)	10	44	[44]
Zn (mg/L)	3.84	1.1, 145, 200	[41,44,45]
SO_4 (mg/L)	2797	-	-

2.3. Characterization

The conductivity and pH were measured using a pH meter and a conductivity device. COD (LCK-514), Cu (LCK-329), Ni (LCK-337), Zn (LCK-360), and SO_4 (LCK-353) were carried out with Hach test kits. Mixed Liquor Suspended Solids (MLSS) were determined according to Standard Methods (1994) (American Public Health Association, 1995). Oil-grease was analyzed by standard methods [43].

2.4. Operating cost

The operating cost was calculated in EUR (€) per cubic meter (m^3) wastewater volume, taking into account the optimum conditions in which the discharge parameters to the sewage were provided. The Equation 1 used in calculating the cost resulting from chemical consumption was given.

$$\text{Operating Cost} = c \times C_{chemical} \quad (1)$$

In Equation 1, c; the amount of chemicals required to treat 1 m^3 of wastewater (kg); $C_{chemical}$; cost of 1 kg of chemicals (€/kg).

3. Results and Discussion

In the metal plating industry, different compounds and concentrations are encountered depending on the chemicals used and the type of process. Ni concentration was 438.10 mg/L in present study and was higher than the literature in Table 2. While the oil-grease concentration was found to be lower than the value encountered in the literature; the Zn concentration (3.84 mg/L) was among the concentrations stated in the literature.

Additionally, COD concentrations were found at different values of 159 mg/L and 4000 mg/L. In present study, the COD value was determined as 1230 mg/L. In the jar tests performed in four sets which is given in Figure 2, the removal efficiency of the samples taken from the supernatant phase was examined.

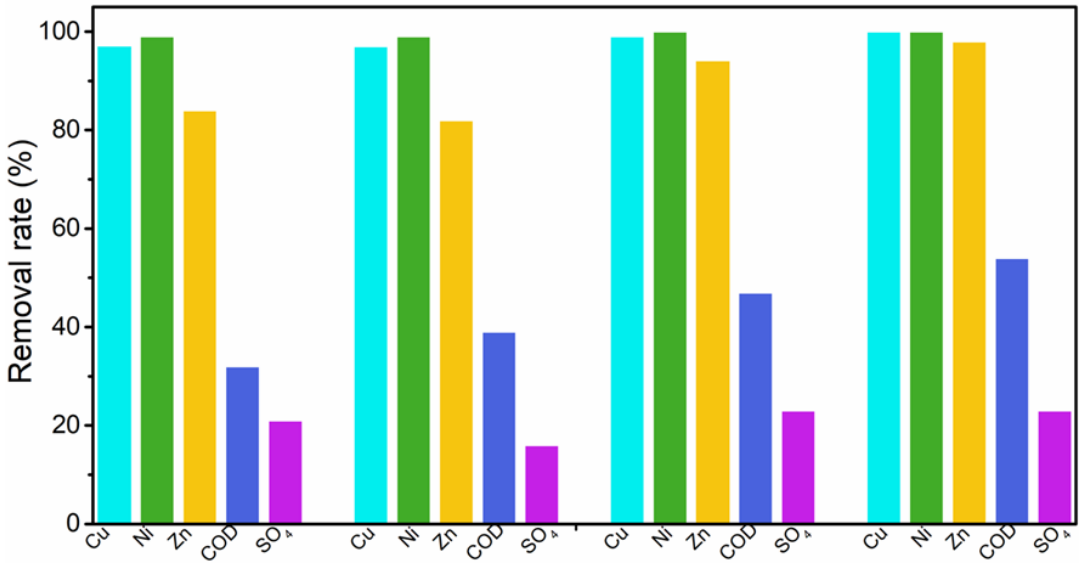


Figure 2. Removal efficiencies of pollutants in result of the jar test.

The pollutants removals were measured based on percent removal efficiency as indicated in the Equation 2 [46].

$$\text{Removal rate (\%)} = (C_0\text{-}C)/\ C_0 * 100 \quad (2)$$

Where C_0 is the concentration of pollutants in the influent wastewater sample and C is the concentration of pollutants treated wastewater sample. When the chemicals used in Set 1 and Set 2 in Figure 2 were applied at the concentration and pH values specified in Table 1, the removal of Cu, Ni and Zn pollutants were achieved but, the discharge standard into the sewage could not be supplied. Additionally, with Set 1, no active settling occurred after 45 minutes. The flocs remained suspended and in a small structure. With Set 2, treatment was more effective, and it was considered to increase the applied dose slightly. In Set 3, the discharge standard to the sewage was supplied, but in order to reduce the amount of chemicals spent and reduce the cost, the concentrations specified in Set 4 were studied and these values were applied. Yatim et al. [47] investigated the removal performance by coagulation and sedimentation processes in plating industry wastewater. Cu, Zn and Cr^{3+} removal efficiencies by hydroxide precipitation at pH 10.5 using 0.8 mL of $FeCl_3$ and 1 mL of NaOH solution, respectively; it was found to be 86.61%, 99.81% and 99.99%. In the same study, Cu was removed by 93.91%, Zn by 99.37% and Cr^{3+} by 99.99% by sulphide precipitation at pH 10 using 0.2 mL of $FeCl_3$ and 0.8 mL of Na_2S solution. In present study, Cu removal was found to be higher. As can be seen from the Figure 2, Cu and Ni removals were in the range of 97-100% in all sets. Also, Zn removal was similar to [47] and was found to be 98% in set 4. Zn removal from aqueous solution by coagulation process was investigated and anionic polyacrylamide was used. Zn removal was found ~100% [48]. It was found that Zn removal gradually increased from 82% to 98% and the highest Zn removal was found in set 4. Cu removal from aqueous solution by flocculation was investigated and acrylamide and acrylic acid-based co-polymers functionalized with hydroxamic

acid groups were used. At the end of the study, 67% Cu removal was founded [49]. Although the wastewater used in present study was real metal plating industry wastewater, Cu was almost completely removed (97-100%). In Table 3, the effluent concentrations of all compounds obtained from four sets and the limit values for discharge to sewage are given.

Figure 2, COD was removed to a limited and its removal remained between 32-54%, but as can be seen from Table 3, COD values met the discharge standard specified in the regulation. Polyaluminum chloride was used as coagulant using automotive wastewater and COD, Fe, Zn and Ni removal efficiencies were found to be 70%, 98%, 83% and 63%, respectively [50]. In the present study, although COD removal was limited to 54% compared to the literature, Zn and Ni removal efficiencies were found ≥97%.

Table 3. Effluent concentrations and limit values obtained from all sets.

Parameter	Effluent concentrations (mg/L)				Limit Values (mg/L) [42]
	Set-1	Set-2	Set-3	Set-4	
SO_4	2220	2344	2143	2150	1500
Cu	3.1	2.24	0,18	0.11	1
Ni	6.03	6	0.41	0.47	3
COD	838	751	654	568	3000
Zn	0.63	0.71	0.23	0.09	5
MLSS	-	-	-	< 10	1

SO_4 concentrations were found to be high due to the H_2SO_4 used in production. For SO_4 where the discharge standard could not be supplied, (5%) $BaCl_2$ (4 g, 8 g, 12 g and 16 g) was added which was the last stage of present study. When $BaCl_2$ was added at 4 g, 8 g, 12 g and 16 g the effluent concentrations were found to be 1618 mg/L, 545 mg/L, 297 mg/L and 153 mg/L, respectively. Since 545 mg/L SO_4 concentrations were reached with the 8 g $BaCl_2$ dose, the limit value (1500 mg/L) was supplied. pH values in the solution affect the efficiency of coagulation and flocculation. Optimum pH values for different metal types are stated as Ni: 10.5-11; Cu: 8-8.2; Al: 11; Cr: 8.5-9; Fe^{3+}: 3-4; Fe^{2+}: 8-9; Pb: 9-10; Zn: 9-10; Ag: 12; Cd: 9-11 [51]. While the pH values that provide the best removal of Cu through coagulation/flocculation processes in plating industry wastewater were ≥8; The best removal efficiency in Ni was at pH ≥10 [52]. In parallel with the literature, in present study, the applied pH value was kept in the range of 11-12 and the removal efficiency of Cu and Ni were 100%.

3.1. Operating cost

In the removal of pollutants, removal efficiency and operating cost are important parameters in the selection of the wastewater treatment process. In the coagulation flocculation process, the operating cost consists of the coagulant cost. As a result of the study, the cost of the chemicals used in Set 4 values and the standard of discharge to sewage was calculated. In Table 4 and 5, the consumption of coagulants added in 250 mL volumes is calculated. Since the chemicals used in set 4 supplied the criteria for discharge to sewage, the cost analysis was calculated over set 4. In addition, the dose of $BaCl_2$ used for reducing SO_4 concentration in Set 4 to supplied the discharge criteria was calculated and given in Table 6.

The amounts consumed for NaOH and $FeCl_3$ were found to be 4.11 and 2.08 g, respectively.

The amounts consumed for $Ca(OH)_2$ and $((NH4)_2S_2O_8$ were 0.55 and 0.0065 g, respectively. Table 6 shows the coagulant doses and costs for 1 m^3 of water. The total cost was found using Equation 1.

Unit Costs were added from the cash purchase price list dated 14.11.2023 of one of the chemical suppliers that the company purchases to operate the treatment plant. It was observed that the highest cost among the chemicals used was due to the use of NaOH due to the high consumption. If it is assumed that the plant works for 6 days and generates 150 m^3 of wastewater per week, 25 m^3 of wastewater will be generated daily. In this case, the cost will be 25*12.74=318.5 € (cost of coagulant used per day.)

Sludge disposal is one of the main cost elements in treatment processes. Sludge disposal methods in treatment plants spend 20-30% of the total construction cost of the plant and 50-70% of the operating cost on sludge disposal. The thickening process increases the solids density in the sludge and reduces the volume of the sludge by removing the water [53]. The chemical treatment sludge coming out of the plant is taken into the thickening process and converted into sludge cake with the help of filter presses to reduce the moisture content.

Using 250 mL sample, it was observed that 67 mL of sludge was formed at the bottom by settling the chemicals applied in set 4. Since the plant flow is 24 m_3/day 6.4 m^3/day of sludge ((24 m^3/day *67 mL)/250 mL) was formed.

Although the amount of dry solids in the sludge varies according to the treatment processes and the difference in the techniques applied, it has been reported that sludges generally contain between 5-12% dry solids, [54]

When the dry solids content is accepted as 5% in present study;

Solid content in sludge:

6.4 m^3/day*(1000 kg/m^3)*(0.05) = 320 kg/day of dry sludge generated in one day due to the chemicals used in the plant.

In filter presses, the dry matter content reaches up to 40% [55]. The sludge transformed into an easily transportable form is sent to sludge incineration plants.

The sludge from the filter press was calculated as:

Sludge from filter press: 320 kg/day / 0.4

800 kg sludge was sent to incineration plants.

Table 4. Consumption of NaOH and $FeCl_3$ according to Set 4 for 250 mL sample.

Coagulants	Concentration (%)	Consumption (mL)	Density (g/cm^3)	Consumption (g)
NaOH	45	2.75	1.495	4.11
$FeCl_3$	40	1.45	1.435	2.08

Table 5. Consumption of $Ca(OH)_2$ and $(NH_4)_2S_2O_8$ according to Set 4 for 250 mL sample.

Coagulants	Concentration/Solution (by mass) (%)	Solution Consumption (g)	Consumption calculation of the chemical used in the solution (g)
$Ca(OH)_2$	5	11	11*(5/100) = 0.55
$(NH_4)_2S_2O_8$	0.1	6.5	6.5*(0.1/100) = 0.0065
$BaCl_2$	0.5	8	8*(5/100) = 0.40

Table 6. Consumption and cost of coagulants used for 250 mL jar test in 1 m^3 wastewater.

Coagulants	Consumption in 1 m^3 Water (kg)	Unit Cost per kg (€)	Total Cost (€)
NaOH	16.44	0.45	7.40
$Ca(OH)_2$	2.2	0.10	0.22
$FeCl_3$	8.32	0.27	2.25
$(NH_4)_2S_2O_8$	0.26	2.75	0.07
$BaCl_2$	1.6	1.75	2.80
Cost of coagulants to be used for 1 m^3 wastewater			12.74 €

4. Conclusion

Coagulation-flocculation process has been investigated in order to operate the removal of heavy metals and COD from real wastewater in metal plating industry. Cu, Ni and Zn removal efficiencies were above 97% in set 4 by achieved the discharge criteria to sewage. In terms of COD removals, sewage discharge criteria were achieved. It was found that high concentration of SO_4 is detected due to the use of sulphuric acid in rinsing baths. SO_4 concentrations were also reduced by using $BaCl_2$.

The operating cost of the procedure has been studied. The total cost of the chemicals used in terms of the treatment plant is 318.5 € per day. The outcomes show that wastewater in metal plating industry have been successfully studied and heavy metals from wastewater has been removed and integrated to the wastewater treatment plant. No other coagulant other than $FeCl_3$ was used in present study. It will be more advantageous to make comparisons using different coagulants in the metal plating industry.

Sludge disposal is an important step in wastewater treatment processes, accounting for more than half of the total cost. Sludge disposal and management is therefore a serious challenge. Although the coagulation-flocculation process is one of the most widely used treatment processes in industrial wastewater due to its easy applicability and high efficiency, sludge generated as a result of the coagulation-flocculation process contains high concentrations of heavy metals can cause problems in the sludge treatment-disposal stage. Moreover, synthetic polymer flocculants can create large amounts of highly toxic sludge. For these reasons, the sludge causes serious problems in terms of environment and health. Also, heavy metals and toxicity are major disadvantages in the conversion of waste into valuable resources.

Green materials such as waste-based products and renewable materials can be used to produce environmentally friendly and biodegradable coagulants. Thus, negative impacts on the environment can be avoided and minimized and the cost of treatment can be significantly reduced.

Author contributions

Inci Karakas: Data curation, Methodology, Writing-Original draft preparation, Investigation, Writing-Reviewing and Editing.
Soner Kızıl: Writing-Original draft preparation, Investigation, Writing and Reviewing.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Palansooriya, K. N., Yang, Y., Tsang, Y. F., Sarkar, B., Hou, D., Cao, X., ... & Ok, Y. S. (2020). Occurrence of contaminants in drinking water sources and the potential of biochar for water quality improvement: A review. Critical Reviews in Environmental Science and Technology, 50(6), 549-611. https://doi.org/10.1080/10643389.2019.1629803
2. Crini, G., & Lichtfouse, E. (2019). Advantages and disadvantages of techniques used for wastewater treatment. Environmental Chemistry Letters, 17, 145-155. https://doi.org/10.1007/s10311-018-0785-9
3. Karakas, I., Sam, S. B., Cetin, E., Dulekgurgen, E., & Yilmaz, G. (2020). Resource recovery from an aerobic granular sludge process treating domestic wastewater. Journal of Water Process Engineering, 34, 101148. https://doi.org/10.1016/j.jwpe.2020.101148
4. Zamora-Ledezma, C., Negrete-Bolagay, D., Figueroa, F., Zamora-Ledezma, E., Ni, M., Alexis, F., & Guerrero, V. H. (2021). Heavy metal water pollution: A fresh look about hazards, novel and conventional remediation methods. Environmental Technology & Innovation, 22, 101504. https://doi.org/10.1016/j.eti.2021.101504

5. Yati, I., Kizil, S., & Bulbul Sonmez, H. (2019). Cellulose-based hydrogels for water treatment. In Cellulose-Based Superabsorbent Hydrogels, 1015-1037. Springer, Cham.
6. Qasem, N. A., Mohammed, R. H., & Lawal, D. U. (2021). Removal of heavy metal ions from wastewater: A comprehensive and critical review. Npj Clean Water, 4(1), 36. https://doi.org/10.1038/s41545-021-00127-0
7. Shrestha, R., Ban, S., Devkota, S., Sharma, S., Joshi, R., Tiwari, A. P., ... & Joshi, M. K. (2021). Technological trends in heavy metals removal from industrial wastewater: A review. Journal of Environmental Chemical Engineering, 9(4), 105688. https://doi.org/10.1016/j.jece.2021.105688
8. Tchounwou, P. B., Yedjou, C. G., Patlolla, A. K., & Sutton, D. J. (2012). Heavy metal toxicity and the environment. Molecular, Clinical and Environmental Toxicology: Volume 3: Environmental Toxicology, 133-164. https://doi.org/10.1007/978-3-7643-8340-4_6
9. Zhou, Q., Yang, N., Li, Y., Ren, B., Ding, X., Bian, H., & Yao, X. (2020). Total concentrations and sources of heavy metal pollution in global river and lake water bodies from 1972 to 2017. Global Ecology and Conservation, 22, e00925. https://doi.org/10.1016/j.gecco.2020.e00925
10. Briffa, J., Sinagra, E., & Blundell, R. (2020). Heavy metal pollution in the environment and their toxicological effects on humans. Heliyon, 6(9), e04691. https://doi.org/10.1016/j.heliyon.2020.e04691
11. Zhu, Y., Fan, W., Zhou, T., & Li, X. (2019). Removal of chelated heavy metals from aqueous solution: A review of current methods and mechanisms. Science of the Total Environment, 678, 253-266. https://doi.org/10.1016/j.scitotenv.2019.04.416
12. Nguyen, M. K., Pham, T. T., Pham, H. G., Hoang, B. L., Nguyen, T. H., Nguyen, T. H., ... & Ngo, H. H. (2021). Fenton/ozone-based oxidation and coagulation processes for removing metals (Cu, Ni)-EDTA from plating wastewater. Journal of Water Process Engineering, 39, 101836. https://doi.org/10.1016/j.jwpe.2020.101836
13. Hosseini, S. S., Bringas, E., Tan, N. R., Ortiz, I., Ghahramani, M., & Shahmirzadi, M. A. A. (2016). Recent progress in development of high performance polymeric membranes and materials for metal plating wastewater treatment: A review. Journal of Water Process Engineering, 9, 78-110. https://doi.org/10.1016/j.jwpe.2015.11.005
14. Katsumata, H., Kaneco, S., Inomata, K., Itoh, K., Funasaka, K., Masuyama, K., ... & Ohta, K. (2003). Removal of heavy metals in rinsing wastewater from plating factory by adsorption with economical viable materials. Journal of Environmental Management, 69(2), 187-191. https://doi.org/10.1016/S0301-4797(03)00145-2
15. Kurniawan, T. A., Chan, G. Y., Lo, W. H., & Babel, S. (2006). Physico–chemical treatment techniques for wastewater laden with heavy metals. Chemical Engineering Journal, 118(1-2), 83-98. https://doi.org/10.1016/j.cej.2006.01.015
16. Hosseini, S. S., Bringas, E., Tan, N. R., Ortiz, I., Ghahramani, M., & Shahmirzadi, M. A. A. (2016). Recent progress in development of high performance polymeric membranes and materials for metal plating wastewater treatment: A review. Journal of Water Process Engineering, 9, 78-110. https://doi.org/10.1016/j.jwpe.2015.11.005
17. Oden, M. K., & Sari-Erkan, H. (2018). Treatment of metal plating wastewater using iron electrode by electrocoagulation process: Optimization and process performance. Process Safety and Environmental Protection, 119, 207-217. https://doi.org/10.1016/j.psep.2018.08.001
18. Akbal, F., & Camcı, S. (2012). Treatment of metal plating wastewater by electrocoagulation. Environmental Progress & Sustainable Energy, 31(3), 340-350. https://doi.org/10.1002/ep.10546
19. Zoungrana, A., Çakmakci, M., Zengin, İ. H., İnoğlu, Ö., & Elcik, H. (2016). Treatment of metal-plating waste water by modified direct contact membrane distillation. Chemical Papers, 70(9), 1185-1195. https://doi.org/10.1515/chempap-2016-0066
20. Noulas, C., Tziouvalekas, M., & Karyotis, T. (2018). Zinc in soils, water and food crops. Journal of Trace Elements in Medicine and Biology, 49, 252-260. https://doi.org/10.1016/j.jtemb.2018.02.009
21. Kumar, A., Kumar, A., MMS, C. P., Chaturvedi, A. K., Shabnam, A. A., Subrahmanyam, G., ... & Yadav, K. K. (2020). Lead toxicity: health hazards, influence on food chain, and sustainable remediation approaches. International Journal of Environmental Research and Public Health, 17(7), 2179. https://doi.org/10.3390/ijerph17072179
22. Witkowska, D., Słowik, J., & Chilicka, K. (2021). Heavy metals and human health: Possible exposure pathways and the competition for protein binding sites. Molecules, 26(19), 6060. https://doi.org/10.3390/molecules26196060
23. Kesari, K. K., Soni, R., Jamal, Q. M. S., Tripathi, P., Lal, J. A., Jha, N. K., ... & Ruokolainen, J. (2021). Wastewater treatment and reuse: a review of its applications and health implications. Water, Air, & Soil Pollution, 232, 1-28. https://doi.org/10.1007/s11270-021-05154-8
24. Hunsom, M., Pruksathorn, K., Damronglerd, S., Vergnes, H., & Duverneuil, P. (2005). Electrochemical treatment of heavy metals (Cu^{2+}, Cr^{6+}, Ni^{2+}) from industrial effluent and modeling of copper reduction. Water Research, 39(4), 610-616. https://doi.org/10.1016/j.watres.2004.10.011
25. Ilhan, F., Ulucan-Altuntas, K., Avsar, Y., Kurt, U., & Saral, A. (2019). Electrocoagulation process for the treatment of metal-plating wastewater: Kinetic modeling and energy consumption. Frontiers of Environmental Science & Engineering, 13, 1-8. https://doi.org/10.1007/s11783-019-1152-1
26. Baskar, A. V., Bolan, N., Hoang, S. A., Sooriyakumar, P., Kumar, M., Singh, L., ... & Siddique, K. H. (2022). Recovery, regeneration and sustainable management of spent adsorbents from wastewater treatment streams: A review. Science of the Total Environment, 822, 153555. https://doi.org/10.1016/j.scitotenv.2022.153555

27. Kaur, J., Sengupta, P., & Mukhopadhyay, S. (2022). Critical review of bioadsorption on modified cellulose and removal of divalent heavy metals (Cd, Pb, and Cu). Industrial & Engineering Chemistry Research, 61(5), 1921-1954. https://doi.org/10.1021/acs.iecr.1c04583
28. Braga, W. L. M., Roberto, J. A., Vaz, C., Samanamud, G. R. L., Loures, C. C. A., Franca, A. B., ... & Naves, F. L. (2018). Extraction and optimization of tannin from the flower of Musa sp. applied to the treatment of iron ore dump. Journal of Environmental Chemical Engineering, 6(4), 4310-4317. https://doi.org/10.1016/j.jece.2018.05.058
29. El Gaayda, J., Rachid, Y., Titchou, F. E., Barra, I., Hsini, A., Yap, P. S., ... & Akbour, R. A. (2023). Optimizing removal of chromium (VI) ions from water by coagulation process using central composite design: Effectiveness of grape seed as a green coagulant. Separation and Purification Technology, 307, 122805. https://doi.org/10.1016/j.seppur.2022.122805
30. Huang, J., Yuan, F., Zeng, G., Li, X., Gu, Y., Shi, L., ... & Shi, Y. (2017). Influence of pH on heavy metal speciation and removal from wastewater using micellar-enhanced ultrafiltration. Chemosphere, 173, 199-206. https://doi.org/10.1016/j.chemosphere.2016.12.137
31. Chang, S., Ahmad, R., Kwon, D. E., & Kim, J. (2020). Hybrid ceramic membrane reactor combined with fluidized adsorbents and scouring agents for hazardous metal-plating wastewater treatment. Journal of Hazardous Materials, 388, 121777. https://doi.org/10.1016/j.jhazmat.2019.121777
32. Agridiotis, V., Forster, C. F., & Carliell-Marquet, C. (2007). Addition of Al and Fe salts during treatment of paper mill effluents to improve activated sludge settlement characteristics. Bioresource Technology, 98(15), 2926-2934. https://doi.org/10.1016/j.biortech.2006.10.004
33. Al-Shannag, M., Al-Qodah, Z., Bani-Melhem, K., Qtaishat, M. R., & Alkasrawi, M. (2015). Heavy metal ions removal from metal plating wastewater using electrocoagulation: Kinetic study and process performance. Chemical Engineering Journal, 260, 749-756. https://doi.org/10.1016/j.cej.2014.09.035
34. Sinha, S., Yoon, Y., Amy, G., & Yoon, J. (2004). Determining the effectiveness of conventional and alternative coagulants through effective characterization schemes. *Chemosphere*, *57*(9), 1115-1122. https://doi.org/10.1016/j.chemosphere.2004.08.012
35. Inam, M. A., Khan, R., Akram, M., Khan, S., Park, D. R., & Yeom, I. T. (2019). Interaction of arsenic species with organic ligands: Competitive removal from water by coagulation-flocculation-sedimentation (C/F/S). Molecules, 24(8), 1619. https://doi.org/10.3390/molecules24081619
36. Folens, K., Huysman, S., Van Hulle, S., & Du Laing, G. (2017). Chemical and economic optimization of the coagulation-flocculation process for silver removal and recovery from industrial wastewater. Separation and Purification Technology, 179, 145-151. https://doi.org/10.1016/j.seppur.2017.02.013
37. Zhao, C., Shao, S., Zhou, Y., Yang, Y., Shao, Y., Zhang, L., ... & Luo, L. (2018). Optimization of flocculation conditions for soluble cadmium removal using the composite flocculant of green anion polyacrylamide and PAC by response surface methodology. Science of the Total Environment, 645, 267-276. https://doi.org/10.1016/j.scitotenv.2018.07.070
38. Sun, Y., Chen, A., Pan, S. Y., Sun, W., Zhu, C., Shah, K. J., & Zheng, H. (2019). Novel chitosan-based flocculants for chromium and nickle removal in wastewater via integrated chelation and flocculation. Journal of Environmental Management, 248, 109241. https://doi.org/10.1016/j.jenvman.2019.07.012
39. Yang, K., Wang, G., Chen, X., Wang, X., & Liu, F. (2018). Treatment of wastewater containing Cu2+ using a novel macromolecular heavy metal chelating flocculant xanthated chitosan. Colloids and Surfaces A: Physicochemical and Engineering Aspects, 558, 384-391. https://doi.org/10.1016/j.colsurfa.2018.06.082
40. El Gaayda, J., Rachid, Y., Titchou, F. E., Barra, I., Hsini, A., Yap, P. S., ... & Akbour, R. A. (2023). Optimizing removal of chromium (VI) ions from water by coagulation process using central composite design: Effectiveness of grape seed as a green coagulant. Separation and Purification Technology, 307, 122805. https://doi.org/10.1016/j.seppur.2022.122805
41. Kılıç, M. Y., & Kumbasar, P. (2023). Çinko-nikel alaşım kaplama atıksularının kimyasal arıtabilirliğinin araştırılması. Uludağ Üniversitesi Mühendislik Fakültesi Dergisi, 28(1), 307-316. https://doi.org/10.17482/uumfd.1249112
42. DOSAB (2019). DOSAB Atıksu Altyapı Yönetim Talimatı 2019.
43. Baird, R., Rice, E., & Eaton, A. (2017). Standard methods for the examination of water and wastewaters. Water Environment Federation, Chair Eugene W. Rice, American Public Health Association Andrew D. Eaton, American Water Works Association.
44. Öztel, M. D., Kuleyin, A., & Akbal, F. (2020). Treatment of zinc plating wastewater by combination of electrocoagulation and ultrafiltration process. Water Science and Technology, 82(4), 663-672. https://doi.org/10.2166/wst.2020.357
45. Yatim, S. R. M., Zainuddin, N. A., Mokhtar, N. S., Syahjidan, H. N., & Kamsuri, S. N. H. (2021, February). Competitiveness in removing copper, zinc and chromium trivalent in plating industrial effluent by using hydroxide precipitation versus sulphide precipitation. In IOP Conference Series: Materials Science and Engineering, 1053(1), 012084. https://doi.org/10.1088/1757-899X/1053/1/012084
46. Al-Shannag, M., Al-Qodah, Z., Bani-Melhem, K., Qtaishat, M. R., & Alkasrawi, M. (2015). Heavy metal ions removal from metal plating wastewater using electrocoagulation: Kinetic study and process

performance. Chemical Engineering Journal, 260, 749-756. https://doi.org/10.1016/j.cej.2014.09.035
47. Yatim, S. R. M., Zainuddin, N. A., Mokhtar, N. S., Syahjidan, H. N., & Kamsuri, S. N. H. (2021). Competitiveness in removing copper, zinc and chromium trivalent in plating industrial effluent by using hydroxide precipitation versus sulphide precipitation. In IOP Conference Series: Materials Science and Engineering, 1053(1), 012084. https://doi.org/10.1088/1757-899X/1053/1/012084
48. Xu, D., Zhou, B., & Yuan, R. (2019). Optimization of coagulation-flocculation treatment of wastewater containing Zn (II) and Cr (VI). In IOP Conference Series: Earth and Environmental Science, 227, 052049. https://doi.org/10.1088/1755-1315/227/5/052049
49. Arun, Y., Daifa, M., & Domb, A. J. (2021). Polyhydroxamic acid as an efficient metal chelator and flocculant for wastewater treatment. Polymers for Advanced Technologies, 32(2), 842-852. https://doi.org/10.1002/pat.5135
50. Bakar, A. F. A., & Halim, A. A. (2013). Treatment of automotive wastewater by coagulation-flocculation using poly-aluminum chloride (PAC), ferric chloride ($FeCl_3$) and aluminum sulfate (alum). In AIP conference proceedings, 1571(1), 524-529. https://doi.org/10.1063/1.4858708
51. Hoffland Environmental (2020). https://heienv.com/hydroxide-precipitation-of-metals/#:~;:text=It%20is%20common%20to%20utilize.9.5%20to%20precipitate%20both%20metals
52. Nguyen, M. K., Pham, T. T., Pham, H. G., Hoang, B. L., Nguyen, T. H., Nguyen, T. H., ... & Ngo, H. H. (2021). Fenton/ozone-based oxidation and coagulation processes for removing metals (Cu, Ni)-EDTA from plating wastewater. Journal of Water Process Engineering, 39, 101836. https://doi.org/10.1016/j.jwpe.2020.101836
53. Yüksekdağ, M., Gökpınar, S., & Yelmen, B. (2020). Atıksu Arıtma Tesislerinde Arıtma Çamurları ve Bertaraf Uygulamaları. Avrupa Bilim ve Teknoloji Dergisi, (18), 895-904. https://doi.org/10.31590/ejosat.699952
54. Spinosa, L. & Vesilind, P. A. (2001). Sludge into biosolids. IWA Publishing.
55. Cevik, A. (2017). Atık su arıtma çamuru susuzlaştırma kurutma ve yakmada kullanılan yöntem ve ekipmanlar.

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

A novel strategy to avoid local optimum: Army-inspired genetic algorithm (AIGA)

Müslüm Kilinc[1], Emrah Atılgan[*2], Cengiz Atiş [1]

[1] *Erciyes University, Department of Civil Engineering, Türkiye, kilinc@erciyes.edu.tr; cdatis@erciyes.edu.tr*
[2] *Eskişehir Osmangazi University, Department of Computer Engineering, Türkiye, emrah.atilgan@ogu.edu.tr*

Cite this study: Kılınç, M., Atılgan, E., & Atış, C. (2024). A novel strategy to avoid local optimum: Army-inspired genetic algorithm (AIGA). Turkish Journal of Engineering, 8 (3), 436-446

https://doi.org/10.31127/tuje.1412271

Keywords
Stochastic optimization
Evolutionary algorithm
Genetic algorithm
Army-inspired strategy
Local optima

Research Article

Received: 31.12.2023
Revised: 27.02.2024
Accepted: 29.02.2024
Published: 05.07.2024

Abstract

Objective functions of which an analytical solution is very difficult or time-consuming are solved using stochastic optimization algorithms. Those optimization algorithms compute an approximate solution for objective functions. For a specific search space, the objective function might have one or more local optima along with the global optimum. When a comparison is made among the algorithms, one optimization algorithm could be more effective than others in finding a solution for certain objective functions. The most important factors affecting the success of optimization algorithms are the greatness of search space and the complexity of the objective function. Reaching the global optimum in huge search spaces is very difficult. In complex objective functions that have many local optima or where the differences between global optimum and local optima are very small, the probability of trapping into the local optimum is high. Existing optimization algorithms could be improved using the search space scanned more successfully to give a better performance. To achieve this aim, we present a novel algorithm, called Army-Inspired Genetic Algorithm (AIGA), which is inspired from military movement. The presented algorithm, apart from other optimization algorithms, searches global optima effectively by dividing the entire search area into territories instead of searching in one piece. Thus, the probability of getting trapped in a local optimum reduces and the probability of finding the global optimum increases. The presented algorithm was tested on well-known benchmark problems. The results shows that AIGA is more efficient algorithm in finding the global optimum than traditional algorithms.

1. Introduction

It is known that there are many functions used in different fields of science. Sometimes it is hard (or impossible) to obtain the exact solution of a function's optimum. Thus, optimization methods have been used for approximation of optimal solutions. Many optimization methods in the fields like engineering, computer science, business, bioinformatics, and statistics, etc., have been applied to obtain the optimal solution [1]. When it is hard to achieve the exact solution and computationally expensive, optimization methods speed up the process. Particularly, while the optimal solution is under investigation in a large search space, optimization algorithms systematically search the best solution. Based on the optimization problem, the main aim is to minimize or maximize the objective function.

Stochastic optimization techniques are the member of optimization methods which use random variables. The process is started with random initial variables and randomly searches the fitness landscape for the possible optimal solution. Over the years, many different algorithms and methods have been developed by researchers for optimization [2-6]. In the following, a brief literature review was given on the heuristic algorithms that used in this study for comparison. Grey Wolf Optimization (GWO) [7] is a metaheuristic inspired by the social structure and hunting patterns of grey wolves. GWO is particularly effective in solving optimization problems with a continuous search space. The algorithm emulates the organizational structure of a wolf pack, where the most dominant members are symbolized by alpha, beta, and delta wolves. By iteratively updating the positions of these wolves, GWO seeks to converge towards the optimal solution. Particle Swarm Optimization (PSO) [8] is a swarm intelligence technique inspired by the collective behavior of birds and fish. In PSO, a group of particles explores a search space, modifying their positions based on both individual and collective experiences. The velocity of each particle is

adapted according to its own best-known position and the overall best-known position of the entire swarm. This enables effective global exploration and exploitation during the optimization process. Firefly Algorithm (FA) [9] draws inspiration from the natural flashing patterns exhibited by fireflies. Fireflies are attracted to each other, and this attraction is used as the basis for optimization. The algorithm utilizes the brightness of fireflies to represent the objective function values. Firefly movements are then governed by attractive and repulsive forces, guiding the search towards optimal solutions in the optimization landscape. Bat Algorithm (BA) [10] draws inspiration from the echolocation behavior of bats. In this method, simulated bats navigate through the exploration space by modifying both their frequencies and loudness. The echolocation pulses determine the proximity of bats to potential solutions. Additionally, randomness is introduced to simulate the foraging behavior of bats, improving the algorithm's capacity to efficiently navigate and take advantage of the search space. Crow Search Algorithm (CSA) [11] is based on the social foraging behavior of crows. The algorithm introduces multiple flocks of crows, each representing a candidate solution. Crows share information about their positions, allowing the algorithm to balance exploration and exploitation. The algorithm incorporates a global exploration strategy to discover diverse solutions and a local exploitation mechanism for refining the search around promising regions. Cuckoo Search (CS) [12] draws inspiration from the reproductive habits of cuckoo birds, which involve depositing their eggs in the nests of different bird species, and the host birds may reject eggs that deviate from their expected characteristics. This rejection behavior is mimicked in the algorithm to generate new solutions. Cuckoo Search aims to balance exploration and exploitation by incorporating random walks and Levy flights [13] to navigate the search space efficiently. Flower Pollination Algorithm (FPA) [14] draws inspiration from the pollination process in plants, where flowers share genetic information to achieve successful reproduction. In the algorithm, flowers represent candidate solutions, and the pollination process corresponds to the exchange of information between flowers. This approach improves the algorithm's capacity to navigate the search space and move towards achieving optimal solutions.

One of the most powerful optimization algorithms is Genetic Algorithms (GAs) inspired by natural selection. GA works according to the Darwin principle: individuals best adapted to development will survive. This improvement is achieved through generations with genetic operators called crossover and mutation. Individuals newly generated using these operators are sorted based on their fitness values, the good ones are kept in the next generation, and those who cannot adapt are removed from the population. Genetic algorithms are among the most preferred search algorithms and optimization algorithms. One of the main disadvantages of GAs is getting trapped into local optima. Although much work has been done to overcome this problem [15-18], it has not yet been fully solved.

The main issue of such stochastic algorithms is largeness search space boundaries. The narrower search space results with more approximate solutions in less computational time. In such algorithms, the solution can be searched in narrower search spaces with the clustering method, however, in this case the computational time might be longer than as expected due to decrease of the population size for each clustered space. As an advantage, the clustering method prevents being trapped with a local solution. If there is no global optimum in a cluster, this will waste time and effort. Moreover, assigning individuals in the population in a non-result cluster will delay reaching the global optimum. The proposed study utilizing modified clustering method based on army movement strategy prevents idle operation in search subspaces and avoids being trapped in a local solution. Army-Inspired Genetic Algorithm (AIGA) was adapted to the traditional genetic algorithm. The efficiency of AIGA has been illustrated by solving the benchmark optimization problems given in the literature [19-20].

2. Method

The main idea of the army-inspired method is inspired from capturing a specific area with a limited number of soldiers. During a war, the army cannot attack the enemy at a single point if enemy's main base is unknown. Priority target is to locate and conquer the headquarters of the enemy. This can be achieved by scanning the whole enemy area. Therefore, the commander should split the whole region into small areas and send a certain number of pioneer soldiers equally to gather information about the quantity and location of the enemy soldiers and bases. Based on feedback information from pioneer soldiers, more soldiers are to be dispatched to the possible main base, and less soldiers are kept at the other bases. Thus, while fighting the main enemy in a region, the army is also on guard against enemy threats from other regions. If a larger enemy base is detected in another area, the army needs to attack there by dispatching more troops immediately.

In the proposed algorithm AIGA, each soldier represents an individual while the army and troops correspond to the whole population and sub-populations, respectively. As well as greater or stronger enemies are symbolized as the optimal points of optimization problems and greatness of these enemies are defined by the fitness function value. Whilst, main enemy base is represented by global optimum point, smaller bases correspond to local optima points. The search space defines the whole region, and the sub-search space represents the sub-regions where the troops are located. The most important criterion in this method is population number namely the number of soldiers. Depending on the size of the number of soldiers at present, either the entire area or only a limited area can be conquered.

An ordinary stochastic optimization algorithm begins with a defined search space and a certain population size. Apart from other algorithms, the proposed algorithm, AIGA, initiates by dividing the entire search space into smaller sub-search spaces. Since there is no specific information or trail available at the beginning of a search,

the population is partitioned equally into the sub-populations in similar manner for the search space partitioning. Starting from the first sub-search space to the last sub-search space, TGA runs using the number of generations, for each sub-search space and sub-population. After each certain number of generations, the sub-population sizes in each sub-search space are determined using a roulette-wheel method based on evaluation of the population. Since the total population size is constant, the population in some sub-search spaces could be increased, and some could be decreased. Thus, the search continues with a higher population in the sub-search space where the current optimum point is, while in other subspaces the search continues with less individuals. If a new current optimum point is found in a sub-search space with a smaller number of individuals, the majority of the population is directed here to reach a better point in this sub-search space. An illustration of individuals' movement is shown in Figure 1.

In this method, depending on the present situation, it is aimed to search in the sub-search spaces with the sub-population instead of using all population in the entire search space. The method aims to obtain the result in a sub-search space which may contain the highly probable solution. While the majority of the population searches in one subspace, the exploration persists across other sub-search spaces with fewer individuals to prevent becoming confined to a local optimum. The pseudocode code of AIGA is given with Algorithm 1 and 2.

Figure 1. Illustration of AIGA with 4-subspaces over generations.

Algorithm 1: Army-Inspired Genetic Algorithm (AIGA).

```
Set Parameters
Partitioning:
    All Search Space is divided into Sub-Search Spaces. (Number of Sub-Search Spaces is user-defined.)
Distribution:
    All Population is distributed into the sub-space searches. (Initially, equal distribution)
Stopping condition ← false.
while stopping condition==false
    foreach search space
        set options of population size, and initial population.
        run Genetic Algorithm (GA) (details in Alg. 2)
    end
    Gathering:
        Sub-Populations come together.
    Distribution:
        Whole population is distributed into the sub-space searches with the Roulette Wheel Method.
        (Sub-population size of every sub-space search is determined based on current best fitness value.)
    Check if the stopping condition is true.
end
```

Algorithm 2. Genetic Algorithm (GA).

```
Set Initial Population
Evaluation of Fitness Values
Sorting of Fitness Values
Termination criteria ← false.
while Termination criteria is not met
  foreach generation (Generation Number is optional)
    Selection: (Method is optional, Tournament, Roulette Wheel, etc.)
    Crossover: (Method is optional, One point, two points, multi-points, arithmetic, uniform, etc.)
    Mutation: (Mutation Rate, Population Size, Mutation Limits, etc. are optional)
    Evaluation of Fitness Values
    Sorting of Fitness Values
  end
  Check if Termination criteria is met?
end
```

3. Results

In this study, to demonstrate the success of the proposed method, multi-modal benchmark optimization problems that contain global and local optima have been tested. Therefore, two-dimensional multimodal functions having a few local optima were employed for visuality and illustration.

When addressing an optimization challenge, various factors come into play, influencing the algorithm's overall performance. These parameters may vary depending on the algorithm. For example, the success of the conventional genetic algorithm is influenced by various factors, including but not limited to the population size, number of generations, selection method, crossover method, elite population rate, mutation rate, and mutation amount. In fact, an optimization algorithm for two different problems may show different performance with the same parameters because of the nature of the stochastic search. Under this acknowledgment, AIGA has been compared to a traditional genetic algorithm using the same parameters for a fair comparison. For this, Damavandi function [21], which is well known as one of the difficult optimization problems in the literature, and Gaussian-Like function were chosen.

For each case study, same operators and parameters were used in the traditional genetic algorithm and AIGA. For instance, the tournament method was chosen for selection operator, and the arithmetic crossover method is chosen for crossover operator. Furthermore, the traditional genetic algorithm and AIGA share identical settings for parameters such as population size, number of generations, elite population rate, mutation rate, and total number of evaluations.

AIGA is not a new algorithm, but a novel strategy that has not been proposed before. Processing AIGA for a single group corresponds to running a traditional genetic algorithm (TGA). Therefore, the sample optimization problems were processed by increasing the number of sub-spaces to show the effect of AIGA. Accordingly, the results were presented on graphics and tables.

3.1 Case Study 1: Damavandi Function

As the first case study, the multi-model Damavandi function was considered as a test function with 2-dimensions. The equation for the Damavandi function is given in Equation 1.

The boundaries of the defined function have been set as 0 and 14 for each variable, x_1 and x_2. As seen in Figure 2, the function has a local optimum along with the global optimum.

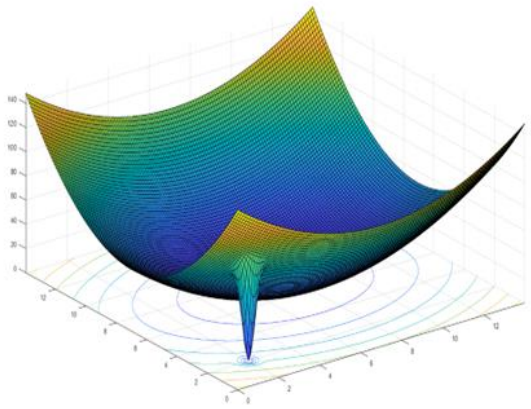


Figure 2. Damavandi function.

$$f(x) = \left[1 - \left|\frac{sin[\pi(x_1 - 2)]sin[\pi(x_2 - 2)]}{\pi^2(x_1 - 2)(x_2 - 2)}\right|^5\right][2 + (x_1 - 7)^2 + (x_2 - 7)^2] \quad (1)$$

Since the area, where the global optimum exists, has small amplitude, the solution to be obtained from any optimization algorithm is highly probable to be stuck in the local optimum. For this reason, Damavandi function is a very exceptional optimization problem to use in comparing or testing the performance of optimization algorithms.

For this optimization problem, the total number of evaluations was set to 20,000. TGA and AIGA were compared under different population sizes from 300 to 1800 by incrementing 100. In a similar manner, the number of groups as a parameter of AIGA were assigned values from 1 to 25. AIGA was run 1,000 times for each population size and the number of groups. The results obtained are presented by both in tabular form and the graphs as the success rate (%) in achieving global optimum for Damavandi function.

In Figure 3, each subplot demonstrates success rates as percentage (y-axis) value versus the number of groups (x-axis) at which AIGA runs at different population sizes on Damavandi function. As seen in Figure 3, the success rate increases (in most cases) as the number of groups increases for each population size. When comparison made between TGA (corresponds to 1-group AIGA) and AIGA (2 or more groups), the increase in success of AIGA is seen prominently as the number of groups increases.

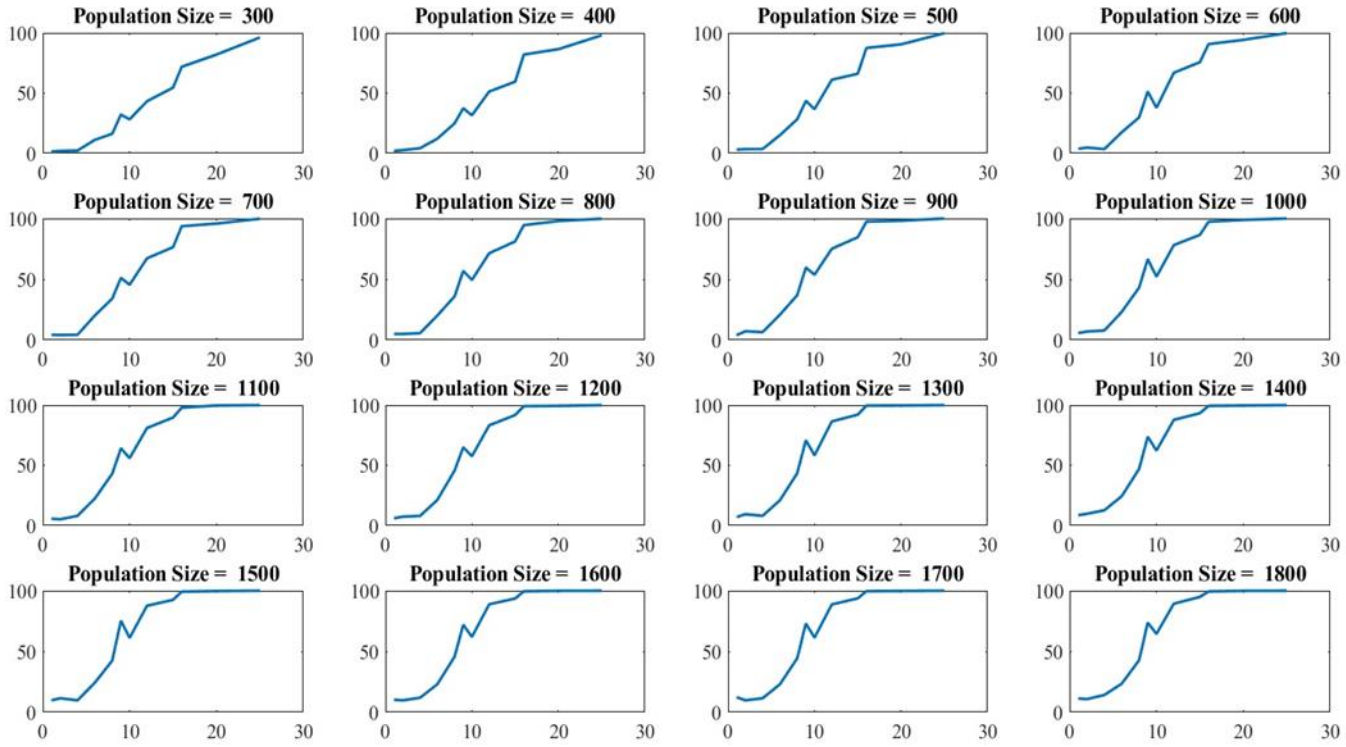


Figure 3. Success Rates (%) vs Number of Groups for Different Population Size

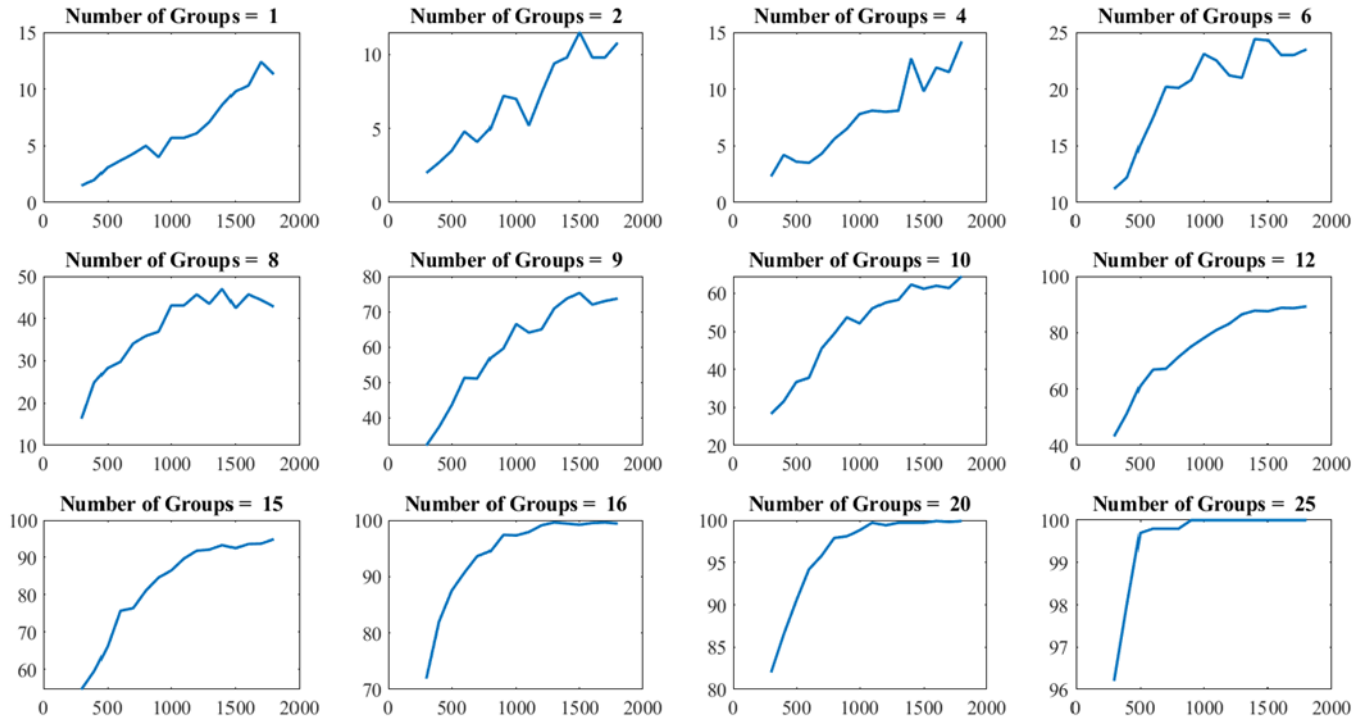


Figure 4. Success rates (%) vs population size for different number of groups.

In Figure 4, each subplot demonstrates success rates as percentage (y-axis) value versus the population size (x-axis) at which AIGA runs at different numbers of groups on Damavandi function. It can be seen in the Figure 4; the success rate increases as population size increases for each group number. However, the increase in success rate here is not as much as the increase in success rate achieved by increasing the number of groups. In multimodal functions, it is normal to expect that for any optimization algorithm the larger population size the better result. Since population is distributed into zones as in AIGA, augmenting the quantity of groups results in a reduction of individuals within each group. Thus, increase in population size becomes more effective in success of finding global optimum as the number of groups increases.

The success rates of AIGA in finding the global optimum of the Damavandi function dependent on both group number and population size are shown in Table 1. As the average success rates are considered, the increase in population size or number of groups or increase in both result with an enhancement in the success rate of finding global optimum. The largest group number exhibits a success rate of 99.60 percent on average, whereas the average success rate for the largest population size is 60.40 percent. Namely, the number of groups as a parameter of AIGA is much more effective than the population size as a decisive parameter to obtain a better success rate. While the population size for 1-group corresponding to TGA was increased from 300 to 1800, the success rate reached 11.3 percent from 1.5 percent. Similarly, while the population size for 25-groups of AIGA was increased from 300 to 1800, the success rate reached 100 percent from 96.2 percent. As a result, AIGA even in small population size performs much better than TGA.

Table 1. Success rates (%) of AIGA for 2d Damavandi function.

Population	Number of Groups												
Size	#1	#2	#4	#6	#8	#9	#10	#12	#15	#16	#20	#25	Average
300	1.5	2.0	2.3	11.2	16.3	32.2	28.3	43.2	54.6	71.9	82.0	96.2	36.8
400	2.0	2.7	4.2	12.2	24.9	37.4	31.6	51.3	59.5	82.0	86.5	98.0	41.0
500	3.1	3.5	3.6	15.1	28.3	43.6	36.7	61.1	66.1	87.5	90.5	99.7	44.9
600	3.7	4.8	3.5	17.5	29.8	51.3	37.8	66.9	75.7	90.7	94.2	99.8	48.0
700	4.3	4.1	4.3	20.2	34.1	51.1	45.5	67.2	76.4	93.6	95.8	99.8	49.7
800	5.0	5.0	5.6	20.1	35.9	56.9	49.4	71.4	81.1	94.5	97.9	99.8	51.9
900	4.0	7.2	6.5	20.8	36.9	59.6	53.7	75.1	84.6	97.4	98.1	100.0	53.7
1000	5.7	7.0	7.8	23.1	43.1	66.5	52.1	78.1	86.5	97.3	98.8	100.0	55.5
1100	5.7	5.2	8.1	22.5	43.1	64.1	56.0	81.0	89.7	97.9	99.7	100.0	56.1
1200	6.1	7.4	8.0	21.2	45.7	65.0	57.6	83.2	91.8	99.1	99.4	100.0	57.0
1300	7.1	9.4	8.1	21.0	43.5	70.9	58.3	86.5	92.1	99.6	99.7	100.0	58.0
1400	8.6	9.8	12.7	24.4	47.0	73.7	62.3	87.8	93.3	99.4	99.7	100.0	59.9
1500	9.8	11.5	9.8	24.3	42.5	75.3	61.2	87.6	92.5	99.2	99.7	100.0	59.5
1600	10.3	9.8	11.9	23.0	45.7	72.0	62.0	88.8	93.6	99.5	99.9	100.0	59.7
1700	12.4	9.8	11.5	23.0	44.4	73.0	61.4	88.7	93.7	99.6	99.8	100.0	59.8
1800	11.3	10.8	14.2	23.5	42.8	73.7	64.5	89.3	94.9	99.4	99.9	100.0	60.4
Average	6.3	6.9	7.6	20.2	37.8	60.4	51.2	75.5	82.9	94.3	96.4	99.6	

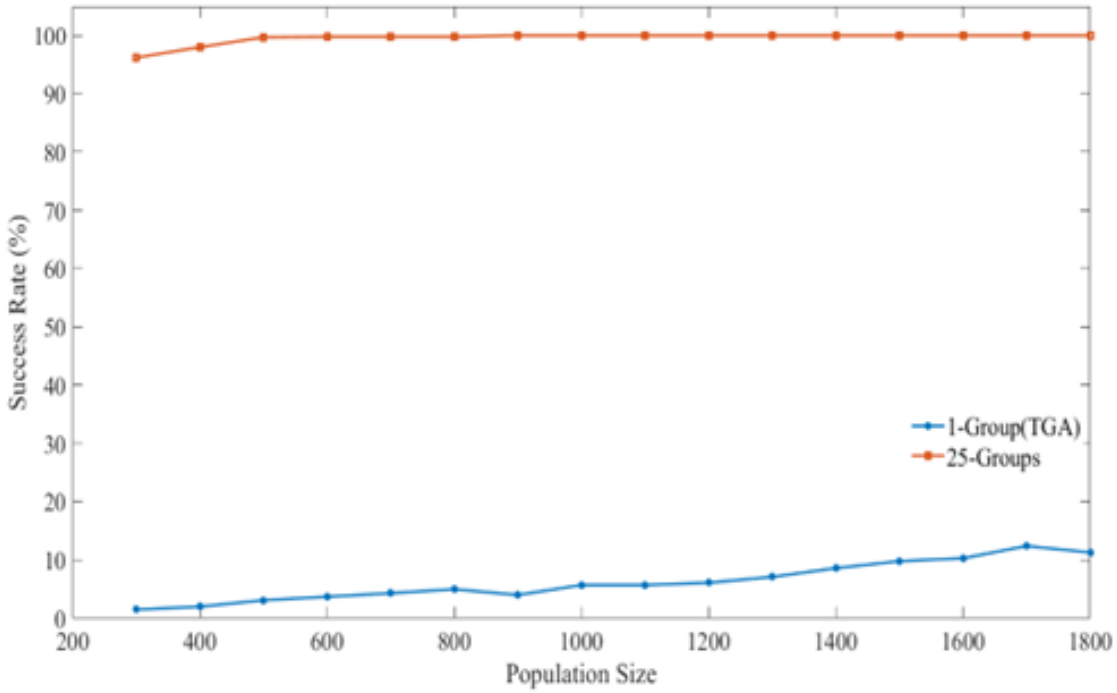


Figure 5. Success rates vs population size for TGA and best number of groups of AIGA.

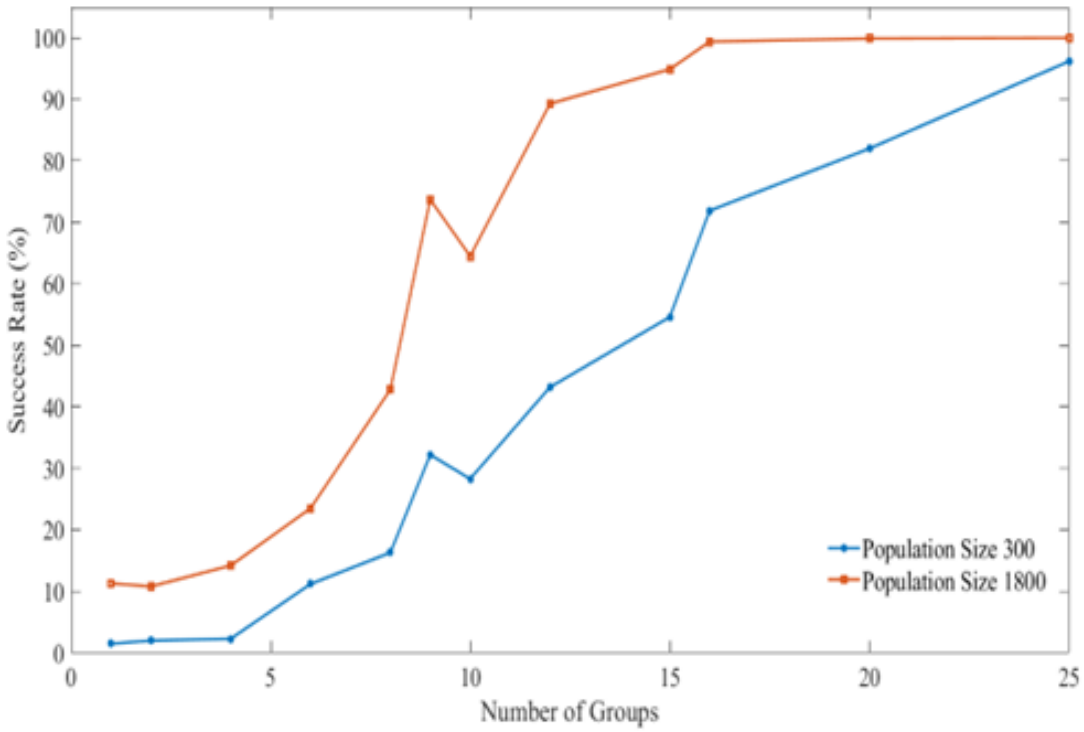


Figure 6. Success rates vs number of groups of AIGA for min. and max. population size.

In Figure 5, two lines on plot show success rates versus population size for the TGA (AIGA; 1-group) and AIGA; 25-groups. Success rates vs population size for TGA and different number of groups of AIGA were shown in subplots of Figure 2, before. Two lines on plot, one corresponds to TGA and other one corresponds to the best group AIGA, were selected to make comparison in Figure 5. Although AIGA distributes the population to different zones depending on the number of groups, it is more successful than TGA even with a small population size.

In Figure 6, two lines on plot show success rates versus the number of groups for the selected minimum and maximum population size as 300 and 1800, respectively. TGA (AIGA; 1-group) has been shown the lowest success rate when compared with all number of groups in both lines.

3.2 Case Study 2: Gaussian-Like Function

As the second case study, a 2-dimensional multimodal test function with different amplitudes and peaks derived from the Gaussian-Like function was used. Kilinc and Caicedo [22] derived the Gaussian function as an example optimization problem to obtain multiple optimal solutions. The equation for the test function is given in Equation 2.

$$f(x) = 4e^{\left(\frac{(-x_1^2-x_2^2)}{50}\right)} - 5e^{\left(\frac{(-x_1^2-x_2^2)}{8}\right)} + 2e^{(-2(x_1-3)^2-2(x_2+2)^2)} + 7e^{\left(\frac{-25(x_1+3)^2-25(x_2-2)^2}{2}\right)} - 3e^{\left(\frac{-25(x_1+3)^2-25(x_2-2)^2}{18}\right)} - 4e^{(-50(x_1-3)^2-50(x_2-2)^2)} \quad (2)$$

The function given in Equation 2 has two local minima with higher amplitude than the global minimum at the intervals of both variables [-5, 5]. It is known that the smaller the amplitude of the global minimum the harder it is to find. Therefore, the test function given in Equation 2 is very plausible for testing optimization algorithms. The graph of the test function is shown in Figure 7.

In this case study, the maximum allowed number of evaluations to achieve the global optimum was taken as 40,000. The population size of both TGA and AIGA were selected as 300, 600, 900, and 1200. In a similar manner, the number of groups as a parameter of AIGA were selected as 1, 4, 9, and 16. AIGA was run 400 times for each population size and the number of groups. The results obtained are presented by both in tabular form and the graphs as the success rate.

In Figure 8, each line on plot demonstrates success rates as percentage (y-axis) value versus the number of groups (x-axis) at which AIGA runs at different population sizes on the Gaussian Like function. Plots in Figure 8 shows that the success of AIGA is seen markedly as the number of groups increases different population sizes.

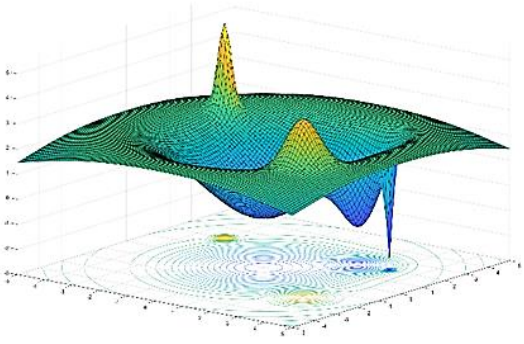


Figure 7. Gaussian-Like Function.

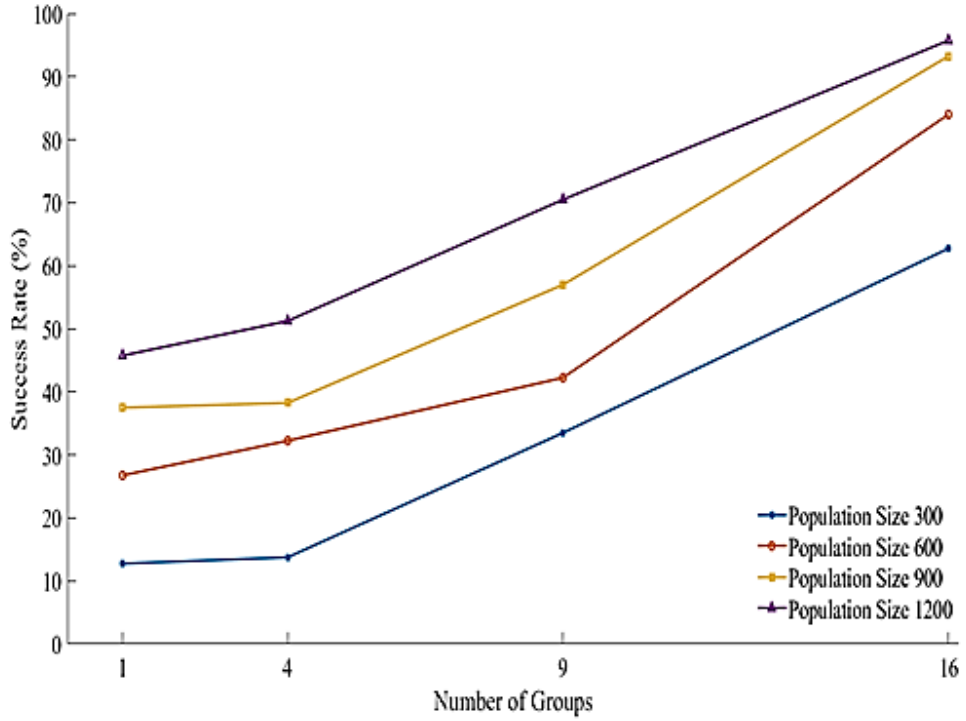


Figure 8. Success rates (%) vs number of groups for different population.

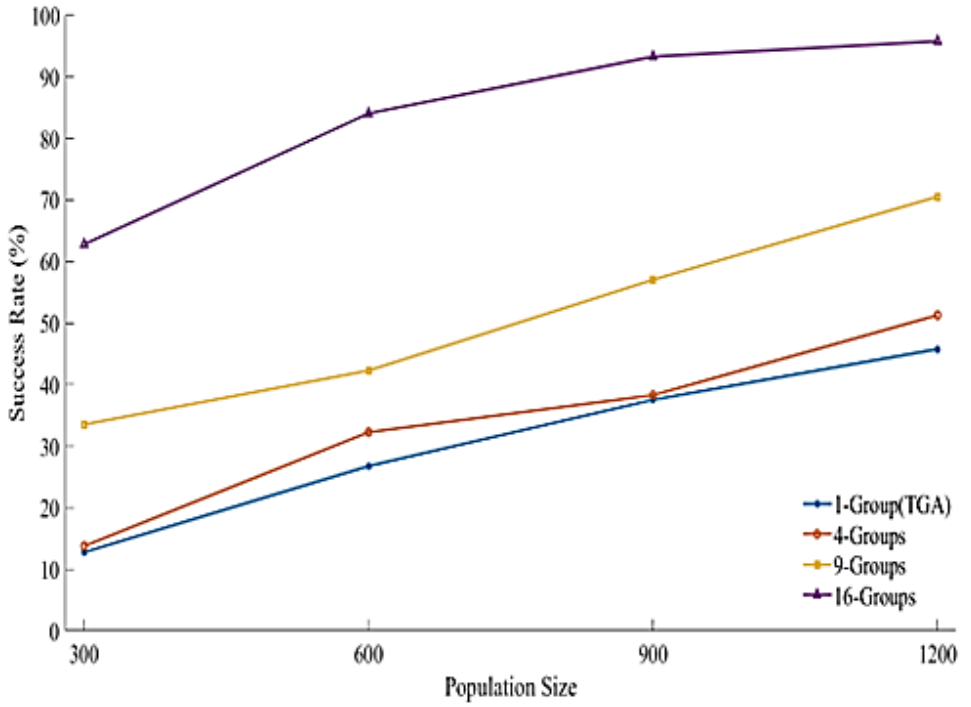


Figure 9. Success rtes (%) vs population size for different number of groups.

In Figure 9, each line on plot demonstrates success rates as percentage (y-axis) value versus the population size (x-axis) at which AIGA runs at different numbers of groups for Gaussian like function. Lines in Fig. 10 show that the success of both TGA (1-group AIGA) and AIGA is enhanced as population size increases for each group number. In the derived 2D gaussian function, the increase in population size was seen that improved the result. Nevertheless, an increase in the number of groups was found to be more effective than that of population size as concluded for Damavandi Function in case study 1. As a result, higher success rate can be achieved to obtain global optimum by using AIGA with the same population size instead of increasing the population size for TGA.

The success rates of AIGA in finding global optimum for the derived 2D Gaussian function dependent on both number of groups and population size are shown in Table 2. In terms of success rate, the increase in population size or number of groups or increase in both result with higher success rate of finding global optimum. While the average success rate for to the largest group number is 89.80 percent, the average success rate for to the largest population size is 44.60 percent. While the population

size for TGA was increased from 300 to 1200, the success rate reached 6.1 percent from 1.5 percent. Similarly, while the population size for Group 16 of AIGA was increased from 300 to 1200, the success rate reached almost 100 percent from 71.9 percent. Therefore, it is concluded that the increase in the number of groups is much more effective than the population size to increase the success rate. Based on the above discussion, it should be noted that 1-group corresponds to TGA, while n-group corresponds to AIGA which the search space is divided into n groups, thus using AIGA means better than using TGA to achieve finding global optima.

Table 2. Success rates (%) of AIGA for 2d Gaussian-like function.

Population Size	Number of Groups				
	#1	#4	#9	#16	Average
300	1.5	2.3	32.2	71.9	27.0
600	3.7	3.5	51.3	90.7	37.3
900	4.0	6.5	59.6	97.4	41.9
1200	6.1	8.0	65.0	99.1	44.6
Average	3.8	5.1	52.0	89.8	

4. Discussion

The performance of the AIGA was tested with widely recognized benchmark functions and compared with existing well-known algorithms such as Grey Wolf Optimizer (GWO) [7], Particle Swarm Optimization (PSO) [8], Firefly Algorithm [9], Bat Algorithm [10], Crow Search Algorithm [11], Cuckoo Search Algorithm [12], Flower Pollination Algorithm (FPA)[14]. Comparisons were carried out using functions with 2 and 30 dimensions having a significant search space size difference. For optimization, complex functions with different mathematical characteristics were selected. For instance, Damavandi function is continuous, differentiable, non-scalable and multimodal, Schwefel function is continuous, differentiable, scalable and has many local optima, Griewank function is continuous, differentiable, non-separable and has many local optima, Rosenbrock function is continuous, differentiable, non-separable and unimodal.

Firstly, Damavandi, Griewank, Schwefel, Rosenbrock, and Gaussian functions set to two dimensions shown in Table 3 were run 30 times with below algorithms.

The average and standard deviation of the obtained results for each function and algorithm were presented in the Table 3. Damavandi and Gaussian functions have been studied in detail above in the success of AIGA over TGA. In comparison to other algorithms, it is understood from the evaluation of average and standard deviation results (Table 4) that AIGA performs mostly better in capturing the global optimum.

Table 3. Benchmark functions used for 2-dimension experiments.

Function	Equation	dim	Range	f_{min}
Damavandi	$f(x)=\left[1-\left\lvert\frac{sin[\pi(x_1-2)]sin[\pi(x_2-2)]}{\pi^2(x_1-2)(x_2-2)}\right\rvert^5\right][2+(x_1-7)^2+(x_2-7)^2]$	2	[0,14]	0
Griewank	$f(x)=\sum_{i=1}^{d}\frac{x_i^2}{400}-\prod_{i=1}^{d}\cos\left(\frac{x_i}{\sqrt{i}}\right)+1$	2	[-600,600]	0
Schwefel	$f(x)=418.9829d-\sum_{i=1}^{d}x_i\sin\left(\sqrt{\lvert x_i\rvert}\right)$	2	[-512, 512]	0
Rosenbrock	$f(x)=\sum_{i=1}^{d-1}[100(x_{i+1}-x_i^2)^2+(x_i-1)^2]$	2	[-5,5]	0
Gaussian	$f(x)=4e^{\left(\frac{(-x_1^2-x_2^2)}{50}\right)}-5e^{\left(\frac{(-x_1^2-x_2^2)}{8}\right)}+2e^{(-2(x_1-3)^2-2(x_2+2)^2)}+7e^{\left(\frac{-25(x_1+3)^2-25(x_2-2)^2}{2}\right)}-3e^{\left(\frac{-25(x_1+3)^2-25(x_2-2)^2}{18}\right)}-4e^{(-50(x_1-3)^2-50(x_2-2)^2)}$	2	[-5,5]	0

Table 4. Comparison of algorithms for selected benchmark functions (dim = 2).

Function / Algorithm	Damavandi		Griewank		Schwefel		Rosenbrock		Gaussian	
	Avg	Std	Avg	Std	Avg	Std	Avg	Std	Avg	Std
AIGA	0.6667	0.9589	0.0000	0.0000	0.0000	0.0000	0.0093	0.0162	-2.6560	0.0000
GWO	1.8674	0.5047	0.0025	0.0035	3.9491	21.6237	0.0650	0.2179	-2.6008	0.3023
PSO	1.9333	0.3651	0.0003	0.0013	3.9480	21.6238	33.2036	140.0460	-2.6560	0.0000
Firefly	1.9552	0.2452	0.0019	0.0031	68.5925	46.6370	55.0196	74.5664	-1.5939	0.3217
BAT	1.9747	0.3066	0.3271	0.2342	24.9053	38.2936	642.8919	518.1514	-1.4556	0.3472
Crow S.	1.6735	0.6231	0.0003	0.0003	0.0000	0.0000	0.0096	0.0151	-2.6560	0.0000
Cuckoo S.	0.2000	0.6103	0.0009	0.0012	0.0002	0.0003	1.5223	1.3912	-2.6560	0.0001
FPA	0.4811	0.7417	0.0076	0.0036	0.2090	0.2028	8.3209	6.8515	-2.6278	0.0271

Lastly, Griewank, Schwefel, Rastrigin, Michalewicz and Styblinski-Tang functions set to thirty dimensions shown in the Table 5 were run 30 times with below algorithms as made above.

The mean and the variability of the obtained results for each function defined in a much larger search space, and for each algorithm are shown in Table 6. Likewise, AIGA outperformed over the other algorithms in case of

significantly larger (higher dimension) search space as well. The large differences in average values and standard deviation seen in Table 6 indicate that other algorithms are trapped more on local optima.

This proposed method is more effective in many engineering problems where the number of variables is high and a definitive result cannot be achieved, and in problems where there is a possibility of getting stuck in a local result. Planar and space truss structures optimization can be given as an example of concrete engineering problems in these studies.

Table 5. Benchmark functions used for 30-dimension experiments.

Function	Equation	dim	Range	f_{min}
Griewank	$f(x)=\sum_{i=1}^{d}\frac{x_i^2}{400}-\prod_{i=1}^{d}\cos\left(\frac{x_i}{\sqrt{i}}\right)+1$	30	[-600,600]	0
Schwefel	$f(x)=418.9829d-\sum_{i=1}^{d}x_i\sin\left(\sqrt{\lvert x_i\rvert}\right)$	30	[-512, 512]	0
Rastrigin	$f(x)=10d+\sum_{i=1}^{d}\left[x_i^2-10\cos(2\pi x_i)\right]$	30	[-5.12, 5.12]	0
Michalewicz	$f(x)=-\sum_{i=1}^{d}\sin(x_i)\sin^{2m}\left(\frac{ix_i^2}{\pi}\right)$	30	$[0,\pi]$	-1.8013
Styblinski-Tang	$f(x)=\frac{1}{2}\sum_{i=1}^{d}\left(x_i^4-16x_i^2+5x_i\right)$	30	[-5,5]	-39.16599d

Table 6. Comparison of algorithms for selected benchmark functions (dim = 30).

Function / Algorithm	Griewank		Schwefel		Rastrigin		Michalewicz		Styblinski-Tang	
	Avg	Std	Avg	Std	Avg	Std	Avg	Std	Avg	Std
AIGA	0.0090	0.0027	94.7817	99.0926	0.0117	0.0068	-29.5855	0.0322	-1174.9822	0.0009
GWO	0.0025	0.0047	5933.4899	795.9782	2.1056	3.3369	-17.1622	3.5241	-1017.4006	50.0187
PSO	0.0094	0.0103	5605.3116	665.5943	41.5892	16.2007	-26.6819	0.9912	-1035.9739	31.9540
Firefly	0.0298	0.0161	9236.3183	393.6007	18.7905	9.5962	-17.7670	2.0266	-750.5202	0.1449
BAT	185.3600	51.3961	4862.3753	4673.7892	319.2250	24.9510	-8.3810	0.7271	-669.7952	46.0701
Crow S.	0.8102	0.1110	5374.9974	684.3050	8.2851	6.5580	-19.2515	1.9559	-1012.5750	27.0560
Cuckoo S.	19.0544	3.0362	4273.5146	159.2521	146.1697	9.0925	-16.1591	0.5792	-969.3548	12.0875
FPA	155.0082	22.9206	6322.0616	186.1977	256.9353	9.6748	-11.7977	0.3635	-765.4912	18.3458

5. Conclusion

In this study, a new strategy proposed to be used for optimization algorithms. The new strategy has been tested on a genetic algorithm and aim was achieved by overcoming the problem of trapping the local optimum. In other words, the proposed method AIGA is not a new algorithm, rather is a new methodology that can be used in any optimization algorithm. Similar problems exist in other optimization algorithms such as trapping local optimum and premature convergence problems in genetic algorithms. By employing this suggested approach, exploration will persist across all regions within the search space, ensuring that the quest for the global optimum point extends to the majority of the population.

The performance of AIGA is verified using different optimization problems revealing the weakness of TGA in the case studies. For optimization problems, 2-dimensional multimodal Damavandi and Gaussian-Like functions having very small amplitude of global optimum and particularly local optima with larger amplitude have been selected. Finding the global optimum in stochastic optimization algorithms poses a significant challenge due to the low likelihood or difficulty in escaping local optima. The success of AIGA regardless of the precision of the result has been tested by observing if it reaches the global optimum in case studies. When validating case studies, consistency was maintained across all parameters and operators for both TGA and AIGA, with a fixed number of fitness function evaluations. In this context, detailed conclusions have been acquired as below:

In general, the probability of finding the global optimum increases as an increase in population size. In the proposed algorithm AIGA, even with fixed population size, the probability of reaching the global optimum significantly increases much more as the number of groups increases. It has been concluded that the increase in the number of groups is more effective than the increase in population size, thus AIGA is superior to TGA. Based on the results of this study, it can be stated that this approach can also be applied to improve other similar optimization algorithms.

Author contributions

Muslum Kilinc: Conceptualization, Software, Methodology, Validation, Writing-Original draft preparation. **Emrah Atilgan:** Methodology, Validation, Writing-Original draft preparation. **Cengiz Atis:** Validation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Atilgan, E., & Hu, J. (2018). First-principle-based computational doping of SrTiO 3 using combinatorial genetic algorithms. Bulletin of Materials Science, 41(1), 1. https://doi.org/10.1007/s12034-017-1515-9
2. S., V. C. S., & S., A. H. (2022). Nature inspired meta heuristic algorithms for optimization problems. Computing, 104(2), 251-269. https://doi.org/10.1007/s00607-021-00955-5
3. Fister Jr, I., Yang, X. S., Fister, I., Brest, J., & Fister, D. (2013). A brief review of nature-inspired algorithms for optimization. Neural and Evolutionary Computing, 80(3), 116-122. https://doi.org/10.48550/arXiv.1307.4186
4. Darwish, A. (2018). Bio-inspired computing: Algorithms review, deep analysis, and the scope of applications. Future Computing and Informatics Journal, 3(2), 231-246. https://doi.org/10.1016/j.fcij.2018.06.001
5. Yang, X. S. (2020). Nature-inspired optimization algorithms: Challenges and open problems. Journal of Computational Science, 46, 101104. https://doi.org/10.1016/j.jocs.2020.101104
6. Stork, J., Eiben, A. E., & Bartz-Beielstein, T. (2022). A new taxonomy of global optimization algorithms. Natural Computing, 21(2), 219-242. https://doi.org/10.1007/s11047-020-09820-4
7. Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. Advances in Engineering Software, 69, 46-61. https://doi.org/10.1016/j.advengsoft.2013.12.007
8. Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In Proceedings of ICNN'95-International Conference on Neural Networks, 4, 1942-1948. https://doi.org/10.1109/ICNN.1995.488968
9. Yang, X. S. (2010). Nature-inspired metaheuristic algorithms. Luniver press.
10. Yang, X. S., & Hossein Gandomi, A. (2012). Bat algorithm: a novel approach for global engineering optimization. Engineering Computations, 29(5), 464-483. https://doi.org/10.1108/02644401211235834
11. Askarzadeh, A. (2016). A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm. Computers & Structures, 169, 1-12. https://doi.org/10.1016/j.compstruc.2016.03.001
12. Yang, X. S., & Deb, S. (2009). Cuckoo search via Lévy flights. In 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), 210-214. https://doi.org/10.1109/NABIC.2009.5393690
13. Chechkin, A. V., Gonchar, V. Y., Klafter, J., & Metzler, R. (2006). Fundamentals of Lévy flight processes. Fractals, Diffusion, and Relaxation in Disordered Complex Systems: Advances in Chemical Physics, Part B, 439-496.
14. Yang, X. S. (2012). Flower pollination algorithm for global optimization. In International Conference on Unconventional Computing and Natural Computation, 7445, 240-249. https://doi.org/10.1007/978-3-642-32894-7_27
15. Rocha, M., & Neves, J. (1999). Preventing premature convergence to local optima in genetic algorithms via random offspring generation. In Multiple Approaches to Intelligent Systems: 12th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems IEA/AIE-99, Cairo, Egypt, May 31-June 3, 1999. Proceedings 12, 127-136. https://doi.org/10.1007/978-3-540-48765-4_16
16. Dang, D. C., Friedrich, T., Kötzing, T., Krejca, M. S., Lehre, P. K., Oliveto, P. S., ... & Sutton, A. M. (2017). Escaping local optima using crossover with emergent diversity. IEEE Transactions on Evolutionary Computation, 22(3), 484-497. https://doi.org/10.1109/TEVC.2017.2724201
17. Doerr, B. (2020). Does comma selection help to cope with local optima?. In Proceedings of the 2020 Genetic and Evolutionary Computation Conference, 1304-1313. https://doi.org/10.1145/3377930.3389823
18. Oliveto, P. S., Paixão, T., Pérez Heredia, J., Sudholt, D., & Trubenová, B. (2018). How to escape local optima in black box optimisation: when non-elitism outperforms elitism. Algorithmica, 80, 1604-1633. https://doi.org/10.1007/s00453-017-0369-2
19. Sharma, P., & Raju, S. (2024). Metaheuristic optimization algorithms: A comprehensive overview and classification of benchmark test functions. Soft Computing, 28(4), 3123-3186. https://doi.org/10.1007/s00500-023-09276-5
20. Cheng, R., Li, M., Tian, Y., Xiang, X., Zhang, X., Yang, S., ... & Yao, X. (2018). Benchmark functions for the cec'2018 competition on many-objective optimization.
21. Deb, L. (1993). Multimodal deceptive functions. Complex Systems, 7, 131-153.
22. Kilinc, M., & Caicedo, J. M. (2019). Finding plausible optimal solutions in engineering problems using an adaptive genetic algorithm. Advances in Civil Engineering, 2019(1), 7475156. https://doi.org/10.1155/2019/7475156

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Effect of dimension reduction with PCA and machine learning algorithms on diabetes diagnosis performance

Yavuz Bahadır Koca [*1], **Elif Aktepe** [2]

[1]*Afyon Kocatepe University, Department of Electrical Engineering, Türkiye, ybkoca@gmail.com*
[2]*Afyon Kocatepe University, Department of Electronics and Automation, Türkiye, eaktepe@aku.edu.tr*

Cite this study: Koca, Y. B., & Aktepe, E. (2024). Effect of dimension reduction with PCA and machine learning algorithms on diabetes diagnosis performance. Turkish Journal of Engineering, 8 (3), 447-456

https://doi.org/10.31127/tuje.1413087

Keywords
Diabetes
Data analysis
Machine learning
PCA
Random forest

Research Article

Received: 01.01.2024
Revised: 23.01.2024
Accepted: 29.01.2024
Published:05.07.2024

Abstract

Diabetes, a long-term metabolic disorder, causes persistently high blood sugar and presents a significant global health challenge. Early diagnosis is of vital importance in mitigating the effects of diabetes. This study aims to investigate diabetes diagnosis and risk prediction using a comprehensive diabetes dataset created in 2023. The dataset contains clinical and anthropometric data of patients. Data simplification was successfully applied to clean unnecessary information and reduce data dimensionality. Additionally, methods like Principal Component Analysis were applied to decrease the number of variables in the dataset. These analyses rendered the dataset more manageable and improved its performance. In this study, a dataset encompassing health data of a total of 100,000 individuals was utilized. This dataset consists of 8 input features and 1 output feature. The primary objective is to determine the algorithm that exhibits the best performance for diabetes diagnosis. There was no missing data during the data preprocessing stage, and the necessary transformations were carried out successfully. Nine different machine learning algorithms were applied to the dataset in this study. Each algorithm employed various modelling approaches to evaluate its performance in diagnosing diabetes. The results demonstrate that machine learning models are successful in predicting the presence of diabetes and the risk of developing it in healthy individuals. Particularly, the random forest model provided superior results across all performance metrics. This study provides significant findings that can shed light on future research in diabetes diagnosis and risk prediction. Dimensionality reduction techniques have proven to be valuable in data analysis and have highlighted the potential to facilitate diabetes diagnosis, thereby enhancing the quality of life for patients.

1. Introduction

High blood sugar, the most important characteristic marker of diabetes causes severe damage, especially to kidneys, eyes, nerves and heart [1,2]. Type 2 diabetes, usually seen in adults, occurs when the body becomes resistant to insulin or cannot produce enough insulin. Presently, prevalence of type 2 diabetes has increased significantly [1,3]. Approximately 422 million people worldwide have diabetes, and the majority live in low- and middle-income countries [4]. Every year, 1.5 million people die due to diabetes-related causes. Diabetes cases and prevalence have been increasing steadily in recent years [5]. For people living with diabetes, access to affordable treatment and medications such as insulin is vital. There is a global target to stop the increase in diabetes and obesity by 2025 [6,7].

The key to living a good life with diabetes is early diagnosis. People living with undiagnosed and untreated diabetes often have worsened health outcomes. Therefore, it is important that basic diagnostics, such as blood sugar testing are easily accessible. By analysing diabetes related data Machine Learning (ML) can offer advanced methods for early diagnosis [8]. Additionally, by using the knowledge that patients with diabetes need periodic expert evaluations and treatment a ML supported system can play an important role in monitoring and managing patients. In recent years, ML based approaches have come to the fore in many health and medical application fields.

Different approaches are employed in ML to process data and derive meaningful insights. These ae supervised, unsupervised, semi-supervised and reinforcement. Supervised learning involves training

models using labelled datasets, associating input features with specific outcomes. For instance, a dataset could use individual attributes like weight and height to predict the onset of health conditions such as diabetes. On the other hand, unsupervised learning focuses on pattern recognition and exploration without predefined target variables. In this context, it is effective for identification to the detection of new disease mechanisms, genotypic variations or phenotypic patterns. Semi-supervised learning utilizes datasets containing both labelled and unlabelled information. It aims to strike a balance between the two. This approach is particularly effective in heterogeneous data structures, leveraging the advantages of both supervised and unsupervised techniques. Reinforcement learning stands apart by combining concepts from both supervised and unsupervised learning methodologies. And it is based on trial-error exploration rather than solely relying on structured data, reflects the human learning experience. Through iterative learning and feedback mechanisms, reinforcement learning optimizes decisions and actions to achieve desired outcomes. This process results in outputs closely resembling human learning through adaptation and interaction.

Machine learning offers an approach to creating and evaluating learning ability based on data obtained regarding diabetes [9]. In recent years, the use of ML algorithms for analysis and prediction in the field of healthcare has attracted great attention among researchers. ML algorithms increase diagnostic accuracy thanks to their ability to handle large data sets by combining information from various data sources. These algorithms have the potential to provide more accurate, faster and more economical results in health diagnosis [8]. With the combination of wearable technologies and embedded systems, real-time monitoring opportunities in the field of healthcare have increased significantly. In this way, human health parameters can be monitored instantly instead of periodic follow-ups. The integration of wearable technologies and management of diseases such as blood pressure and diabetes offer the opportunity to provide a more comfortable life for individuals. For the field of diabetes, instantaneous changes in glucose levels and other health parameters have become available with wearable sensors placed on the body instead of blood tests [10,11]. These applications offer significant opportunities to improve patients' quality of life by using artificial intelligence techniques such as ML.

Many studies on diabetes diagnosis are carried out to predict the presence of progression of the disease using the obtained data sets. These studies provide important insights into diabetes diagnosis using data analytics methods and ML algorithms.

Mujumdar and Vahidehi [12] conducted a study using various machine learning algorithms for diabetes diagnosis. In this study, a large data set was analysed and predictions were made for the diagnosis of diabetes using patients' clinical and anthropometric data. Researchers have determined that the logistic regression model achieved a high accuracy rate of 96% in diagnosing diabetes.

Kopitar et al. [13] compared regression models, which are commonly used in the prediction of undiagnosed Type 2 diabetes and ML-based prediction models (LightGBM, XGBoost, RF, Glment). Researchers have observed that no clinically meaningful improvement is achieved when more complex prediction models are used. Also, Kumar et al. [14] employed an unsupervised learning approach, the Deep Neural Network (DNN) classifier, for diagnosing type 2 diabetes. Additionally, they utilized a feature importance model packaged with extra trees and random forest for feature selection. Another study aimed to create a prediction model to better identify individuals at risk of diabetes. Predictive models have been developed using ML techniques such as Gradient Boosting Machine (GBM) and Logistic Regression (LR). Additionally in the study, the discriminatory ability of the models was measured by evaluating the area under the Receiver Operating Characteristic Curve (AROC). Researchers have emphasized that GBM and LogR models exhibit superior performance than Random Forest (RF) and Decision Tree (DT) models [15].

Sowah et al. [1] combined multiple artificial intelligence algorithms to address various factors affecting the health status of individuals with diabetes. Soni and Varma [16], various ML algorithms such as SVM, KNN, RF, DT, LR and GBM were used. The results showed that the RF classifier had the highest performance with 77% classification accuracy. Tasin et al. [17] studied machine learning classification models in the training and testing stages. The results revealed that the XGBoost classifier with the ADASYN approach exhibited the highest performance with an accuracy rate of 81%. In other studies, using different machine learning methods on diabetes, researchers have conducted studies on advanced prediction and approaches [18–22].

This study uses various machine learning algorithms by performing data simplification to predict the presence of diabetes and the risk of developing it in healthy individuals in the future, using 100,000 patient data taken from the Kaggle database. The dataset includes patients' clinical and anthropometric data. Data simplification was carried out to reduce the size of the dataset and purify it from unnecessary information. Later, various machine learning algorithms such as LogR, GBM, RF, DT and SVM were used to predict the presence of diabetes and the risk of developing it in the future in healthy individuals. The results obtained show that machine learning models are successful in predicting the presence of diabetes and the risk of developing it in the future in healthy individuals. In particularly, it was determined that the random forest classifier exhibited the highest performance. These findings reveal the potential to improve patients' quality of life by increasing early diagnosis and intervention opportunities for diabetes.

The results highlight the significant opportunities provided by using artificial intelligence techniques on diabetes and provide healthcare professionals with valuable information on the diagnosis of diabetes and its likelihood of progression. Analyses and prediction models show that it can be an effective tool in the management and treatment of diabetic patients.

2. Materials and Method

Algorithms used in fields such as machine learning and artificial intelligence perform predictive modelling processes to forecast future outcomes by utilizing data and statistics. Diabetes commonly presents with abnormal metabolism and elevated blood sugar levels. It can lead to specific complications in different body parts such as the eyes, kidneys, and nervous system. Such symptoms can be employed for data collection purposes, followed by a modelling process based on factors such as age and gender. Data analytics plays a significant role in the management and diagnosis of chronic diseases like diabetes within healthcare. For instance, medical records of past diabetes patients, data on blood sugar levels, treatment methods, and age can be utilized to predict future diabetes risk. By processing these data with machine learning algorithms, personalized risk profiles can be constructed, enabling earlier diagnosis and the development of more effective treatment plans for patients. Therefore, in healthcare issues such as diabetes diagnosis and management, data analytics and predictive modelling serve as powerful tools to enhance patient care and treatment processes. These data play a critical role in the development of algorithms used to achieve improved health outcomes and enhance the quality of life for patients.

2.1. Dataset and data preprocessing

In this study, the diabetes dataset used was obtained from the Kaggle data sharing platform [23]. The dataset contains a total of 100,000 records belonging to individuals. Out of these records, 8,500 represent diabetes patients, while 91,500 do not have diabetes. The dataset consists of a total of 9 features that include symptoms related to diabetes. Eight of these features are used as input parameters and one is the output parameter. The attributes and their values for the dataset are detailed in Table 1.

During the preprocessing stage of the dataset, checks were conducted to identify any missing, scattered or mis-defined data. However, no missing data was detected in the dataset. The necessary transformations to convert the data into a processable form were carried out based on the information provided in Table 1. In the subsequent stage, 20% of the dataset were set aside to evaluate the model's performance. The remaining data points were utilized to train the model. This division of the data allowed for the assessment of the model's accuracy and generalization ability.

Table 1. Attributes and values in the dataset.

No	Attribute	Values
1	Gender	0 – Male 1 – Female 2 - Other
2	Age	[0-80]
3	Hypertension	0-No hypertension 1- There is hypertension
4	Heart disease	0- No heart disease 1- There is heart disease
5	Smoking history	.....
6	Body mass index	[10-95,7]
7	HbA1c level	[3,5-9]
8	Blood glucose level	[80-300]
9	Diabetes	0-No diabetes 1-Have diabetes

The statistics for the features in the dataset are presented in Table 2. This table provides statistical summaries of different attributes in a dataset, including measures such as mean, standard deviation, minimum, and maximum values.

Figure 1 displays the correlation matrix, which is employed to assess the relationship between diabetes and other parameters. Correlation coefficients depict the direction and strength of a linear relationship between two variables. A positive correlation signifies a scenario where both variables increase or decrease concurrently. At the same time, a negative correlation indicates that one variable increases as the other decreases.

2.2. Machine learning algorithms

Machine learning offers various approaches and solution methods for different types of problems. The selection of the most suitable algorithm for each problem type depends on the structure of the dataset and the nature of the intended solution. The algorithms provided in Figure 2 offer different approaches for tasks such as data classification, regression, clustering, dimensionality reduction, and other types of problems.

This study was developed using the Python programming language within the Anaconda platform, specifically in the Spyder environment. Scikit-learn and TensorFlow libraries were utilized for programming. A comprehensive analysis was conducted for the diagnosis of diabetes, involving nine different supervised machine

Table 2. Dataset analysis.

Characteristic	Average	Standard deviation	Minimum Value	Maximum Value
Gender	0.59	0.49	0.00	2.00
Age	41.89	22.52	0.08	80.00
Hypertension	0.07	0.26	0.00	1.00
Heart disease	0.04	0.19	0.00	1.00
Smoking history	1.39	1.59	0.00	5.00
Body mass index	27.32	6.64	10.01	95.69
HbA1c level	5.53	1.07	3.50	9.00
Blood glucose level	138.06	40.71	80.00	300.00
Diabetes	0.09	0.28	0.00	1.00

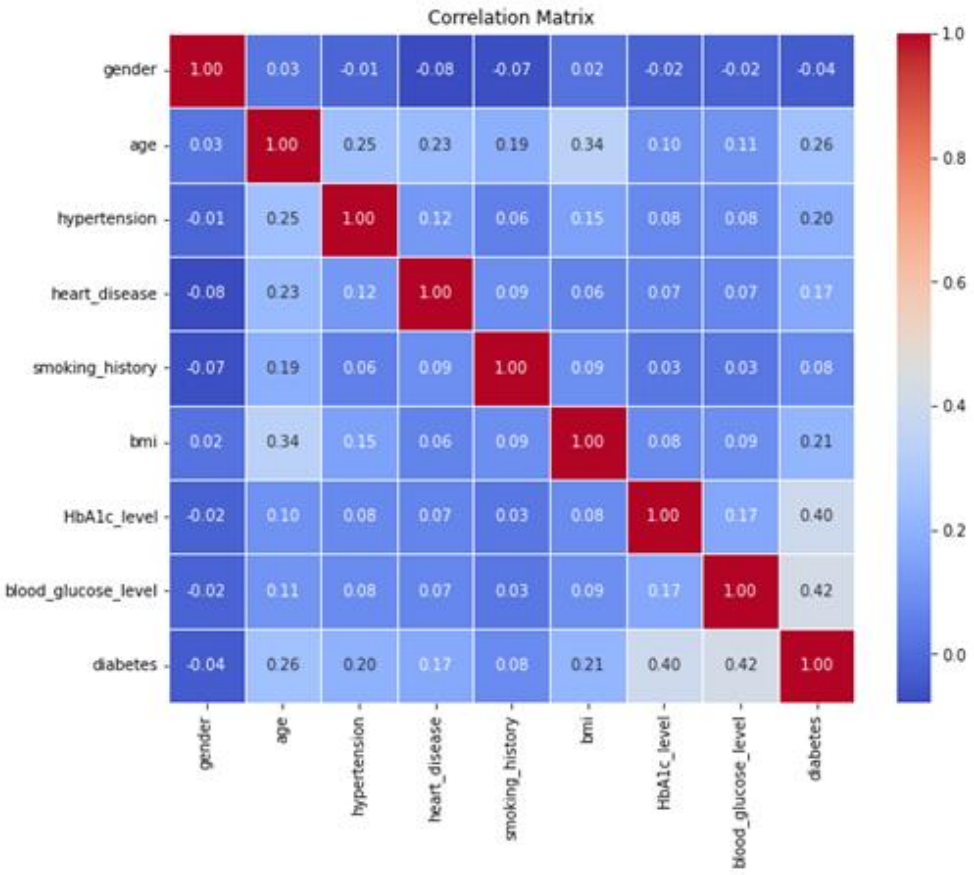


Figure 1. Correlation matrix.

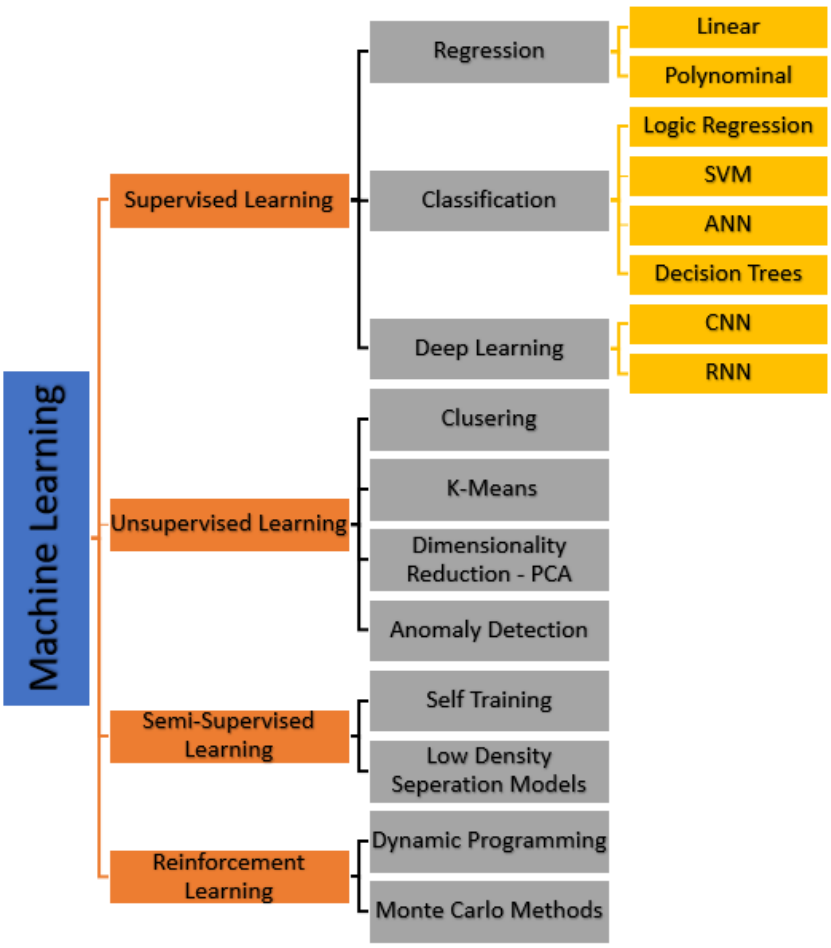


Figure 2. Machine learning algorithms tree.

learning algorithms along with the results of Principal Component Analysis (PCA). The purpose of this analysis was to enhance the accuracy of diabetes diagnosis and determine the best-performing algorithm. The process employed for the development and evaluation of predictive models is illustrated in Figure 3.

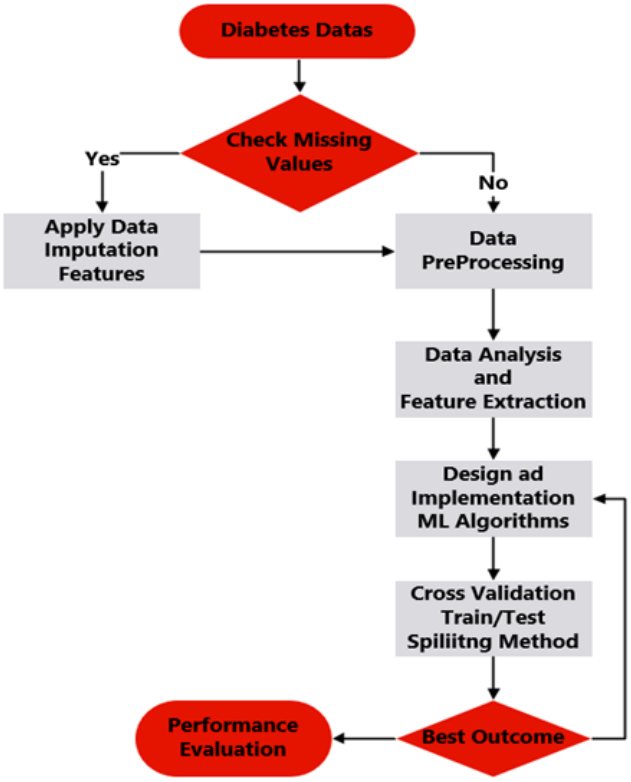


Figure 3. Algorithms process.

The dataset was analysed using various machine learning algorithms, including Linear Regression (LR), Polynomial Regression (PR), Logistic Regression (LogR), Artificial Neural Networks (ANN), Support Vector Machines (SVM), Decision Trees (DT), K-Nearest Neighbors (K-NN), Random Forest (RF), and Recurrent Neural Networks (RNN) for diabetes diagnosis. Each algorithm learned from the dataset and made predictions to evaluate its performance in diagnosing diabetes.

PCA (Principal Component Analysis) method was applied to reduce the dimensionality of the dataset, and the performance of algorithms was measured by experimenting with different numbers of components. This helped simplify the dataset, making it more manageable. Changes in the performance of algorithms with different numbers of components from PCA were analysed in detail. The results underwent a comprehensive evaluation to compare the performance of each algorithm and PCA with different numbers of components. This analysis determined which algorithm performed best and provided the most accurate predictions for diabetes diagnosis under different component scenarios.

2.2.1. Linear regression

Linear Regression (LR), predicting the dependent variable when the values of independent variables are given. Using the LR model, data is modelled with the best-fitting linear line [24,25]. LR model named *lr_model* is defined using the *LinearRegression()* function this dataset. The model is trained with training data using the *fit()* function. Then, the trained model makes predictions on the test data, and these predictions are rounded to 0 or 1. The accuracy value is calculated by comparing the rounded predictions with the actual class labels.

2.2.2. Polynomial regression

Polynomial Regression (PR) expresses the relationship between dependent and independent variables with a non-linear, polynomial equation [26]. PR provides more flexibility and can better fit the true structure of the data. Polynomial features were created by increasing the degree of the data in this study. The data is split into training and test data. Then, a PR model is created and fitted to the training data. When training is completed, predictions are made on the test data. The predictions are rounded to the nearest integer, and the accuracy value is calculated by comparing these rounded predictions with the actual test data.

2.2.3. Logistic regression

Logistic Regression (LogR) is a technique applied to classification problems. This method attempts to express the correlation between dependent and independent variables with a curve that is similar to a straight line. This line-like curve is used to predict the categorical values of the dependent variable. Independent variables are measured with binary or multiple values [19]. A LogR model is defined using the *LogisticRegression()* function in this study. The model is trained with the training data. Predictions are made, and the accuracy value is calculated by comparing these predictions with the actual class labels.

2.2.4. Support vector machines

Support Vector Machines (SVMs) are a supervised machine learning algorithm that can be used for classification or regression tasks. SVMs work by finding a hyperplane that separates two classes of data points in a high-dimensional space. The hyperplane is chosen to maximize the margin between the two classes, which helps to ensure that the model is robust to noise and outliers [16]. In this study, an SVM model was trained on a dataset of labeled data. The model was then used to make predictions on a test dataset. The accuracy of the model was evaluated by comparing the predicted labels to the true labels.

2.2.5. Artificial neural networks

Artificial Neural Networks (ANN) are known for their ability to learn complex data structures and relationships. An artificial neural network consists of layers organized in a structured manner [27]. The first layer takes input data, while the last produces predictions or classification results. The hidden layers in between process the data to capture higher-level features. In this study, the model consists of two layers.

In this architecture, data from the input layer is processed in the first layer, and the output is obtained as a probability value between 0 and 1 at the output layer. Training is done for ten epochs (cycles) and in each epoch. Once training is complete, the model makes predictions on the test data. These predictions are then rounded to obtain 0 or 1 values. After making predictions, the accuracy value is calculated by comparing them with the actual class labels.

2.2.6. Decision trees

Decision Trees (DT) start with a root node initially. This root node splits the data into child nodes under specific conditions. These child nodes, in turn, further split into sub-nodes based on conditions and decisions. This process continues until it reaches leaf nodes, where the data is classified or predicted [27,28] . Decision Trees are highly effective in understanding patterns and relationships in a dataset. Each node has the ability to classify or make predictions by evaluating data features. In this study, the model is trained with the training data. After making predictions, the accuracy value is calculated by comparing these predictions with the actual class labels.

2.2.7. K-Nearest neighbors

K-Nearest Neighbors (K-NN) algorithm's fundamental principle is to determine the class or value of a data point by using the majority of its closest neighbors. Data points are positioned in space, and when a new point arrives, its closest neighbors are examined. The K-NN algorithm makes predictions using the labels or values of these neighbors [16]. The model is trained with the training data. After making predictions, the accuracy value is calculated by comparing these predictions with the actual class labels.

2.2.8. Random forest

Random Forest (RF) is a powerful machine learning algorithm used for classification and regression problems. This algorithm is constructed by combining multiple decision trees for accurate predictions and classifications. It achieves this through a voting process where each tree votes based on its own analysis and the final result reflects the consensus of the majority [17,29]. Each decision tree is trained with different subsets of data and random feature selection. This allows each tree to learn from different features and data points. The RF algorithm aims to obtain more reliable and stable results by combining the predictions of these trees. In this study, a Random Forest model is defined using *RandomForestClassifier()*. The model is trained with the training data. After making predictions, the accuracy value is calculated by comparing these predictions with the actual class labels.

2.2.9. Recurrent neural network

Recurrent Neural Network (RNN) is a type of artificial neural network designed to process data such as time series and sequential data. RNNs have the ability to internally store information from previous time steps, making them suitable for the analysis of sequential data. In this study, an RNN model was employed to process time series data [30]. The model is trained with training data. During training, the model utilizes information from previous time steps to understand and learn the temporal dependencies within the data. Once training is completed, the model makes predictions on test data. These predictions are then compared to the actual class labels, and an accuracy value is calculated. This process is employed to assess how the model processes and predicts time series data.

2.2.10. Principal component analysis

Principal Component Analysis (PCA) is a statistical technique employed for dimensionality reduction in multi-dimensional datasets, aiming to encapsulate the core information within the data [31]. Its fundamental aim is to express the data with a reduced number of eigenvalue components, simplifying the dataset. The main idea of PCA is to create new transformed principal components. It aims maximize the variance in the dataset. These new components are linear combinations of the original variables. PCA achieves this by selecting the components that carry the highest variance, effectively filtering out unnecessary or low-variance variables, thereby expressing the dataset in a lower dimension for improved interpretation and analysis. PCA finds applications in data compression, visualization, noise reduction, and enhancing the performance of certain models. In this study, the impact of PCA is investigated by selecting different numbers of components (ranging from 8 to 1). Consequently, the dimensionality of the dataset is sequentially reduced from 8 to 1, and the outcomes of this dimension reduction process are evaluated. The experiments conducted with various numbers of components aim to provide a detailed analysis of PCA's effect on the dataset.

3. Results

The study evaluated nine supervised ML algorithms. Evaluating their accuracy, precision, recall and F1 scores. According to Table 3, most models demonstrated notably high levels of these metrics when applied to classify the dataset. These results indicate that the machine learning algorithms used successfully classified symptoms related to diabetes in the dataset.

The high values of F1 score, recall, accuracy and precision indicate that the models were effective in identifying and classifying cases related to diabetes. This suggests that the selected algorithms, such as LogR, PR, SVM, ANN, DT, KNN, RF and RNN all showed promising performance in diabetes diagnosis.

Furthermore, it's important to note that these algorithms were evaluated using various performance metrics to ensure a comprehensive assessment of their effectiveness in diabetes diagnosis. The combination of these metrics provides a well-rounded view of each algorithm's performance, highlighting their strengths and suitability for different aspects of diabetes diagnosis.

Table 3. Performance evaluation of machine learning algorithms.

ML Algorithms	Accuracy	Precision	Recall	F1 Score
LR	0,938	0,941	0,938	0,922
PR	0,952	0,955	0,952	0,944
LogR	0,952	0,948	0,952	0,947
SVM	0,947	0,950	0,947	0,936
ANN	0,959	0,959	0,959	0,955
DT	0,951	0,952	0,951	0,952
KNN	0,952	0,950	0,952	0,947
RF	0,969	0,969	0,969	0,968
RNN	0,963	0,963	0,963	0,959

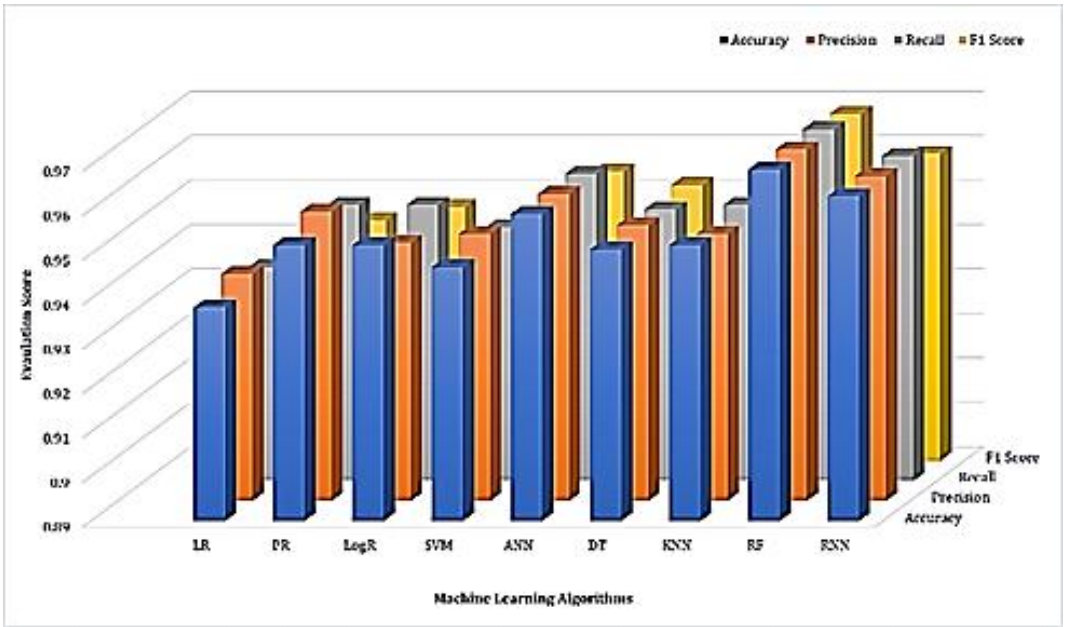


Figure 4. Machine learning algorithms performance.

Figure 4 shows the performance analysis of machine learning algorithms.

In conclusion, the study's findings demonstrate that the application of supervised machine learning algorithms, along with the evaluation of multiple performance metrics can significantly contribute to accurate diabetes diagnosis. This has the potential to improve understanding in the field and enhance access to correct diagnosis and treatment for patients.

Especially the RF model has demonstrated the highest performance in all metrics. This means that this model achieved high values not only in accuracy but also in precision, recall, and F1 score, indicating both a high rate of correct classification and a high rate of correctly classified positive cases. Additionally, it outperformed in terms of the rate of detecting all true positives (recall) and F1 score. This indicates that the RF algorithm can be effectively used in critical medical conditions such as diabetes diagnosis. However, the performance of other algorithms is also notably high. For example, ANN stands out with its high accuracy and classification success. Similarly, DT and KNN algorithms have also yielded good results. These results emphasize the effectiveness of various machine learning algorithms in diagnosing diabetes. The choice of the most suitable algorithm may depend on specific requirements, such as the balance between precision and recall or computational efficiency. Therefore, this study provides valuable insights into the selection and application of machine learning algorithms in the field of medical diagnosis, particularly for diabetes.

Table 4 displays the test accuracies achieved by each model on the PCA-transformed dataset for different numbers of components. Generally, PCA is used with the aim of dimensionality reduction, intending to represent the dataset in a lower dimension. This may lead to some information loss, but it can reduce noise in the dataset and potentially improve the performance of some models. As seen in Table 4, as the number of components decreases, the accuracy value generally tends to decrease. This indicates that fewer components result in more significant information loss. However, in some cases, reducing very high-dimensional datasets to smaller dimensions can accelerate model training and help with better generalization. Therefore, the choice of the number of components to use should be made carefully, depending on the specific application and dataset.

The accuracy rate of machine learning algorithms given in Table 4 increases as the number of components increases, as shown in the graph in Figure 5. When working with the real dataset with 8 components, the accuracy values on the PCA-transformed dataset are generally low or similar. However, the PCA transformation for the RNN model produces results that are almost the same as the real dataset's accuracy. When dimensionality reduction is performed with PCA for component numbers 7, 6, 5, 4, and 3, the accuracy values increase or remain almost the same for most models. In this case, it has been observed that performance can be improved by reducing the complexity of the model and

representing the dataset in a lower dimension. When reducing the number of components to 2 or lower levels, the performance of the models significantly decreases. This leads to the loss of important features of the dataset and makes it difficult for models to learn the data effectively.

In conclusion, when PCA transformation is used by carefully selecting the right number of components, it can improve the performance of some models and reduce computation time by representing the dataset in a lower dimension. However, the choice of the correct number of components should be made carefully, depending on the structure of the dataset and model performance.

Table 4. Effect of component number on accuracy.

Number of Components	LR	PR	LogR	SVM	ANN	DT	KNN	RF	RNN
Real Data (n=8)	0,938	0,952	0,952	0,947	0,953	0,951	0,952	0,970	0,966
PCA (n=7)	0,937	0,952	0,958	0,960	0,967	0,953	0,952	0,969	0,971
PCA (n=6)	0,936	0,952	0,957	0,960	0,968	0,953	0,952	0,96	0,970
PCA (n=5)	0,936	0,952	0,958	0,960	0,967	0,953	0,952	0,968	0,968
PCA (n=4)	0,929	0,952	0,939	0,948	0,945	0,923	0,941	0,944	0,947
PCA (n=3)	0,929	0,952	0,939	0,948	0,947	0,921	0,941	0,941	0,947
PCA (n=2)	0,928	0,952	0,938	0,948	0,947	0,923	0,943	0,938	0,948
PCA (n=1)	0,928	0,952	0,940	0,947	0,947	0,912	0,943	0,912	0,944

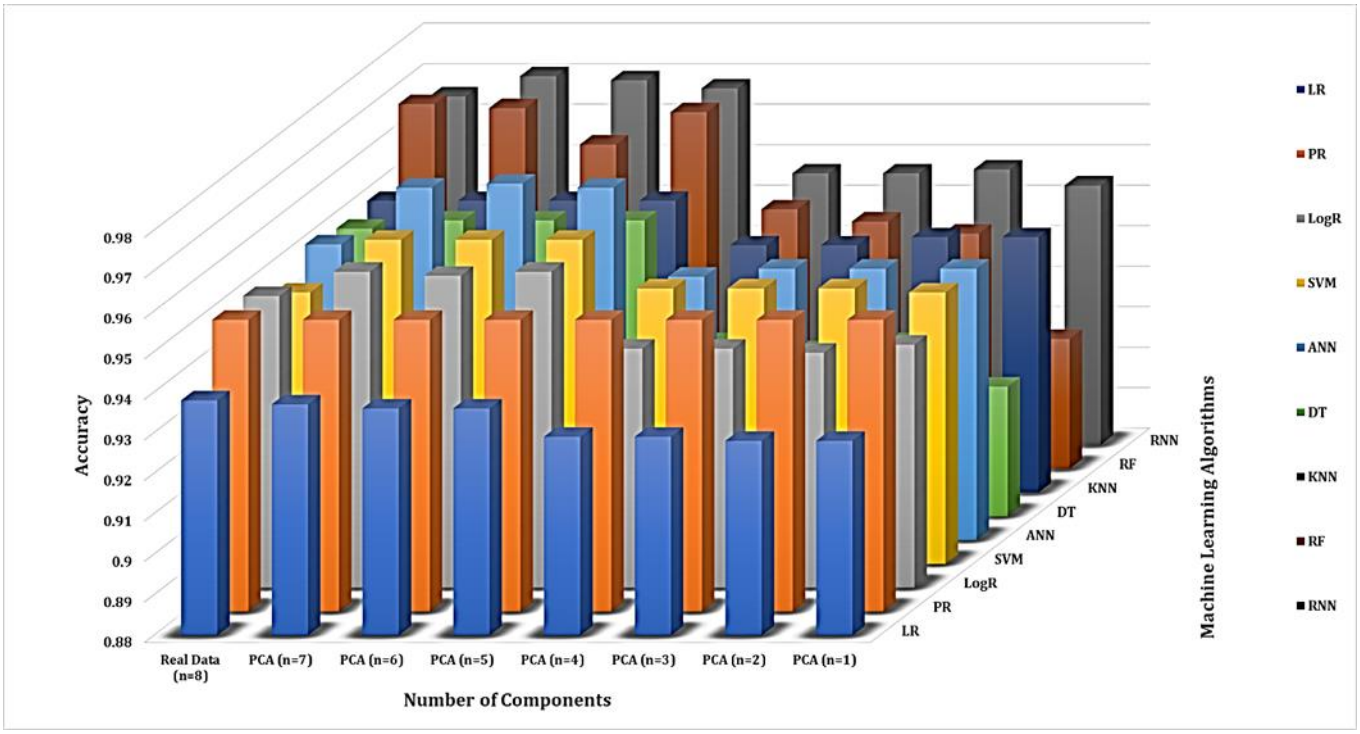


Figure 5. Accuracy of machine learning algorithms according to PCA

4. Conclusion

In this study, data mining approaches were used to evaluate the performance of different machine learning algorithms for diabetes diagnosis. Additionally, the impact of reducing the dataset's dimensionality using the PCA method was examined. The performance of diabetes diagnosis was assessed with 9 different supervised machine learning algorithms applied to the dataset. The analyses revealed that the RF algorithm had the highest accuracy, precision, recall, and F1 score values compared to other algorithms. This result indicates that the RF algorithm can be successfully used in critical medical situations like diabetes diagnosis. Furthermore, ANN, DT, and KNN algorithms also demonstrated good performance. However, the dataset's dimensionality was reduced using the PCA method, and the performance of algorithms was measured with different numbers of components. It was shown that when the PCA transformation is used by selecting the correct number of components, it can enhance the performance of some models and reduce computation time by representing the dataset in a lower dimension. Nevertheless, the choice of the right number of components should be made carefully, considering the dataset's structure and model performance.

Among the limitations of the study are the limited number of features in the dataset and its lack of representation of different types of diabetes. Additionally, the dataset's imbalance should be taken into account, as it may have some impact on classification performance. Future studies are recommended to use larger and balanced datasets and include different diabetes types. Moreover, exploring combinations of different machine learning algorithms and more advanced feature selection methods could further enhance performance.

Author contributions

Yavuz Bahadır Koca: Visualization, Conceptualization, Software, Writing-Reviewing and Editing, Validation.
Elif Aktepe: Methodology, Data curation, Writing-Original draft preparation, Software.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Sowah, R. A., Bampoe-Addo, A. A., Armoo, S. K., Saalia, F. K., Gatsi, F., & Sarkodie-Mensah, B. (2020). Design and development of diabetes management system using machine learning. International Journal of Telemedicine and Applications, 2020(1), 8870141. https://doi.org/10.1155/2020/8870141
2. Yahyaoui, A., Jamil, A., Rasheed, J., & Yesiltepe, M. (2019, November). A decision support system for diabetes prediction using machine learning and deep learning techniques. In 2019 1st International Informatics and Software Engineering Conference (UBMYK), 1-4. https://doi.org/10.1109/UBMYK48245.2019.8965556
3. Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. Computational and Structural Biotechnology Journal, 15, 104-116. https://doi.org/10.1016/j.csbj.2016.12.005
4. Abudnejad, N., Salehpour, M., & Saadati, Z. (2023). Theoretical evaluation of boron carbide nanotubes as non-enzymatic glucose sensors. Chemical Physics Letters, 823, 140510. https://doi.org/10.1016/j.cplett.2023.140510
5. Başer, B. Ö., Yangın, M., & Sarıdaş, E. S. (2021). Makine öğrenmesi teknikleriyle diyabet hastalığının sınıflandırılması. Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 25(1), 112-120. https://doi.org/10.19113/sdufenbed.842460
6. World Health Organization WHO European Regional Obesity Report (2022). World Health Organization. Regional Office for Europe. ISBN 9289057734.
7. Sun, J., Ren, J., Hu, X., Hou, Y., & Yang, Y. (2021). Therapeutic effects of Chinese herbal medicines and their extracts on diabetes. Biomedicine & Pharmacotherapy, 142, 111977. https://doi.org/10.1016/j.biopha.2021.111977
8. Hasanzad, M., Aghaei Meybodi, H. R., Sarhangi, N., & Larijani, B. (2022). Artificial intelligence perspective in the future of endocrine diseases. Journal of Diabetes & Metabolic Disorders, 21(1), 971-978. https://doi.org/10.1007/s40200-021-00949-2
9. Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. Computational and Structural Biotechnology Journal, 15, 104-116. https://doi.org/10.1016/j.csbj.2016.12.005
10. Cappon, G., Acciaroli, G., Vettoretti, M., Facchinetti, A., & Sparacino, G. (2017). Wearable continuous glucose monitoring sensors: a revolution in diabetes treatment. Electronics, 6(3), 65. https://doi.org/10.3390/electronics6030065
11. Zherebtsov, E. A., Zharkikh, E. V., Kozlov, I. O., Loktionova, Y. I., Zherebtsova, A. I., Rafailov, I. E., ... & Rafailov, E. U. (2019, June). Wearable sensor system for multipoint measurements of blood perfusion: pilot studies in patients with diabetes mellitus. In European Conference on Biomedical Optics, 11079_62. https://doi.org/10.1117/12.2526966
12. Mujumdar, A., & Vaidehi, V. (2019). Diabetes prediction using machine learning algorithms. Procedia Computer Science, 165, 292-299. https://doi.org/10.1016/j.procs.2020.01.047
13. Kopitar, L., Kocbek, P., Cilar, L., Sheikh, A., & Stiglic, G. (2020). Early detection of type 2 diabetes mellitus using machine learning-based prediction models. Scientific Reports, 10(1), 11981. https://doi.org/10.1038/s41598-020-68771-z
14. Nadesh, R. K., & Arivuselvan, K. (2020). Type 2: diabetes mellitus prediction using deep neural networks classifier. International Journal of Cognitive Computing in Engineering, 1, 55-61. https://doi.org/10.1016/j.ijcce.2020.10.002
15. Lai, H., Huang, H., Keshavjee, K., Guergachi, A., & Gao, X. (2019). Predictive models for diabetes mellitus using machine learning techniques. BMC Endocrine Disorders, 19, 1-9. https://doi.org/10.1186/s12902-019-0436-6
16. Soni, M., & Varma, S. (2020). Diabetes prediction using machine learning techniques. International Journal of Engineering Research & Technology (IJERT), 9(09), 921-925.
17. Tasin, I., Nabil, T. U., Islam, S., & Khan, R. (2023). Diabetes prediction using machine learning and explainable AI techniques. Healthcare Technology Letters, 10(1-2), 1-10. https://doi.org/10.1049/htl2.12039
18. Cahn, A., Shoshan, A., Sagiv, T., Yesharim, R., Goshen, R., Shalev, V., & Raz, I. (2020). Prediction of progression from pre-diabetes to diabetes: development and validation of a machine learning model. Diabetes/metabolism Research and Reviews, 36(2), e3252. https://doi.org/10.1002/dmrr.3252
19. Dinh, A., Miertschin, S., Young, A., & Mohanty, S. D. (2019). A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Medical Informatics and Decision Making, 19(1), 1-15.

https://doi.org/10.1186/s12911-019-0918-5
20. Hasan, M. K., Alam, M. A., Das, D., Hossain, E., & Hasan, M. (2020). Diabetes prediction using ensembling of different machine learning classifiers. IEEE Access, 8, 76516-76531. https://doi.org/10.1109/ACCESS.2020.2989857
21. Kaur, H., & Kumari, V. (2022). Predictive modelling and analytics for diabetes using a machine learning approach. Applied Computing and Informatics, 18(1/2), 90-100. https://doi.org/10.1016/j.aci.2018.12.004
22. Birjais, R., Mourya, A. K., Chauhan, R., & Kaur, H. (2019). Prediction and diagnosis of future diabetes risk: a machine learning approach. SN Applied Sciences, 1, 1-8. https://doi.org/10.1007/s42452-019-1117-9
23. Nandy, S. (2023). Kaggle. https://www.kaggle.com/datasets/sharmisthanandy/diabetes
24. Abdelaziz, A., Elhoseny, M., Salama, A. S., & Riad, A. M. (2018). A machine learning model for improving healthcare services on cloud computing environment. Measurement, 119, 117-128. https://doi.org/10.1016/j.measurement.2018.01.022
25. Schober, P., & Vetter, T. R. (2021). Statistical Minute Logistic Regression in Medical Research.
26. Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends, 1(2), 140-147. https://doi.org/10.38094/jastt1457
27. Choudhury, A., & Gupta, D. (2019). A survey on medical diagnosis of diabetes using machine learning techniques. In Recent Developments in Machine Learning and Data Analytics: IC3 2018, 740, 67-78. https://doi.org/10.1007/978-981-13-1280-9_6
28. Vijayan, V. V., & Anjali, C. (2015, December). Prediction and diagnosis of diabetes mellitus—A machine learning approach. In 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS), 122-127. https://doi.org/10.1109/RAICS.2015.7488400
29. Ghosh, P., Azam, S., Karim, A., Hassan, M., Roy, K., & Jonkman, M. (2021). A comparative study of different machine learning tools in detecting diabetes. Procedia Computer Science, 192, 467-477. https://doi.org/10.1016/j.procs.2021.08.048
30. Al-Askar, H., Radi, N., & MacDermott, Á. (2016). Recurrent neural networks in medical data analysis and classifications. In Applied Computing in Medicine and Health, 147-165. https://doi.org/10.1016/B978-0-12-803468-2.00007-2
31. Maćkiewicz, A., & Ratajczak, W. (1993). Principal components analysis (PCA). Computers & Geosciences, 19(3), 303-342. https://doi.org/10.1016/0098-3004(93)90090-R

Turkish Journal of Engineering

https://dergipark.org.tr/en/pub/tuje

e-ISSN 2587-1366

Efficient prediction of compressive strength in geotechnical engineering using artificial neural networks

Ali Ulvi Uzer[*1]

[1]*Kayseri University, Department of Construction, Türkiye, aliulviuzer@kayseri.edu.tr*

Cite this study: Uzer, A. U. (2024). Efficient prediction of compressive strength in geotechnical engineering using artificial neural networks. Turkish Journal of Engineering, 8 (3), 457-468

https://doi.org/10.31127/tuje.1415931

Keywords
Artificial intelligence
Regression learner
UCS
ANN

Research Article

Received:07.01.2024
Revised: 20.02.2024
Accepted: 21.02.2024
Published: 05.07.2024

Abstract

In recent years, artificial neural networks (ANNs) have emerged as highly effective tools for addressing the intricate challenges encountered in geotechnical engineering. ANNs find application in a variety of geotechnical problems, showcasing promising outcomes. This study aims to improve the efficiency of predicting intermediate values from unconfined compressive strength (UCS) data obtained from laboratory tests through the use of ANNs. The modelling of artificial neural networks was carried out using the Regression Learner program, integrated with the Matlab 2023a software package, offering a user-friendly graphical interface for AI model development without the need for coding. The ANNs' validation and training were based on UCS test data obtained from the Geotechnical Laboratory of Iowa State University, USA. These laboratory tests focused on engineering properties, specifically the UCS of soils treated with biofuel co-products (BCPs). The dataset, organized in a matrix of size 216 × 5, features columns providing information on soil type (Soil 1; Soil 2; Soil 3; Soil 4), sample type (pure soil-untreated; 12% BCP- treated soil; 3% cement; 6% cement; 12% cement treated soil), time (1, 7, and 28 days), moisture content (OMC-4%, OMC%, and OMC+4%), and corresponding UCS peak stress (psi) values. The AI predictions for the test data output achieved an outstanding R^2 score of 0.93, showcasing the potential of employing ANNs to efficiently acquire a substantial amount of data with fewer experiments and in less time. This approach holds promise for applications in geotechnical engineering.

1. Introduction

In recent years, the application of artificial neural networks (ANNs) has garnered significant attention as a powerful and effective methodology for tackling the complex challenges prevalent in geotechnical engineering. The evolution of computational geotechnical engineering analyses closely aligns with the advancements in computational methods. In the initial stages of geotechnical engineering, analytical methods and a simple limit equilibrium approach, combined with engineering expertise, were employed to construct physical models for addressing geotechnical engineering challenges. Over time, more complex problems necessitated the use of finite element methods, finite difference methods, and discrete element methods. Nevertheless, the efficacy of these approaches in geotechnical engineering faces impediments, primarily stemming from the difficulty in formulating precise constitutive models and addressing the spatial variability of soil, particularly in complex situations like liquefaction and pile capacity issues. As a result, statistically derived empirical methods and semi-empirical methods, grounded in analytical approaches, have become prevalent. Their success hinges significantly on the selected statistical or theoretical model that corresponds to the system under analysis [1].

The unpredictable behaviours of soil and rock, arising from complex physical processes in their formation, pose a challenge that traditional engineering models often simplify [2]. The methodological fit of artificial neural networks (ANNs) becomes evident in modelling complex problems where the relationships among variables are unknown [3]. Functioning as a computational emulation of the human brain's physiological structure, ANNs depart from conventional signal reasoning and logical thinking. This machine learning technique excels in handling challenges related to incomplete associative memory, faulty pattern recognition, and autonomous learning. ANNs boast three key advantages: rapid computational speed, robust fault-tolerance, and proficiency in solving problems governed by intricate solution rules [4].

The predominant advancements in artificial intelligence predominantly stem from statistical models, with artificial neural networks standing out as the

forefront contributor. This model emulates the learning mechanisms of the human brain within a computational framework. Artificial neural networks comprise computational units known as neurons, akin to the neural structure in the human brain. Neurons are interconnected through synapses, featuring weighted connections, further mirroring the intricate network observed in biological systems [5].

The calculation in neurons is provided by multiplying the information in the neuron with the weight in the synapse. In artificial neural networks, with the bond of synapses between neurons and the formation of layers of neurons, it creates network models as in the human brain. A comprehensive literature review highlights that artificial neural networks (ANNs) have been successfully applied in various geotechnical disciplines. These applications include, but are not limited to, pile capacity estimation, soil behavior modeling, site characterization, analysis of soil retaining structures, settlement evaluation of structures, slope stability analysis, design of tunnels and underground openings, assessment of liquefaction susceptibility, determination of soil permeability and hydraulic conductivity. It covers a wide range such as evaluation of soil compaction, investigation of soil swelling phenomenon and classification of soil [6]. The effectiveness of these empirical and semi-empirical methods is contingent on the selected statistical or theoretical model, as well as the statistical methods used to determine the model parameters [7]. The complexity and uncertainty of soil parameters often make it challenging to develop theoretical or statistical models, prompting a preference for a data-driven approach over the traditional model-oriented one.

To tackle the formidable challenges encountered in various disciplines, the integration of artificial intelligence (AI) into computational methods has become increasingly prevalent. This infusion of AI has sparked significant research growth, offering innovative solutions to real-life problems while simultaneously unveiling both the latent capabilities and drawbacks of these advanced techniques. The utilization of Artificial Neural Networks (ANNs) in the realm of geotechnical engineering saw its inception in the early 1990s, marked notably by the pioneering work of Goh [8-9]. Goh's research, in particular, demonstrated the remarkable capability of ANNs to predict the intricate liquefaction potential of soil and articulate the intrinsic constitutive relationship of sand using this innovative technology.

The evolution of AI applications in geotechnical engineering signifies a paradigm shift in how we approach and address complex issues within this field. By leveraging the inherent learning and adaptability of ANNs, researchers have made strides in comprehending and predicting intricate phenomena, such as soil liquefaction, which were historically challenging to model accurately. The foundational contributions of Goh and subsequent researchers have laid the groundwork for a more nuanced understanding of the interplay between AI and geotechnical challenges. This not only opens new avenues for solving longstanding problems but also prompts a critical examination of the potential limitations and ethical considerations associated with the widespread adoption of AI in geotechnical research and practice. As the field continues to evolve, these insights and advancements are poised to shape the future landscape of geotechnical engineering, paving the way for more effective and informed decision-making processes.

The heightened attention on Artificial Neural Networks (ANNs) in geotechnical engineering reflects the growing recognition of their efficacy as a potent and adaptable statistical technique for resolving intricate problems. Pioneering applications, as exemplified by Shahin et al. [6] and Das [10], have vividly illustrated the versatility of ANNs in successfully addressing a spectrum of challenges within the field. In a comprehensive overview by Shahin et al. [11], the current status and future prospects of ANNs were discussed, underscoring their increasing reliability in comparison to traditional statistical methods.

Despite this surge in interest and application, there exists a conspicuous gap in the literature – a lack of a comprehensive critical evaluation of the modeling aspects of ANNs in geotechnical engineering. This gap underscores the necessity for a thorough exploration and examination of the nuances associated with employing ANNs in this specific domain. While the potential of ANNs is evident, it is crucial to acknowledge, as noted by Yang et al. [12], that the efficiency of numerical methods is inherently problem-dependent. No singular technique can universally serve as a panacea for solving all types of geotechnical problems. Consequently, this study aims to fill this void by scrutinizing the modeling aspects of ANNs in geotechnical engineering, shedding light on their strengths, limitations, and optimal applications. By doing so, it seeks to contribute to a more nuanced understanding of the role of ANNs in tackling the complexities inherent in geotechnical challenges and provide insights for refining their utilization in future endeavors within the field.

Grima & Babuška [13] introduced a data-driven approach for modeling the unconfined compressive strength of rock samples using fuzzy logic. This method represents a nonlinear relationship through a concatenation of local linear sub-models obtained via fuzzy clustering. Compared to conventional statistical models, the fuzzy model exhibits superior accuracy while providing insights into the nonlinear relationship, a feature lacking in other black-box approaches such as neural networks.

Ceylan et al. [14] reported that two different biopolymer blends containing 12% provided maximum durability increase in the UCS test at 1 day and 7 days recovery times. The results showed that this ratio was the most effective ratio for soil stabilization. In addition, Ceylan et al. [14] conducted UCS tests of sulfur-free lignin-containing biopolymer blends under saturation and semi-saturation conditions. In these tests, it was observed that 12% of the blend provided a significant durability increase, especially with liquid type applications. In their seminal work, conducted UCS tests on specimens subjected to both saturation and half-saturation, noting a substantial increase in strength with the two biofuel co-products (BCPs) treatments, particularly with the liquid-type treatment.

Kurugodu [15] investigated the impact of density, moisture, and fiber content on the unconfined compressive strength (UCS) of silty sand. Employing a distinctive multi-gene genetic programming (MGGP) approach coupled with ANN, they aimed to formulate transparent models illuminating the intricate relationships between UCS and crucial soil parameters. Noteworthy for its superior performance in sensitivity and parametric analyses, the MGGP model, with its transparent formulation, outperformed conventional methods. This study not only contributed theoretical insights but also presented practical implications for optimizing input values in geotechnical infrastructure design.

2. Materials and Method

2.1. Materials

Unconfined Compressive Strength (UCS) is a key mechanical property used to characterize the strength of a material, particularly soils and rocks. It is a measure of the ability of a material to withstand axial loads or uniaxial compression without lateral confinement. The test is commonly used in geotechnical engineering to assess the strength of soil and rock specimens. In general, Unconsolidated-Undrained (UU) tests are conducted for the rapid loading assessment of clay soils or post-construction stability analyses of embankment dams [16]. The test involves applying a uniaxial load to a cylindrical specimen of the material until failure occurs. During the test, the specimen is free to deform laterally (without confinement), and the applied axial stress is gradually increased until the specimen fails. The maximum axial stress at failure is recorded as the unconfined compressive strength.

To assess the Unconfined Compressive Strength (UCS) of the studied materials, a comprehensive testing methodology was employed, adhering to established standards and procedures. For soils, the unconfined compressive strength is often used as an indicator of the material's load-bearing capacity and its suitability for construction purposes. It is a critical parameter in slope stability analysis, foundation design, and other geotechnical assessments. The testing procedure strictly adhered to the guidelines outlined in ASTM D 2166, titled "Standard Test Method for UCS of Cohesive Soil." This standard provided a robust framework for conducting UCS tests on soil specimens. Figure 1 shows the automated computer control system used in this study to determine the soil UCS at the Iowa State University Geotechnical Laboratory [17].

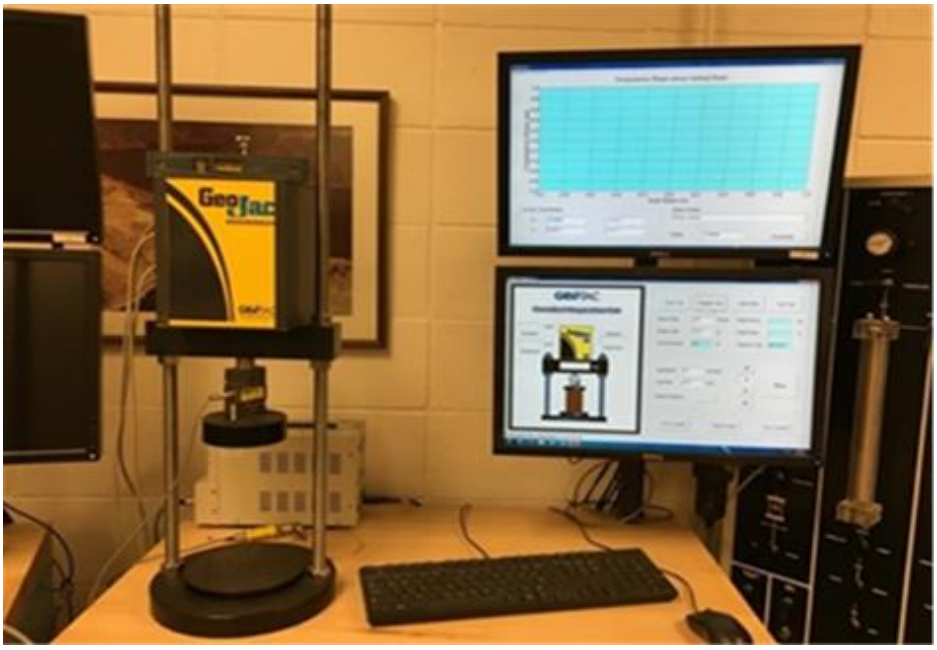


Figure 1. Automated Geotac system for unconfined compressive strength testing.

The equipment utilized a strain-controlled load rate, ensuring a constant axial strain rate during the testing process. Rectangular specimens measuring 2 inches by 2 inches were meticulously prepared for testing. Post-curing, these specimens were loaded into the frame of the automated testing apparatus. The prepared specimens underwent sustained force within the testing apparatus until failure. Throughout this process, the load cell indicator and strain gauge diligently recorded stress and strain, offering a comprehensive dataset on the specimen's behaviour under axial loading. The stress applied to the specimen exhibited a characteristic pattern, escalating with the increase in strain until reaching a peak. Subsequently, the stress decreased due to the sample crush. The automated computer control system facilitated the plotting of the specimen's strain-stress relationship, allowing for the identification and display of the peak stress.

The peak stress of the specimen was pinpointed as the stress value when the specimen reached the 15% strain limit without experiencing crush. This parameter served as a crucial indicator of the specimen's maximum strength under uniaxial compression. Statistical methods were employed to analyse the data, drawing meaningful conclusions regarding the effectiveness of the tested additives in soil stabilization. Multiple tests were

conducted to ensure result reproducibility. Precision and accuracy were maintained throughout the testing process to enhance the reliability of the acquired data.

A comprehensive laboratory experimental test program was performed, comparing the unconfined compressive strength of BCP treated on four different representative Iowa soil types. The engineering properties of the soil samples are shown in Table 1 [18].

Table 1. Engineering properties for four types of soils.

Property Classification	Soil 1	Soil 2	Soil 3	Soil 4
AASHTO (group index)	A-6 (2)	A-4 (2)	A-4 (1)	A-4 (0)
USCS group symbol	SC	CL-ML	CL-ML	ML
USCS group name	Clayed sand	Sandy Silty with clay	Sandy Silty with clay	Sandy Silty
Grain size distribution				
Gravel (> 4.75 mm. %	7.1	0.1	5.2	3.8
Sand (0.075–4.75 mm) %	54.9	37.2	41.7	45.3
Silt and clay (< 0.075mm) %	38.0	62.7	53.1	50.9
Atterberg limits				
Liquid limit (LL) %	32.8	29.1	27.5	17.2
Plasticity limit (PL) %	17.4	22.9	22.2	15.1
Plasticity index (PI) %	15.4	6.2	5.3	2.1
Proctor test				
Optimum moisture content (OMC) %	14.4	18.2	13.5	12.0
Maximum dry unit weight ($\gamma_{d\,max}$) kg/m^3(pcf)	1.728 (107.9)	1.631 (101.8)	1.818 (113.5)	1.839 (114.8)

Considering lignin's significant presence in plant biomass, researchers at Iowa State University (ISU) have previously proposed employing sulfur-free lignin for soil stabilization, aiming to harness potential economic benefits from lignocellulosic biorefineries. In their study, sandy lean clay (CL) soil underwent treatment with two distinct Biopolymer Composite Polymers (BCPs) containing sulfur-free lignin, one in liquid form and the other as a yellow powder. The researchers introduced each BCP into the soil under various moisture conditions: dry side (OMC−4%), optimum moisture content

(OMC), and wet side (OMC+4%), with an addition of up to 15% dry unit weight. After 1-day and 7-day curing periods, the specimens incorporating 12% of the two BCPs exhibited the maximum strength improvement (UCS). Additionally, UCS tests were conducted under both saturation and half-saturation conditions, revealing substantial strength enhancement, particularly with the liquid-type treatment [19].

In a related context, Puppala & Puppala et al. [20-21] explored the application of two additional BCPs containing sulfur-free lignin, incorporating up to 15% by dry soil weight for the treatment of silt soil. Their findings echoed those of the ISU researchers, indicating that a 12% application rate for both BCPs led to the highest strength improvement after 1-day, 7-day, and 28-day curing periods. Complementing these strength assessments, XRD and SEM analyses were conducted to elucidate physical bonds as the mechanism behind sulfur-free lignin's efficacy in soil stabilization. The results underscored the positive role of sulfur-free lignin in soil stabilization, recommending an application rate of 12% by dry soil weight [17].

2.2. Methods

The layered architecture of ANN draws inspiration from the intricate organization of the human nervous system, forming connections among neurons in diverse topologies. ANNs, designed with this biomimetic approach, possess the capability to be trained for specific functions by optimizing the values of connection multipliers. Each neural node within this network encapsulates input values, weights, addition operations, transfer functions, and output values, mirroring the fundamental elements of the human brain's neural structure. Operating as parallel processors, ANNs exhibit proficiency in receiving, storing, and generalizing information connections, showcasing their adaptability across various domains. The iterative learning process within ANNs involves sophisticated algorithms that adjust weights iteratively to attain the desired outcomes, highlighting their dynamic nature in adapting to complex tasks. This architectural and functional resemblance to the human nervous system underscores the versatility and efficiency of ANNs in tackling intricate challenges in diverse applications [22-23].

Neural networks, whether they manifest as biological entities or artificial constructs, embody intricate systems of interconnected neurons. Biological neural networks consist of natural neurons, while artificial neural networks, fashioned from artificial nodes, are strategically devised to tackle challenges within the realm of artificial intelligence. Within artificial networks, the connections between neurons are simulated through weights, with positive values denoting excitatory links and negative values signifying inhibitory connections [24]. The amalgamation of inputs is accomplished through these weights, and an activation function governs the modulation of output amplitudes, typically within a range of 0 to 1 or -1 to 1. This emulation of neural connectivity in artificial systems reflects the convergence of biological inspiration and technological innovation, contributing to the advancement of artificial intelligence across various domains.

The application of artificial neural networks (ANNs) spans various domains, including predictive modeling, adaptive control, and data-driven training, enabling them to autonomously learn and draw conclusions from intricate information. Motivated by the structural and functional intricacies of the human brain, scientists have crafted models for artificial neurons and networks, giving rise to the field of artificial neural networks (ANN). An ANN is a system inspired by the neural architecture of the human brain, endowed with the capability to perform specific functions. The visual representation of an ANN is encapsulated in Figure 2, presenting a general block diagram that illustrates the network's structure and connectivity. This amalgamation of neuroscience principles and computational techniques fuels the evolution of artificial intelligence, offering versatile applications in problem-solving and decision-making realms [25].

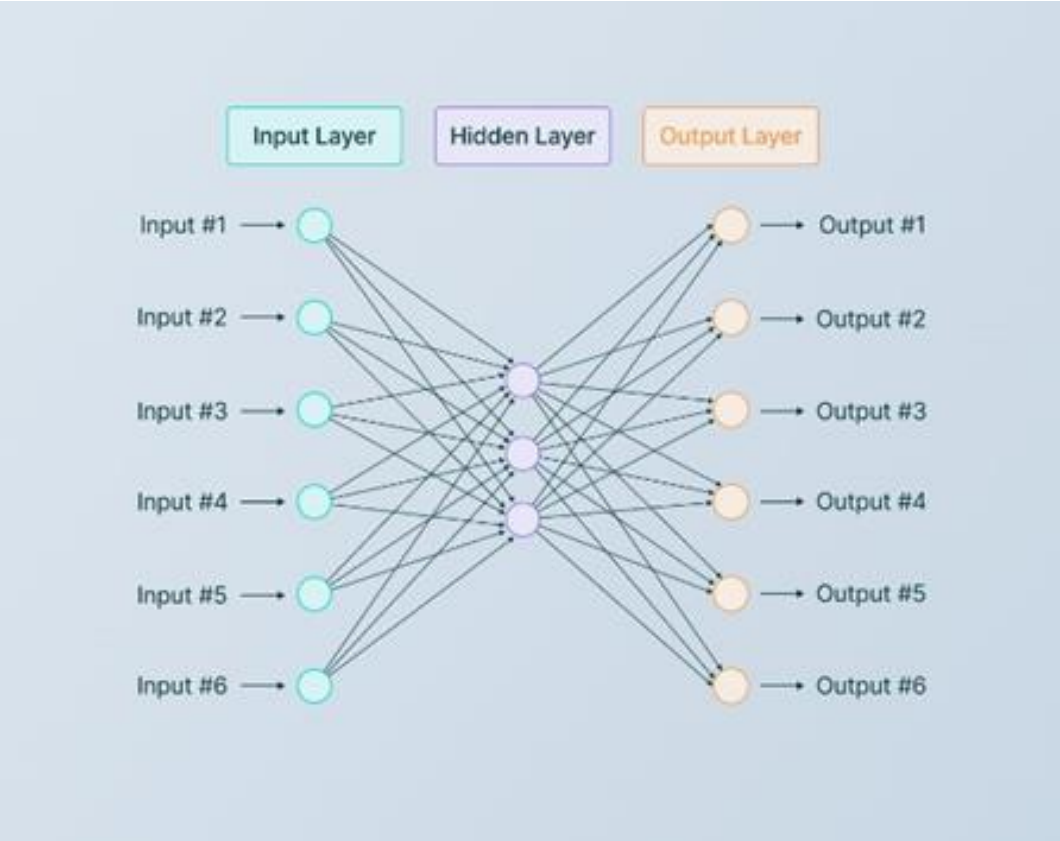


Figure 2. The working structure of neural networks.

Artificial Neural Networks (ANNs) exhibit a distinctive architectural composition comprising three essential layers: the input layer, the hidden layer(s), and the output layer. The inherent design necessitates intricate connections, with nodes in the input layer linked to nodes in the hidden layer and reciprocal connections between each hidden layer node and the nodes in the output layer. Serving as the initial point of interaction, the input layer assimilates data into the network. Subsequently, the hidden layer undertakes the task of receiving and processing raw information from the input layer. The outcome of this processing is then transmitted to the output layer, where further information processing occurs, culminating in the generation of the final output. This layered structure, characterized by interconnected nodes and sequential information processing, embodies the fundamental framework of ANNs, contributing to their versatility in addressing complex problems across various domains [26].

Figure 3 visually depicts the key elements of a neural network: Where:

'n' is the number of neuron inputs (number of neuron weights must be the same)

'w_1, w_2, w_n' are the neuron weight values that are used to describe how strong influence has related input to the neuron,

'x_1, x_2, x_n' are the neuron input values (it is just an array of input values),

'w (weight)' weight is the bias/threshold that is used to stimulate or suppress neuron activity,

'b (bias)' is the “bias” error/deviation is the value reflecting the distance between the data predicted as a result of modelling and the actual data. “Weight” corresponds to in-puts values,

'g' represents sum passing through a neuron activation function, and

'y' is the neuron output value.

Input values, following multiplication by weights and subsequent summation with bias translation, yield an output through a linear or nonlinear transfer function. This model establishes a mathematical relationship between inputs and outputs, necessitating the optimization of 'w' (weights) and 'b' (bias) values to enable the neuron to generate the desired output [27].

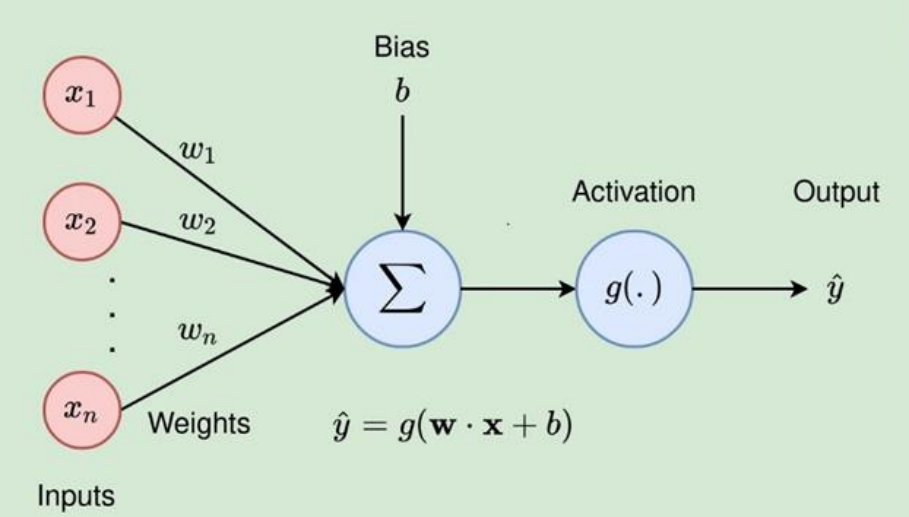


Figure 3. Artificial neuron model.

Regression methods serve as a pivotal tool in statistical modelling, elucidating the intricate relationships and correlations between an output variable and one or more input variables. Within the Matlab software package, the versatile toolboxes empower researchers and analysts to articulate, employ, and explore a spectrum of regression methods. These methods encompass non-linear, generalized, and linear regression techniques, with the flexibility to incorporate cascade and/or mixed models. The capabilities afforded by these toolboxes extend beyond mere analysis; they enable the design of robust models, the prediction of outputs, the assessment of model performance, and the visual expression of complex relationships. The dynamic functionalities embedded in Matlab's toolset elevate regression analysis to a multifaceted approach, providing a comprehensive framework for researchers to unravel the intricacies of variable dependencies in diverse datasets.

The Regression Learner application within the Matlab software package offers a comprehensive suite of tools, incorporating non-parametric methods that enable the integration of intricate prediction curves without explicitly defining the relationship between outputs and a predetermined predictor. Once the training phase concludes, the adeptly trained model can seamlessly generate predictions using novel data, activating its application in the field for real-time decision-making and analysis. This functionality enhances the adaptability and predictive power of regression models, catering to the complexities of diverse datasets and dynamic environments.

Linear regression models serve as a robust statistical approach employed to characterize the intricate relationship between a system's response and its input variables. In the context of a single input variable, this methodology is referred to as simple linear regression, whereas situations involving multiple input variables are encapsulated under the term multiple linear regressions. These models articulate relationships through linear prediction functions derived from empirical data, providing a valuable framework for analyzing and understanding the dependencies within complex systems [28].

Within the realm of statistical modelling, linear regression stands as a foundational tool, indispensable for delving into the connections between a dependent variable, often symbolized as 'y', and one or more independent variables denoted as 'x'. The 'y' variable acts as the response variable, signifying the focal point of interest, while 'x' encapsulates the independent variables, also recognized as explanatory or predictor variables. These predictor variables play a pivotal role in elucidating variations observed in the response variable. Linear regression, therefore, provides an invaluable framework for comprehensively examining and understanding the intricate relationships within datasets, offering insights into the factors influencing the response variable [29].

In conclusion, within the methodological framework, a critical distinction must be made between two fundamental types of independent variables. Continuous explanatory variables, denoted as covariates, are characterized by their measurable and uninterrupted nature. In contrast, categorical independent variables are identified as factors, emphasizing their discrete and categorical attributes. An additional crucial consideration is the incorporation of the matrix 'x,' which encompasses observations related to predictor variables. Referred to as the design matrix, this element assumes a central role in both the formulation and execution of linear regression models. The design matrix serves as a vital tool, encapsulating the intricate relationships among variables and making substantial contributions to the overall structure and interpretability of the model. As researchers navigate the intricacies of their analyses, a comprehensive understanding of these components is imperative for robust and nuanced interpretations within the domain of linear regression modelling [22].

In statistical analysis, the coefficient of determination, commonly denoted as R^2 or r^2, plays a crucial role in assessing the predictability of the dependent variable from one or more independent variables. Error metrics for the neural network model were calculated by the formula in Equation 1-3. This metric holds particular significance in statistical models employed for forecasting future outcomes and testing hypotheses grounded in pertinent information [27-28]. R^2 quantifies

the precision with which observed outcomes align with the model, gauged by the proportion of total outcome variation explained by the model. The R^2 value is confined to the range of 0 to 1, where values closer to 1 signify lower error and a superior fit of the model. This measure provides valuable insights into the effectiveness and reliability of the statistical model in capturing and explaining the observed variability in the data [30].

$$RMSE = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\left(Z_{pred} - Z_{meas}\right)^2} \quad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Z_{pred} - Z_{meas}\right| \quad (2)$$

$$R^2 = \frac{\sum_{i=1}^{n}\left(Z_{i\,meas} - \overline{Z_{i\,meas}}\right)^2.\left(Z_{i\,pred} - \overline{Z_{i\,pred}}\right)^2}{\sum_{i=1}^{n}\left(Z_{i\,meas} - \overline{Z_{i\,meas}}\right)^2.\sum_{i=1}^{n}\left(Z_{i\,pred} - \overline{Z_{i\,pred}}\right)^2} \quad (3)$$

3. Results

In the Unconfined Compressive Strength (UCS) tests, the influence of various additives on soil compressive strength becomes evident. The specimens treated with both by-products and cement showcased notably elevated strength capacities when compared to untreated samples. While all additives contributed to enhancing the inherent strength of natural soil, their impacts were diverse owing to distinct underlying mechanisms. Cement emerged as the most influential additive, imparting a substantial and consistent improvement in strength across all soil types. Generally, the strength of specimens treated with cement exhibited an upward trajectory corresponding to higher cement content, moisture levels, and prolonged curing times, primarily attributed to the process of hydration. These findings underscore the nuanced interplay between different additives and soil properties, offering valuable insights into the optimization of soil stabilization techniques for diverse geotechnical applications.

A comparative analysis of the compressive strength results at three different optimum moisture contents (OMC) was carried out; as shown in Table 2, it was observed in the UCS test with sulfur-free lignin-containing by-products that OMC-4% gave the highest value for all soil types. Remarkably, the addition of 12% (BCP) to the initially pure soil resulted in a significant and consistent increase in shear strength for all soil types. Sulfur-free lignin-containing by-products exhibited a moderate increase in soil strength attributed to the presence of lignin in the untreated soil, ranging from about 20% to 500%. The specimens of Soil 1 treated with BCP B showed the highest improvement in UCS when compared to other soils with BCP B, and the strength of Soil 1 was improved by over 300% for 1-day curing (Table 2).

The dataset presented in Table 2, which was meticulously curated within the confines of this, underwent rigorous modeling procedures utilizing the Regression Learner application within the Matlab 2023a software package. The execution of these modeling endeavors transpired on a desktop computer equipped with an Intel i5 processor boasting 8GB of memory. This computational environment was chosen to ensure the robustness and efficiency of the experiments, allowing for the meticulous exploration and analysis of the intricate relationships embedded within the dataset.

The Matlab Regression Learner application facilitates the development of artificial intelligence models through an intuitive graphical user interface, eliminating the need for manual code creation. The application interface can be accessed by entering the 'Regression Learner' command in the Matlab console. Upon launching the application, a new experiment is initiated using the 'New Session' command. It is essential to import a pre-prepared dataset into the Matlab Workspace.

In the presented dataset, structured in a 216×6 matrix, each row corresponds to a distinct sample, while the columns delineate various parameters. The dataset encompasses information on soil type, categorized as Soil 1, Soil 2, Soil 3, and Soil 4, along with sample types, including pure soil (untreated), 12% BCP treated soil, 3% cement treated soil, 6% cement treated soil, and 12% cement treated soil. Temporal aspects are captured through the time variable, representing intervals of 1, 7, and 28 days. Moisture content is detailed as OMC-4%, OMC%, and OMC+4%, while compressive stress is quantified in pounds per square inch (psi). This structured format, delineated in Table 3, is integral to understanding the interactions between these parameters in the context of soil treatment and the corresponding compressive stress over various temporal durations.

When applying artificial intelligence techniques to a dataset, it is customary to partition the dataset into training and testing sets to evaluate the model's performance. Typically, in practical scenarios, a randomly selected portion, often 90% or 80%, is designated for training, while the remaining 10% or 20% is reserved for testing, commonly known as validation data. The Matlab Regression Learner application facilitates this process through k-fold cross-validation. In our experiments, we employed k = 5, dividing the dataset into five sections and iteratively conducting experiments five times. In each iteration, four sections were utilized for training, while one section was held out for testing. The training and testing phases were systematically executed by sequentially shifting through the partitions. The computed error metrics represent the average of the results obtained across the five experiments. For the experiments outlined in this section, we adopted k = 5,

thereby conducting training with 80% of the data and testing with the remaining 20%.

Figure 4 illustrates the graphical representation of the system output values (ideal compressive stress) for the test data and the graphical representation of the predictions made by the artificial intelligence model.

Table 2. The dataset utilized in MATLAB's regression learner application for unconfined compressive strength (UCS) test values.

Soil type	Sample type	Day (1-7-28)	OMC (-4, 0, +4)	Output (psi)	Soil type	Sample type	Day (1-7,28)	OMC (-4, 0, +4)	Output (psi)	Soil type	Sample type	Day (1-7,28)	OMC (-4, 0, +4)	Output (psi)
1	1	1	-4	84,00	2	1	1	0	13,00	3	1	1	4	26,00
1	2	1	-4	113,00	2	2	1	0	29,00	3	2	1	4	64,00
1	3	1	-4	173,00	2	3	1	0	29,00	3	3	1	4	37,00
1	4	1	-4	122,00	2	4	1	0	87,00	3	4	1	4	111,00
1	5	1	-4	203,00	2	5	1	0	145,00	3	5	1	4	170,00
1	6	1	-4	281,00	2	6	1	0	282,00	3	6	1	4	299,00
1	1	7	-4	89,00	2	1	7	0	16,00	3	1	7	4	23,00
1	2	7	-4	136,00	2	2	7	0	31,00	3	2	7	4	46,00
1	3	7	-4	183,00	2	3	7	0	37,00	3	3	7	4	47,00
1	4	7	-4	156,00	2	4	7	0	122,00	3	4	7	4	135,00
1	5	7	-4	272,00	2	5	7	0	212,00	3	5	7	4	245,00
1	6	7	-4	447,00	2	6	7	0	336,00	3	6	7	4	370,00
1	1	28	-4	93,00	2	1	28	0	12,00	3	1	28	4	24,00
1	2	28	-4	103,00	2	2	28	0	23,00	3	2	28	4	28,00
1	3	28	-4	206,00	2	3	28	0	39,00	3	3	28	4	60,00
1	4	28	-4	234,00	2	4	28	0	238,00	3	4	28	4	158,00
1	5	28	-4	320,00	2	5	28	0	384,00	3	5	28	4	301,00
1	6	28	-4	747,00	2	6	28	0	456,00	3	6	28	4	556,00
1	1	1	0	42,00	2	1	1	4	11,00	4	1	1	-4	27,00
1	2	1	0	71,00	2	2	1	4	19,00	4	2	1	-4	115,00
1	3	1	0	80,00	2	3	1	4	25,00	4	3	1	-4	47,00
1	4	1	0	148,00	2	4	1	4	58,00	4	4	1	-4	72,00
1	5	1	0	262,00	2	5	1	4	129,00	4	5	1	-4	117,00
1	6	1	0	400,00	2	6	1	4	204,00	4	6	1	-4	197,00
1	1	7	0	37,00	2	1	7	4	12,00	4	1	7	-4	27,00
1	2	7	0	110,00	2	2	7	4	22,00	4	2	7	-4	112,00
1	3	7	0	106,00	2	3	7	4	28,00	4	3	7	-4	79,00
1	4	7	0	228,00	2	4	7	4	112,00	4	4	7	-4	147,00
1	5	7	0	369,00	2	5	7	4	188,00	4	5	7	-4	225,00
1	6	7	0	664,00	2	6	7	4	274,00	4	6	7	-4	366,00
1	1	28	0	41,00	2	1	28	4	10,00	4	1	28	-4	35,00
1	2	28	0	102,00	2	2	28	4	16,00	4	2	28	-4	87,00
1	3	28	0	126,00	2	3	28	4	30,00	4	3	28	-4	100,00
1	4	28	0	311,00	2	4	28	4	195,00	4	4	28	-4	181,00
1	5	28	0	648,00	2	5	28	4	256,00	4	5	28	-4	255,00
1	6	28	0	955,00	2	6	28	4	313,00	4	6	28	-4	510,00
1	1	1	4	22,00	3	1	1	-4	68,00	4	1	1	0	12,00
1	2	1	4	57,00	3	2	1	-4	115,00	4	2	1	0	71,00
1	3	1	4	64,00	3	3	1	-4	79,00	4	3	1	0	36,00
1	4	1	4	136,00	3	4	1	-4	128,00	4	4	1	0	45,00
1	5	1	4	226,00	3	5	1	-4	222,00	4	5	1	0	77,00
1	6	1	4	379,00	3	6	1	-4	324,00	4	6	1	0	136,00
1	1	7	4	28,00	3	1	7	-4	67,00	4	1	7	0	14,00
1	2	7	4	57,00	3	2	7	-4	94,00	4	2	7	0	55,00
1	3	7	4	76,00	3	3	7	-4	122,00	4	3	7	0	52,00
1	4	7	4	228,00	3	4	7	-4	155,00	4	4	7	0	131,00
1	5	7	4	341,00	3	5	7	-4	277,00	4	5	7	0	255,00
1	6	7	4	613,00	3	6	7	-4	442,00	4	6	7	0	438,00
1	1	28	4	24,00	3	1	28	-4	66,00	4	1	28	0	15,00
1	2	28	4	60,00	3	2	28	-4	81,00	4	2	28	0	33,00
1	3	28	4	101,00	3	3	28	-4	150,00	4	3	28	0	70,00
1	4	28	4	323,00	3	4	28	-4	214,00	4	4	28	0	194,00
1	5	28	4	698,00	3	5	28	-4	355,00	4	5	28	0	291,00
1	6	28	4	1057,00	3	6	28	-4	538,00	4	6	28	0	597,00
2	1	1	-4	20,00	3	1	1	0	33,00	4	1	1	4	9,00
2	2	1	-4	51,00	3	2	1	0	73,00	4	2	1	4	34,00
2	3	1	-4	41,00	3	3	1	0	48,00	4	3	1	4	21,00
2	4	1	-4	69,00	3	4	1	0	123,00	4	4	1	4	28,00
2	5	1	-4	159,00	3	5	1	0	200,00	4	5	1	4	57,00
2	6	1	-4	243,00	3	6	1	0	334,00	4	6	1	4	119,00
2	1	7	-4	21,00	3	1	7	0	30,00	4	1	7	4	11,00
2	2	7	-4	44,00	3	2	7	0	57,00	4	2	7	4	34,00
2	3	7	-4	47,00	3	3	7	0	90,00	4	3	7	4	28,00
2	4	7	-4	114,00	3	4	7	0	140,00	4	4	7	4	96,00
2	5	7	-4	225,00	3	5	7	0	274,00	4	5	7	4	208,00
2	6	7	-4	294,00	3	6	7	0	474,00	4	6	7	4	406,00
2	1	28	-4	19,00	3	1	28	0	30,00	4	1	28	4	11,00
2	2	28	-4	28,00	3	2	28	0	51,00	4	2	28	4	25,00
2	3	28	-4	49,00	3	3	28	0	112,00	4	3	28	4	40,00
2	4	28	-4	201,00	3	4	28	0	174,00	4	4	28	4	139,00
2	5	28	-4	357,00	3	5	28	0	306,00	4	5	28	4	223,00
2	6	28	-4	435,00	3	6	28	0	607,00	4	6	28	4	452,00

Table 3. Performance evaluation and error metrics of regression models in artificial intelligence developed using Matlab for test data.

Models	Mean Absolute Error (MAE)	R^2	Root Mean Square Error (RMSE)
Neural Network (Model 2.23)	34.6540	0.93	47.9080
Neural Network	35.0330	0.92	48.6610
Neural Network	30.8660	0.92	49.8230
Neural Network	31.9390	0.91	51.4860
Neural Network	32.4260	0.91	51.8030
Neural Network	37.6200	0.91	53.8380
Neural Network	41.4800	0.89	57.9310
Gaussian Process Regression	34.8990	0.89	59.1120
Neural Network	41.7060	0.89	59.5020
Ensemble	37.8610	0.85	68.1650
Ensemble	46.5650	0.79	79.5500
SVM	49.7420	0.80	79.2120
Tree	45.0240	0.79	81.1080
Tree	46.6330	0.76	86.2720

The error values corresponding to the test data are depicted in Figure 5. Notably, some instances exhibit errors in the range of approximately -100 to +100. Given that the ideal compressive stress values are provided within the range of 0 to 1057 (psi), it is evident that a minor degree of prediction error exists for a limited number of examples.

Observing the distribution of errors in the context of the broader range of compressive stress values underscores the overall effectiveness of the predictive model. While discrepancies exist in specific cases, these errors remain within an acceptable margin relative to the overall scale of compressive stress values. This nuanced evaluation emphasizes the model's ability to make reasonably accurate predictions across a diverse range of examples.

The numerical results for the test data, presented in descending order of performance (R^2), are detailed in Table 3. Notably, the neural network method emerged as the most effective among the assessed approaches. As can be seen in Table 3, the highest performance values are obtained for the neural network method regression model. The error metrics for the neural network model are calculated as follows: Mean Absolute Error (MAE) 34.6540, R-Squared (R^2) 0.93, and Root Mean Square Error (RMSE) 47.9080. Remarkably, the model demonstrated remarkable proficiency with an average error of approximately 30.00, showcasing its accuracy in predicting compressive stress values within the expansive range of 0-1057.

In Figure 6, the output parameter (depicted in blue) of the artificial intelligence regression model, trained with compressive stress values for the entire dataset, is presented alongside the corresponding differences (illustrated in orange). The model was trained comprehensively on all available data. As discernible from Figure 6, the errors for many examples are notably low and exhibit a high degree of similarity.

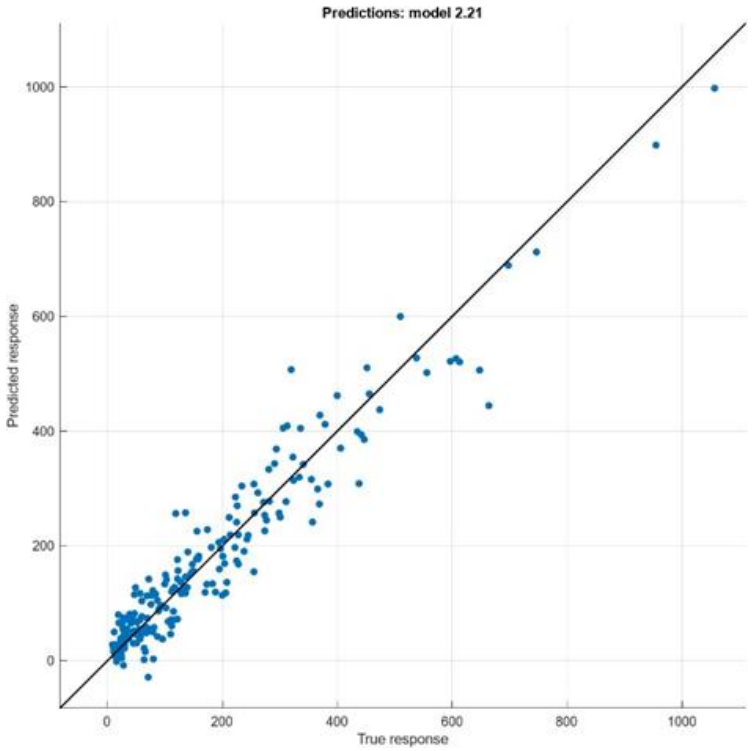


Figure 4. Visualization of actual and predicted responses in test data (psi).

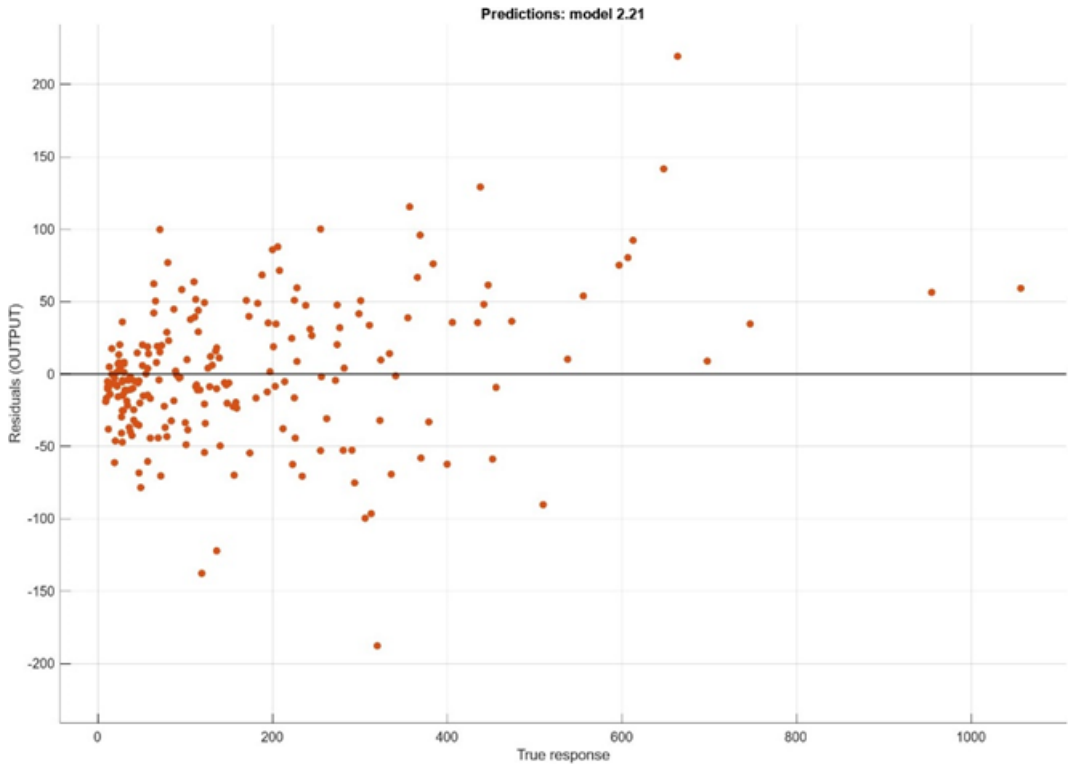


Figure 5. Analysis of error residuals in test data: Discrepancies between true response and model prediction (psi).

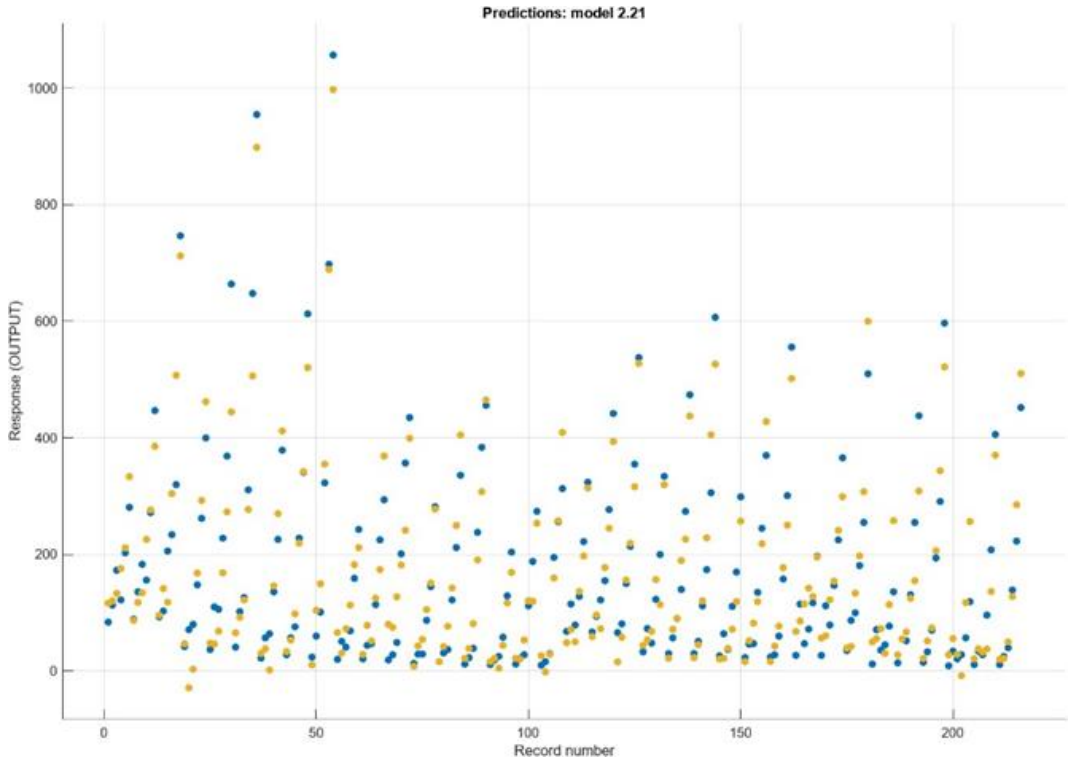


Figure 6. After training, the RMSE (root mean square error) error metric and the system response plot (blue: original output, yellow: model prediction), (psi).

The close alignment between the predicted output and the actual values, as indicated by the minimal differences depicted in orange, attests to the effectiveness of the artificial intelligence regression model. The model demonstrates a high level of accuracy, with errors consistently maintained at low levels across numerous examples. This graphical representation highlights the robust performance of the model in capturing the intricacies of the compressive stress values during training.

4. Discussion

This study delved into the reliability of artificial neural network (ANN) outputs when trained on existing compressive stress data. The training and validation of ANN results were conducted using compressive stress data obtained from the Geotechnical Laboratory at Iowa State University, focusing on the application of biofuel co-products (BCPs) in soil stabilization. The key findings

and conclusions derived from this investigation are summarized as follows:

The data structure, organized with input and output parameters in columns and sample records in rows, constitutes a 216 × 5 matrix.

Input parameters for the dataset, derived from unconfined compressive stress parameters, include soil type, sample type, admixture rate, time, and optimum moisture content. The output parameter represents pressure compressive stress values obtained from experiments.

Through simulations conducted using the Regression Learner application within the Matlab software package, we compared 27 distinct regression methods, employing k = 5 cross-validation for a comprehensive assessment. The results, presented in Table 4, highlight the top 13 methods based on their success rankings.

The Matlab Regression Learner application in Model 2.23, the training outcomes of the neural network revealed a remarkable R^2 value of 0.93, indicating a superior level of model accuracy. This high R^2 value suggests that the neural network effectively captured the linear trend, with minimal deviation of data points from the expected values. The exceptional performance of Model 2.23 underscores its efficacy in the regression task, emphasizing its potential as a robust method for the dataset under consideration.

The most favorable test results were achieved using the neural network method, with error metrics for the model calculated as follows: Mean Absolute Error (MAE) 34.6540, R-Squared (R^2) 0.93, and Root Mean Square Error (RMSE) 47.9080.

5. Conclusion

In summary, the integration of artificial intelligence (AI) has demonstrated exceptional efficacy in predicting system output values for test data, underscored by the noteworthy R^2 value of 0.93, indicating a robust correlation. This achievement holds great promise for the realm of geotechnical engineering, particularly in contexts where experiments are time-consuming and necessitate a large number of trials. The established success of employing artificial neural networks (ANNs) to interpret unconfined compressive test data signifies a considerable advantage for researchers. This approach allows for the efficient acquisition of substantial data with fewer experimental trials and in significantly reduced time frames.

The integration of Artificial Neural Networks (ANN) not only contributes to heightened prediction accuracy but also emerges as a transformative tool in geotechnical engineering research, effectively mitigating the challenges associated with labor-intensive and resource-intensive traditional experiments. The results presented in this study underscore the broader implications of AI-driven solutions, suggesting a paradigm shift in the way data collection and analysis are approached in geotechnical engineering. This technology offers a more efficient and expedited means of extracting meaningful insights, potentially reshaping the methodology for future projects within the field.

Beyond the immediate benefits observed in this study, the broader implications of AI methodologies in geotechnical engineering are promising. As technological advancements persist, the seamless integration of AI is poised to play a pivotal role in shaping the future landscape of geotechnical research and practice. This not only extends to improved accuracy in predictions but also entails a fundamental redefinition of the research and experimentation processes, opening avenues for innovation and efficiency. The evolving synergy between AI and geotechnical engineering holds the potential to usher in a new era of scientific inquiry and problem-solving in the field.

Acknowledgement

The authors gratefully acknowledge the Iowa Highway Research Board and Iowa State University (ISU) for supporting this study.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Ren, J., & Sun, X. (2023). Prediction of ultimate bearing capacity of pile foundation based on two optimization algorithm models. Buildings, 13(5), 1242. https://doi.org/10.3390/buildings13051242
2. Jaksa, M. B. (1995). The influence of spatial variability on the geotechnical design properties of a stiff, overconsolidated clay [Doctoral dissertation, University of Adelaide].
3. Hubick, K. T. (1992). Artificial neural networks in Australia, Technology and Commerce; Commonwealth of Australia: Canberra, Australia.
4. Chao, Z., Ma, G., Zhang, Y., Zhu, Y., & Hu, H. (2018, November). The application of artificial neural network in geotechnical engineering. In IOP conference series: Earth and environmental science, 189, 022054. https://doi.org/10.1088/1755-1315/189/2/022054
5. Kesikoğlu, M. H., Cicekli, S. Y., & Kaynak, T. (2020). The identification of seasonal coastline changes from landsat 8 satellite data using artificial neural networks and k-nearest neighbor. Turkish Journal of Engineering, 4(1), 47-56. https://doi.org/10.31127/tuje.599359
6. Shahin, M. A., Jaksa, M. B., & Maier, H. R. (2001). Artificial neural network applications in geotechnical engineering. Australian Geomechanics, 36(1), 49-62.
7. Das, S. K., & Basudhar, P. K. (2006). Undrained lateral load capacity of piles in clay using artificial neural network. Computers and Geotechnics, 33(8), 454-459. https://doi.org/10.1016/j.compgeo.2006.08.006
8. GOH, A. C. (1994). Nonlinear modelling in geotechnical engineering using neural networks. Transactions of the Institution of Engineers, Australia. Civil engineering, 36(4), 293-297.

9. Ghaboussi, J., & Sidarta, D. E. (1998). New nested adaptive neural networks (NANN) for constitutive modeling. Computers and Geotechnics, 22(1), 29-52. https://doi.org/10.1016/S0266-352X(97)00034-7
10. Das, S. K., & Basudhar, P. K. (2005). Prediction of coefficient of lateral earth pressure using artificial neural networks. Electronic Journal of Geotechnical Engineering, 10.
11. Shahin, M., Jaksa, M., & Maier, H. (2009). Recent advances and future challenges for artificial neural systems in geotechnical engineering applications. Advances in Artificial Neural Systems.
12. Yang, X. S., Gandomi, A. H., Talatahari, S., & Alavi, A. H. (Eds.). (2012). Metaheuristics in water, geotechnical and transport engineering. Newnes.
13. Grima, M. A., & Babuška, R. (1999). Fuzzy model for the prediction of unconfined compressive strength of rock samples. International Journal of Rock Mechanics and Mining Sciences, 36(3), 339-349. https://doi.org/10.1016/S0148-9062(99)00007-8
14. Ceylan, H., Gopalakrishnan, K., & Kim, S. (2010). Soil stabilization with bioenergy coproduct. Transportation Research Record, 2186(1), 130-137. https://doi.org/10.3141/2186-14
15. Vardhan, H., Bordoloi, S., Garg, A., & Garg, A. (2017). Compressive strength analysis of soil reinforced with fiber extracted from water hyacinth. Engineering Computations, 34(2), 330-342. https://doi.org/10.1108/EC-09-2015-0267
16. Öztürk, O., & Türköz, M. (2022). Effect of silica fume on the undrained strength parameters of dispersive. Turkish Journal of Engineering, 6(4), 293-299. https://doi.org/10.31127/tuje.1001413
17. Yang, B. (2015). Performance of bio-based soil stabilizers in transportation earthworks-laboratory investigations. [Master's Thesis, Iowa State University].
18. Uzer, A. U. (2015). Use of biofuel co-product for pavement geo-materials stabilization. Procedia Engineering, 125, 685-691. https://doi.org/10.1016/j.proeng.2015.11.106
19. ASTM D2166. (2006). Standard test method for unconfined compressive strength of cohesive soil. In Annual Book of ASTM standards. West Conshohocken: ASTM International.
20. Zhang, T., Cai, G., Liu, S., & Puppala, A. J. (2014). Stabilization of silt using a lignin-based bioenergy coproduct. Transportation Research Board 93rd Annual MeetingTransportation Research Board.
21. Zhang, T., Liu, S., Cai, G., & Puppala, A. J. (2015). Experimental investigation of thermal and mechanical properties of lignin treated silt. Engineering Geology, 196, 1-11. https://doi.org/10.1016/j.enggeo.2015.07.003
22. Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). Applied linear statistical models. 4th Edition, WCB McGraw-Hill, New York.
23. Seber, G. A., & Lee, A. J. (2012). Linear regression analysis. John Wiley & Sons.
24. Freedman, D. A. (2009). Statistical models: theory and practice. Cambridge University Press.
25. Muggeo, V. M. (2003). Estimating regression models with unknown break-points. Statistics in Medicine, 22(19), 3055-3071. https://doi.org/10.1002/sim.1545
26. Hemanth, D. J., Gupta, D., & Balas, V. E. (Eds.). (2019). Intelligent Data Analysis for Biomedical Applications: Challenges and Solutions. Academic Press.
27. https://dirask.com/posts/JavaScriptartificial-neuron-model-paoM31
28. Draper, N. R., & Smith, H. (1998). Applied regression analysis (Vol. 326). John Wiley & Sons.
29. Glantz, S. A., Slinker, B. K., & Neilands, T. B. (2001). Primer of applied regression & analysis of variance, ed (Vol. 654). McGraw-Hill, Inc., New York.
30. Demir, V., & Doğu, R. (2024). Prediction of elevation points using three different heuristic regression techniques. Turkish Journal of Engineering, 8(1), 56-64. https://doi.org/10.31127/tuje.1257847

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Adsorption-desorption of glyphosate in tropical sandy soil exposed to burning or applied with agricultural waste

Jamilu Garba [*1], **Abd Wahid Samsuri** [2], **Muhammad Saiful Ahmad Hamdani** [3], **Tariq Faruq Sadiq** [4]

[1] *Ahmadu Bello University, Department of Soil Science, Nigeria, jamilugarba96@gmail.com*
[2] *Universiti Putra Malaysia, Department of Land Management, Malaysia, samsuriaw@upm.edu.my*
[3] *Universiti Putra Malaysia, Department of Crop Science, Malaysia, s_ahmad@upm.edu.my*
4 *Salahaddin University, Department of Soil and Water, Iraq, ta.fa1981@gmail.com*

Cite this study: Garba, J., Samsuri, A. W., Hamdani, M. S. A., & Sadiq, T. F. (2024). Adsorption-desorption of glyphosate in tropical sandy soil exposed to burning or applied with agricultural waste. Turkish Journal of Engineering, 8 (3), 469-482

https://doi.org/10.31127/tuje.1428763

Keywords
Glyphosate
Sandy soil
Sorption
Mobility
Contamination

Research Article

Received: 30.01.2024
Revised: 20.03.2024
Accepted: 21.03.2024
Published: 05.07.2024

Abstract

The present study investigated glyphosate adsorption-desorption in Malaysian sandy soil exposed to burning or applied with cow dung or rice husk ash (RHA). One gram each of the adsorbents (control, burnt soil, soil + cow dung, soil + RHA) was separately and in replicate mixed with solution of 20 mL of glyphosate at different initial concentrations (0, 25, 50, 100, 150, 200, 250 and 300 mg L^{-1}), shaken for 24 hours, centrifuged at 10, 000 rpm for 10 minutes and later the decanted supernatants were collected. Desorption study follow immediately through addition of 20 mL of 0.01 M $CaCl_2$ solution and then treated like adsorption study. Glyphosate analysis was done using high performance liquid chromatography with fluorescence detector and isotherm data was fitted to linear, Freundlich, Langmuir and Temkin models. Freundlich best fits the adsorption of glyphosate and application of cow dung or RHA enhanced the soil affinity for glyphosate. Freundlich isotherm constant (K_F) for soils treated with RHA and cow dung were 9.768 mg g^{-1} and 6. 751 mg g^{-1}, respectively as compared to 3.189 mg g^{-1} for the control soil. However, the greatest K_F value (387.238 mg g^{-1}) was recorded by burnt Benta soil. Glyphosate adsorption by both control and treated soils was favourable ($0.044 < R_L < 0.3$). This study suggests that adsorption occurred through physical processes involving diffusion, complexation or ligand exchange. Glyphosate desorption from soils treated with cow dung and RHA was either hysteretic or reversible. Burning this soil should be discouraged due to its effect on increasing glyphosate mobility and possible groundwater contamination. Meanwhile, agricultural waste can be applied to this soil even though it might lead to an increased glyphosate mobility but because of its potential positive effect on the soil's biological functions, the glyphosate ions are bound to be degraded.

1. Introduction

Open-field burning and application of organic manure are widely practiced worldwide. As part of land clearing, Malaysian paddy farmers routinely burn their rice straws *in situ* before mechanized land preparations [1]. There are two seasons of rice cultivation in Malaysia; off-season (mid-March to mid-August) and main season (mid-September to mid-January). Therefore, once harvesting is completed during the off-season farmers are left with a very limited time—around a month and a half— to prepare for the main season [1-2]. This makes open-field burning the easiest way to dispose these straws and save time for field management for the next cropping season. It is also done to destroy sources of pests and diseases [3]. Open burning can result in ash deposits due to oxidation of the rice straws and native soil organic matter which can lead to an increase in pH, addition of Na^+ and K^+, changes in soil minerals and exchangeable cations [4]. However, the burning practice causes severe environmental pollution and land degradation because it contributes to the emission of greenhouse gases such as carbon monoxide, methane and nitrous oxide [3].

To overcome the menace of open burning of soil and agricultural waste a popular waste management approach known as zero waste and zero burning is being advocated [1]. This approach employs direct incorporation of these waste or their compost into the soil as organic fertilizers [1,3,5,6]. Another alternative

includes using these waste for animal feeds, industrial raw materials, biofuel and off-site soil and water quality improvement [1,3,6]. Cow dung and rice husk ash are other agricultural waste abundantly found in Malaysia. An annual estimate of cow dung in Malaysia is 640, 650 tonnes (FAOSTAT, 2017) while annual production of rice husk ash is 1, 200 metric tonnes $mill^{-1}$ [7]. These waste as mentioned earlier are used as industrial raw materials for energy source, production of biochar and/or applied to the soil as organic fertilizers. The advantages of applying organic fertilizers to soils include (i) increase soil biological action which enhances nutrient mobilization and decomposition of toxic substances; (ii) increase soil organic matter which improve exchange and retention capacities of nutrients and toxic substances; and (iii) increase soil aggregate stability and buffering capacity which promote root growth, P supply and increasing soil water retention [5].

Advocating for organic farming formed part of the plans in the Third National Agriculture Policy (NAP3) of Malaysia [5,8]. It involves refraining farmers from the use of chemical fertilizers, pesticide and pharmaceutical, and encouraging them to use organic input such as organic fertilizers and biological pest control for sustained productivity. This campaign was facilitated by launching of Malaysian organic certification program or *Sijil Organic Malaysia* (SOM) in 2003 [5] which helps in increasing attractiveness of Malaysian organic produce in both local and foreign markets. This is through their certification base on prerequisites of the Malaysian standard MS1529. A number of farmers adapted to this practice, however, Tiraieyari et al. [8] reported that majority are not practicing the method and the reason for this includes issues of temporary occupation licenses (TOLs) of lands, expensive and long certification process and inadequate financial support from the government [9]. Alternatively, they apply chemicals as the best solution or employed an integrated approach involving the use of organic and inorganic input. The common basic agrochemicals used in Malaysia are fertilizers, fungicides, insecticides and herbicides. Raja Abdul [10] reported that in Malaysia weed problems are being solve by using herbicide without damage crops and increasing affordable cost. Glyphosate {N (phosphonomethyl) glycine} is among the common herbicides used in weed control on crops and open area. Glyphosate is a post-emergence weed killer consisting of three functional groups-phosphonic, carboxylic amines (Figure 1) and is capable of inhibiting protein synthesis in plant. Glyphosate physico- chemical properties are shown in Table 1. It is water soluble but insoluble in other solvent, odourless with average half-life of 47 days. Furthermore, it has low molecular weight, high surface tension, polarity and affinity as a result, it is non-volatile. It is applied on foliage, therefore, becomes non-active in soil due to its strong affinity soil and organic matter and oxides minerals [11]. Nonetheless, residues of glyphosate were discovered in surface [12,13] and underground water [14], which indicated the influence of soil properties and composition on adsorption and mobility of glyphosate. The main metabolic pathway of glyphosate in soil is its fast degradation to aminomethylphosphonic acid (AMPA), as a result, it is considered less toxic [15]. However, it was recently classified as probably carcinogenic to humans [16]. This makes the experts on pesticide residues in food and environment to re-consider glyphosate and AMPA for toxicological interest.

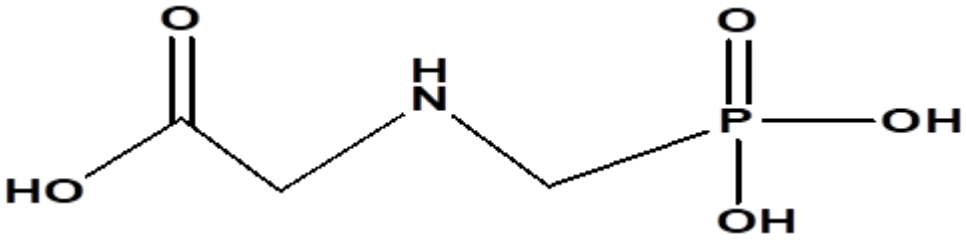


Figure 1. Chemical structure of glyphosate.

Table 1. Physical and chemical properties of glyphosate.

Common name	Glyphosate
Synonyms	N-(Phosphonomethyl) glycine
Chemical formula	$C_3 H_8 NO_5 P$
Molecular weight	169.09
Physical state and color	Crystalline powder, white
Density	1.704 (25°C)
Melting point	200-230°C
Boiling point	Decomposes at 187°C
Water solubility	10,000-15,700 mg L^{-1} at 25°C
Half life	1-174 days
Mobility potential	Very low
Toxicity class	III

In this study, two soil management scenarios were created in the laboratory to imitate open-field burning and application of agricultural waste. They are burning the soil to remove the biota and adding cow dung or rice husk ash to the Malaysian sandy soil. It should be noted that as far as we know the only two papers [4,17] on glyphosate adsorption in soil after heating or removal of organic matter were on temperate soils. Furthermore, both papers did not give an account on the effect of additional organic matter on glyphosate sorption much less its mobility. Therefore, a similar study on tropical soils and taken into account the effect of additional

organic waste will help in increasing the knowledge of adsorption and mobility of glyphosate in soils. The aim of this study was to investigate the adsorption-desorption and potential mobility of glyphosate herbicide in soils exposed to burning or treated with either cow dung or rice husk ash (RHA).

2. Method

2.1 Sampling and chemicals

The present study used Benta soil series which sampled from Sementa Hulu (Lat. 3.841663 ^{0}N, Long. 101.947251 ^{0}E), located at Raub district, Malaysia. The soil was classified as Ultic Hapludalf based on the USDA soil taxonomy [18]. The soil sampling involves random collection of surface soils (0-20 cm) from five different sports which were later bulked to one composite sample. Cow dung was sampled from the animal section of the experimental farm (lat 2° 59′ 28.7″ N, long 101° 42′ 52.9″ E), Faculty of Agriculture, Universiti Putra Malaysia. Conversely, the rice husk ash (RHA) was collected from BERNAS Rice Mill Selangor, Malaysia (Lat. 3°40'32.4"N, Long. 100°59'42.5"E). Both the soil samples and that of agricultural waste were air dried, ground and sieved (2-mm sieve for the soil sample while 1-mm sieve for the cow dung and rice husk ash) then kept in a plastic container before analysis and further studies.

All solutions used in the present study were prepared using Millipore® Direct UV-Q water. Glyphosate (99.7% purity) and 9-flourenylmethlylchloroformate chloride (FMOC-Cl) of 97% purity were purchased from Sigma Aldrich® (Seelze, Germany) whereas acetonitrile was purchased from QREC®, Malaysia. Analytical grades of $CaCl_2$ and $HgCl_2$ were procured from Emsure® Germany and Sigma Aldrich® (India) respectively. The stock solution (500 mg L^{-1}) of glyphosate was prepared by dissolving 25 mg of its analytical grade powder into 50 mL Millipore® Direct UV-Q water. Working solutions for batch equilibrium sorption study were prepared by diluting appropriate stock solutions with Millipore® Direct UV-Q water.

2.2 Soil characterization, burning and application of agricultural waste

Details of soil characterization and its mineralogical determination has been previously `described in Garba et al. [19]. The burnt Benta soil was prepared according to the method described by [17] whereby about 15 grams of the soil sample was weighed into a porcelain crucible then ignited into a F62700 muffle furnace (Electrothermal Engineering, Essex, UK) at 550^{0}C for 2 hours and later allowed to cool at room temperature. The cow dung and RHA were applied at a rate of 10 tonnes ha^{-1} by mixing 1 g of the soil with 0.1 g of cow dung or RHA in a centrifuge tube. For homogeneity, the mixture was incubated for 1 week by maintaining its moisture content at field capacity throughout. After which the soil mixture was kept at room temperature to dry up prior to sorption study.

2.3 Batch equilibrium sorption experiment

Four set of soils in three replicates were used for this experiment and they are 1) the untreated Benta soil as control, 2) burnt Benta soil, 3) Benta soil applied with cow dung and 4) Benta soil amended with RHA. The sorption experiment was done according to the method described by Piccolo et al. [20] with some changes. Briefly, 1 g each of the soil samples were added into 50 mL centrifuge tubes and then a solution of 20 mL of glyphosate at different initial concentrations (0, 25, 50, 100, 150, 200, 250 and 300 mg L^{-1}) were added. The glyphosate solution was prepared in 0.01 M $CaCl_2$ solution consist of 200 mg L^{-1} $HgCl_2$ stand-in as a bioinhibitor. The centrifuge tubes containing the soil mixture and glyphosate solutions were putting into a rotary shaker and shaken for 24 hours at 100 rpm under room temperature. Afterward, the tubes were centrifuged at 10,000 rpm for 10 minutes using a Sartoruis 4–16 centrifugal machine (Sartoruis, Góttingen, Germany). Then, the supernatants were decanted and passed through a 0.45 µm P0377 HmbG syringe filter (Johchem, Malaysia) before analysis. Desorption study follows immediately after decantation of supernatants of the adsorption study. Thus, 20 mL of the background solution (0.01M $CaCl_2$ and 200 mgL^{-1} $HgCl_2$) was put into each of the centrifuge tubes and then treated like adsorption study. The glyphosate analysis in both adsorption and desorption samples was done using a high performance liquid chromatography with fluorescence detector (HPLC-FLD) method previously described by Garba et al. [21]. In brief, the filtered supernatant containing glyphosate was derivatized with 9-flourenylmethlylchloroformate chloride (FMOC-Cl). The glyphosate-FMOC derivatives were then analysed by HPLC-FLD using C18 Agilent® Zorbax Eclipse plus column (4.6 × 150 mm, 5 µm). The mobile phase consists a mixture of acetonitrile and a solution of 0.05 M KH_2PO_4 (30:70; v/v). The flow rate was 0.7 mL min^{-1} using an isocratic mode and injection volume of 20 µL at 40 °C column temperature. The method's excitation and emission wavelengths were 270 nm and 315 nm, respectively. Likewise, the method's lowest limits of detection and quantification were 0.021 mg kg^{-1} and 0.064 mg kg^{-1}, respectively.

2.4 Adsorption-desorption isotherms

The equilibrium concentration of glyphosate [q_e (mg/g)] was calculated using the Equation 1.

$$q_e = \frac{(C_i - C_e)\, x\, V}{W} \quad (1)$$

Likewise, the glyphosate removal efficiency was calculated using the Equation 2.

$$Effeciency\ (\%) = \frac{(C_i - C_e)}{C_i}\, x\, 100 \quad (2)$$

Where C_i and C_e are respective initial and equilibrium glyphosate concentrations (mg/L). While, V

is the volume of the solution (L) and W is the weight (g) of the soil used.

The adsorption behaviour of the adsorbate on the adsorbent was studied by fitting the isotherm data into the linear, Freundlich, Langmuir and Temkin isotherm models. The Table 2 shows both non-linear and linear equations of these isotherm models.

Table 2. Isotherm models and definition of their constants.

Model	Non-linear equation	Linear equation	plot	Constants and their definition
Linear		$q_e = K_d C_e + m$	q_e vs C_e	K_d: partitioning coefficient m: constant related to improved capability
Freundlich	$q_e = K_f {C_e}^{1/n}$	$\log q_e = \log K_f + \frac{1}{n}\log C_e$	log q_e vs log C_e	K_F: adsorption affinity 1/n: adsorption intensity
Langmuir	$q_e = \frac{Q_{max} b C_e}{1 + bC_e}$	$\frac{C_e}{q_e} = \frac{1}{Q_{max}} C_e + \frac{1}{Q_{max} b}$	$\frac{C_e}{q_e}$ vs C_e	Q_{max}: constant related monolayer coverage b : constant related to adsorption affinity
Temkin	$q_e = \frac{RT}{b} \ln (AC_e)$	$q_e = \frac{RT}{b} lnA + \frac{RT}{b} lnC_e$	q_e vs ln C_e	A: constant related to binding energy b: constant related to heat of adsorption

The most suitable isotherm model to represent the experimental data was determined using Marquardt's percent standard deviation (MPSD) (Equation 3):

$$MPSD = \sqrt{\frac{1}{n-p}\sum_{i=1}^{n}\left(\frac{(q_{e,exp} - q_{e,calc})}{q_{e,exp}}\right)_i^2} \quad (3)$$

Where n denote the number of data points, p denote the number of parameters for a given model, $q_{e,exp}$ and $q_{e,calc}$ are the experimental and calculated concentrations of the adsorbate at equilibrium (mg/g), respectively.

The hysteresis index (HI) which describes the nature of adsorption and desorption processes was determined from the values of constant n of Freundlich model for adsorption and desorption of glyphosate (Equation 4):

$$HI = \frac{n\ desorption}{n\ adsorption} \quad (4)$$

The separation factor (R_L) was calculated from the constant of the Langmuir isotherm model and it describes the favorability of the glyphosate adsorption by the soil samples, thus (Equation 5):

$$R_L = \frac{1}{1 + bCo} \quad (5)$$

The mobility of glyphosate was predicted using ground water ubiquity score (GUS) Equation 6.

$$GUS = log_{10}(t_{1/2}soil)\ x\ (4 - log_{10}(k_{oc})) \quad (6)$$

Where $t_{1/2}$ is the glyphosate half-life in soil and K_{oc} stand for organic carbon normalized adsorption coefficient. The K_{oc} was determined from the partitioning coefficient (K_d constant) of the linear isotherm model by using the Equation 7.

$$K_{oc} = \frac{K_d}{\%\ soil\ organic\ carbon} \quad (7)$$

Gustafson [22] reported that, GUS values are universal accepted indices for estimating mobility and persistence of pesticide in soil.

3. Results and discussion

3.1 Soil characteristics

The physicochemical properties of the soil are shown in Table 3. The soil texture was sandy as it had a high sand content (74.17%) but low in silt (5.83%) and clay (20%). The soil had 45% porosity and a bulk density of 1.45 g cm^{-1}, meanwhile it has 23. 88 % moisture content at field capacity. The EC value of the soil was 0.024 dS cm^{-1} and the pH was near neutral (6.73). The organic matter content was 5.23% and the soil cation exchange capacity (CEC) of the soil was 12.677 $cmol_{(+)}$ kg^{-1}. The values of percent free and amorphous oxides of Fe are 0.473 and 0.218, respectively while respective values of free and amorphous oxides of Al are 0.476% and 0.262%. The X-Ray diffractogram (Figure 2) revealed that the soil minerals were dominated by smectite (1.582 nm), mica (1.029 nm), kaolinite (0.720 nm, 0.358 nm), quartz (0.334 nm) and magnetite (0.201nm).

Benta soil was reported to have developed from andesite and rhyolite [23], thus, containing high amount of kaolinite and quartz leading to high sand fraction. The authors further reported that due to intense rainfall and high temperature in Malaysia, chemical weathering dominated the soil forming processes. Therefore, considering the land's undulating topography, erosion occurs at unstable slope leading to surface soil removal and this exposed its parent materials, hence, releasing basic cations [24] which resulted into higher soil pH, EC and base saturation as compared to the ultisols and oxisols soil orders in Malaysia. The low exchangeable Al (0.333 $cmol_{(+)}$ kg^{-1}) led to low Al saturation in the soil. At near neutral pH, most of the Al precipitated as $Al(OH)_3$ and therefore the exchange sites were dominated by the basic cations.

The author also attributed the low available P and higher Fe and Mn content to the soil parent materials. A ratio of amorphous/free iron oxides of > 0.05 (0.46 for the present study) shows that the soil is very active and low in crystallinity which was confirmed by the presence of mica and smectite. A soil bulk density of >1 g cm^{-3} is an indication of the presence of argillic horizon hence, according the USDA soil taxonomy classification this soil was classified as Alfisols [18].

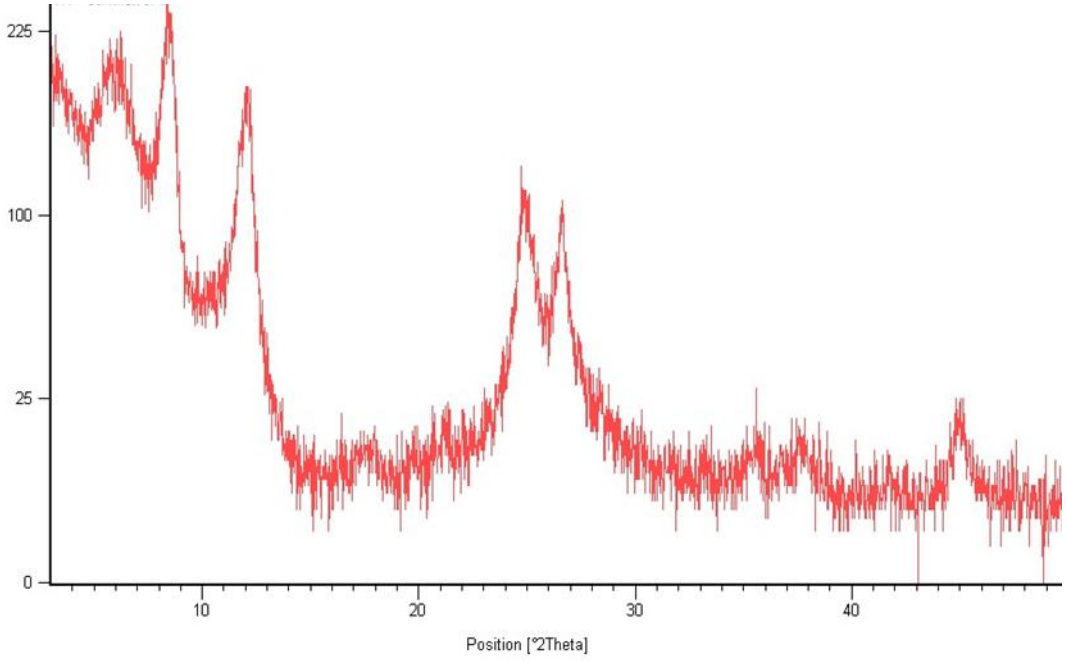


Figure 2. X-ray diffractogram of the soil used in the present study.

Table 3. Properties of the soil used in the present study.

Parameter	Result
Sand (%)	74.17 ± 0.83
Silt (%)	5.83 ± 0.83
Clay (%)	20.00 ± 1.43
Soil type	Sandy soil
Moisture content (%)	23.88 ± 0.89
D_b (g cm $^{-3}$)	1.45 ± 0.14
Porosity (%)	45.00 ± 0.33
pH	6.73 ± 0.03
Ec (dS cm^{-1})	0.024 ± 0.001
Exchangeable Al ($cmol_{(+)}$ kg^{-1})	0.333 ± 0.07
Exchangeable H ($cmol_{(+)}$ kg^{-1})	0.333 ± 0.13
SOM (%)	5.23 ± 0.12
C (%)	1.665 ± 0.00
N (%)	0.164 ± 0.00
P (mg kg^{-1})	6.953 ± 0.16
S (mg kg^{-1})	60.100 ± 1.01
Exchangeable bases ($cmol_{(+)}$ kg $^{-1}$)	
Ca	5.334 ± 0.64
K	0.331 ± 0.02
Mg	1.332 ± 0.07
Na	0.065 ± 0.00
CEC ($cmol_{(+)}$ kg^{-1})	12.667 ± 0.67
Micronutrients (mg kg^{-1})	
Cu	5.600 ± 0.10
Fe	13529 ± 22.94
Mn	547.700 ± 4.22
Zn	87.000 ± 2.16
Free FeO (%)	0.473± 0.01
Amorphous FeO (%)	0.218 ± 0.02
Free AlO (%)	0.476 ± 0.01
Amorphous AlO (%)	0.262 ± 0.09

3.2 Glyphosate removal efficiency

The result of glyphosate removal efficiency from the aqueous solution by the various adsorbent is shown in Table 4. It generally shows that, control and soils amended with the agricultural residues had greater removal efficiencies than burnt soil. Furthermore, an increasing removal efficiency was observed from all the absorbents due to increasing glyphosate initial concentration, even though, the low percent glyphosate removal (28.933%) was observed from burnt soil at 300 mg L^{-1}. By taken the concentration as a whole, high percent removal (89.516%) was obtained from the soil applied with rice husk ash. This was followed by the soil amended with cow dung (86.516%) and control (86.045%), while it was less in the burnt soil (76.304%).

Therefore, burning this soil resulted into 9.746% decreased in glyphosate removal compared to the control soil. Meanwhile, amending this soil with cow dung or rice husk ash led to 0.568% and 3.471% increase of glyphosate removal, respectively.

The present result indicates the influence of organic matter on glyphosate adsorption by this soil. Therefore, open burning with aimed of land clearing can lead to oxidation of its organic matter content. The consequence of this is decreasing glyphosate adsorption capacity and its more desorption from this soil which can results to its mobility and possible ground water contamination. Equally, incorporation/external supplement of organic residues can increase the soil adsorption surfaces, thus, enhanced its glyphosate adsorption capacity. Soil sorption is a complex process because its total adsorption capacity is a function of the summation of individual sorption capacities of various soil components and properties [25]. These components are clay, organic matter and soil minerals content, while the profound soil properties affecting its adsorption are texture, CEC and pH. Therefore, anything affecting these components or properties is expected to affect soil sorption capacity.

Table 4. Glyphosate removal efficiency by the control and amended soils.

Initial concentration (mg l^{-1})	Glyphosate removal efficiency (%)			
	Control	Burnt soil	Soil with cow dung	Soil with rice husk ash
0	0	0	0	0
25	73.915	68.868	71.758	81.678
50	79.839	77.702	85.302	84.957
100	86.208	85.967	87.477	89.445
150	88.544	89.063	88.681	91.033
200	90.462	91.628	90.366	92.549
250	91.242	91.965	90.614	93.105
300	92.103	28.933	92.094	93.843
Mean	86.045	76.304	86.613	89.516

Fire intensity is generally described using fire temperature, therefore, low, medium and high fire intensities had a respective fire temperature of ≤ 200°C, ≤ 400°C and > 675°C [26]. Soil heating to as low as < 250 affect soil biological and chemical properties but the severity prevails more at temperature > 500°C [27,28]. Heating a sandy soil at a temperature above 450°C decreases it aggregate strength [28] which can resulted to a sharp increase in sand particles but decrease in silt and clay. Similarly, the authors reported total loss of soil organic matter, a decrease in kaolinite and resultant increase in pH. The trio significantly affect glyphosate adsorption and mobility in soil, thus, any change on them results in changing soil glyphosate adsorption capacity. The heating temperature in present study is 550°C which falls within the temperatures used by Araya et al [28], suggesting the total loss of organic matter, sharp increase in sand particle, decrease in CEC, dehydration of clay minerals and resultant pH increased due to K^+ and Na^+ ions from organic matter oxidation. The effect of these were suggested to result in low glyphosate removal efficiency by the burnt soil. Meanwhile applying cow dung or rice husk ash can to result to increasing organic matter content of this soil, and so, upsurge its surface charges and enhances adsorption strength [17,29], hence, greater glyphosate removal efficiency. Our previous study on the same soil showed an increased in adsorptive removal of aminomethylphosphonic acid (AMPA) - a metabolite of glyphosate and having the same soil adsorption site- due to amendment with cow dung or rice husk ash [30].

3.3 Glyphosate adsorption isotherms

The adsorption isotherms for glyphosate on different adsorbents were present in Figure 3. The isotherms obtained for all adsorbents were of "C" curve, thus, indicating a constant partitioning of glyphosate amongst the interfacial region of surfaces of the adsorbent and bulk solution. The C-curve isotherm is characterized by an initial slope of adsorptive concentration that remains independently until the maximum adsorption is achieved [31]. From the present result it can be hypnotized that, even with soil burning or application of organic manure glyphosate can be adsorb by the soil of the present study through entrapment in the intra and inter surfaces, therefore, no any binding energy between these surfaces and glyphosate ions.

The glyphosate adsorption isotherm data for both adsorbent were fitted to the four different models listed in Table 2 and the results of these model constant were shown in Table 5. Of all the models, control soil best fitted to Freundlich (r^2= 0.997) which followed by linear (r^2= 0.964), Langmuir (r^2= 0.905) and Temkin (r^2= 0.871). The soil applied with rice husk ash also followed a similar trend, however, the soil + cow dung has the trend of linear > Freundlich > Temkin > Langmuir. Meanwhile, the isotherm data for glyphosate adsorption on the burnt soil very poorly fitted to all models except Langmuir (r^2= 0.959). The soil applied with rice husk ash (178.386 mg g^{-1}) and control (151.325 mg g^{-1}) had higher values of K_d constant of the linear model compared to soil + cow dung (127.983 mg g^{-1}) and burnt soil (0.888 mg g^{-1}). The K_F constant of Freundlich isotherm model was higher in the burnt soil (387.238 mg g^{-1}), then soil amended with rice husk ash (9.768 mg g^{-1}), soil amended with cow dung (6.751 mg g^{-1}) and control was the least (3.187 mg g^{-1}). Meanwhile, the values of 1/n constant of the model showed the trend of control > soil + rice husk ash > soil + cow dung > burnt soil. The burnt soil had higher Q_{max} (233.739 mg g^{-1}) of the Langmuir model. This followed by the soil applied with rice husk ash (36.649 mg g^{-1}), then soil + cow dung (27.937 mg g^{-1}) and control (24.374 mg g^{-1}). The values of b constant have the following trend soil

+ cow dung < control < soil + rice husk ash < burnt soil. The calculated Langmuir separation factor (R_L) for both adsorbents is < 1 and so, ranged from 0.194 to 0.447. The Temkin isotherm constant A for the adsorbents had the following trend burnt soil > soil + rice husk ash > soil + cow dung > control soil. Similar pattern was observed from the b constant which had ranged of 1.275 to 26.585 j mole^{-1}. The result of the calculated MPSD revealed that, linear model had values range of 0.040 to 8.377, Freundlich; 0341 to 0.430, Langmuir; 0.447 and Temkin; 5.663 to 95.310.

It is obvious, the best fitting relationship that measure the distribution of the adsorbate is describe by the correlation coefficient (r^2). It is also used in the corroboration of the consistency of the isotherm models, hence, describe the association between the experimental isotherm data and linear form of the adsorption isotherm model. The major defect of r^2 is that, it does not take into account the changes in error that occur during transformation of the isotherm models into linearized form. This makes it necessary to use an error function analysis to find out the model with best description of the experimental isotherm data. The MPSD is one of the error function use to determine the suitability of the isotherm model to describe experimental data [32]. Therefore, any model having less MPSD values fits the experimental isotherm data best. The order of Freundlich > Langmuir > linear > Temkin was emanated from MPSD values of the present study.

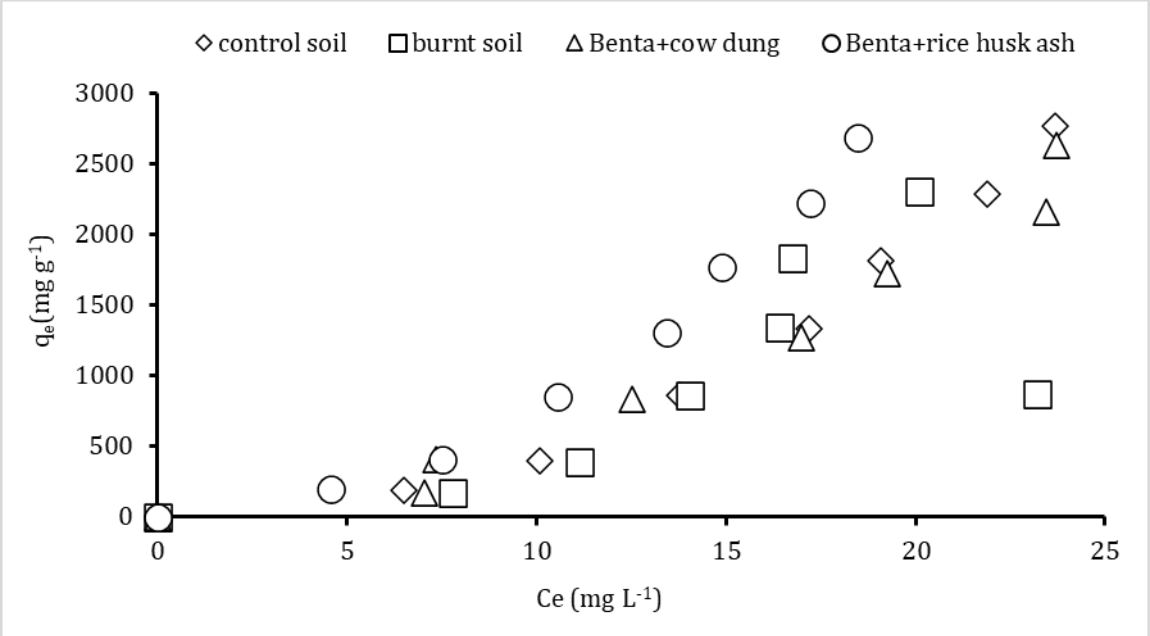


Figure 3. Glyphosate adsorption isotherms for the different adsorbents.

The glyphosate adsorption isotherm data for both adsorbent were fitted to the four different models listed in Table 2 and the results of these model constant were shown in Table 5. Of all the models, control soil best fitted to Freundlich (r^2= 0.997) which followed by linear (r^2= 0.964), Langmuir (r^2= 0.905) and Temkin (r^2= 0.871). The soil applied with rice husk ash also followed a similar trend, however, the soil + cow dung has the trend of linear > Freundlich > Temkin > Langmuir. Meanwhile, the isotherm data for glyphosate adsorption on the burnt soil very poorly fitted to all models except Langmuir (r^2= 0.959). The soil applied with rice husk ash (178.386 mg g^{-1}) and control (151.325 mg g^{-1}) had higher values of K_d constant of the linear model compared to soil + cow dung (127.983 mg g^{-1}) and burnt soil (0.888 mg g^{-1}). The K_F constant of Freundlich isotherm model was higher in the burnt soil (387.238 mg g^{-1}), then soil amended with rice husk ash (9.768 mg g^{-1}), soil amended with cow dung (6.751 mg g^{-1}) and control was the least (3.187 mg g^{-1}). Meanwhile, the values of 1/n constant of the model showed the trend of control > soil + rice husk ash > soil + cow dung > burnt soil. The burnt soil had higher Q_{max} (233.739 mg g^{-1}) of the Langmuir model. This followed by the soil applied with rice husk ash (36.649 mg g^{-1}), then soil + cow dung (27.937 mg g^{-1}) and control (24.374 mg g^{-1}). The values of b constant have the following trend soil + cow dung < control < soil + rice husk ash < burnt soil. The calculated Langmuir separation factor (R_L) for both adsorbents is < 1 and so, ranged from 0.194 to 0.447. The Temkin isotherm constant A for the adsorbents had the following trend burnt soil > soil + rice husk ash > soil + cow dung > control soil. Similar pattern was observed from the b constant which had ranged of 1.275 to 26.585 j mole^{-1}. The result of the calculated MPSD revealed that, linear model had values range of 0.040 to 8.377, Freundlich; 0341 to 0.430, Langmuir; 0.447 and Temkin; 5.663 to 95.310.

It is obvious, the best fitting relationship that measure the distribution of the adsorbate is describe by the correlation coefficient (r^2). It is also used in the corroboration of the consistency of the isotherm models, hence, describe the association between the experimental isotherm data and linear form of the adsorption isotherm model. The major defect of r^2 is that, it does not take into account the changes in error that occur during transformation of the isotherm models into linearized form. This makes it necessary to use an error function analysis to find out the model with best

description of the experimental isotherm data. The MPSD is one of the error function use to determine the suitability of the isotherm model to describe experimental data [32]. Therefore, any model having less MPSD values fits the experimental isotherm data best. The order of Freundlich > Langmuir > linear > Temkin was emanated from MPSD values of the present study.

Table 5. Models' constant for glyphosate adsorption by the adsorbents, correlation coefficients and error functions.

Models	Parameters				
Linear	K_d (mg g^{-1})	m	r^2		MPSD
Control soil	151.325	1050.986	0.964		6.767
Burnt soil	0.888	1145.960	0.008		0.040
Soil + cow dung	127.783	702.188	0.963		6.000
Soil + rice husk ash	178.386	865.281	0.959		8.377
Freundlich	K_F (mg g^{-1})	1/n	r^2		MPSD
Control soil	3.189	2.132	0.997		0.341
Burnt soil	387.238	0.255	0.092		0.430
Soil + cow dung	6.751	1.864	0.941		0.349
Soil + rice husk ash	9.768	1.905	0.994		0.350
Langmuir	Q_{max} (mg g^{-1})	b	r^2	R_L	MPSD
Control soil	24.374	0.037	0.905	0.219	0.447
Burnt soil	233.739	0.262	0.959	0.044	0.447
Soil + cow dung	27.937	0.033	0.560	0.234	0.447
Soil + rice husk ash	36.649	0.044	0.940	0.194	0.447
Temkin	A (mg g^{-1})	b (j $mole^{-1}$)	r^2		MPSD
Control soil	4.389	1.275	0.871		95.310
Burnt soil	987.950	26.585	0.017		5.663
Soil + cow dung	6.000	1.463	0.909		87.072
Soil + rice husk ash	7.664	1.448	0.856		86.024

Freundlich isotherm model assumes a heterogeneous surface coverage hence stronger binding site are occupied first and binding strength increases with increasing surface coverage until equilibrium is attained [33]. The K_F and 1/n constant of the model denotes adsorption affinity and intensity, respectively. Even though applying cow dung or rice husk ash an increase affinity (K_F) for glyphosate adsorption by this soil but a greater increased was observed from the burnt soil. This can be attributed to the structural changes of minerals smectite, mica and kaolinite which was reported to occur at 500°C [4,17,25], a degree lower than that of the present study (550°C). This change result to dihydroxylation which increase soil acidity due to release of H^+ and Al^{3+} from these minerals, and so, an increase the net positive charges on the soil surface for greater glyphosate adsorption. Soil pH, mineral group and cations considerably affect glyphosate adsorption. Result of the physicochemical analysis (Table 3) revealed the pH of this soil to be of near neutral (6.763) and a significant content of Fe, Mn and Zn. The resultant ash of burning this soil might increase its pH and contents of these cations and many more, hence, forming a strong complex with glyphosate phosphonic and carboxylic moieties. Glyphosate had amphoteric properties, and so, a good chelating agent therefore it can form strong complexes with metal ion, especially at near-neutral pH levels due to deprotonation of its carboxylate and phosphonate moieties [34]. Unlike K_F values, amending this soil resulted into decrease of 1/n values compared to control, thus, indicating an increasing intensity. The 1/n values of burnt soil was 0.255 and was < 1, accordingly, glyphosate adsorption by the burnt soil was favourable. Meanwhile the 1/n values for both control and soil amended with cow dung or rice husk ash were between 1.864 and 2.132 which denotes a cooperative adsorption.

Langmuir isotherm model assumes monolayer coverage of adsorbate occurring at a homogenous site on adsorbent. It further presumes equal adsorption energy, therefore, no transmigration of adsorbate in the adsorbent's surfaces [33]. The Q_{max} which represent a monolayer coverage and b which denote adsorption affinity are the model constant. In addition to these, the model has a dimensionless separation factor (R_L) which describe it favourability. Based on this R_L, the adsorption of glyphosate was more favourable by soil + cow dung > control soil > soil + rice husk ash > burnt soil. The present study revealed that both application of these agricultural waste or burning its natural organic matter led to an increase in glyphosate adsorption by this soil. The resultant increase in monolayer coverage owing to applying cow dung or rice husk ash can be attributed to additional functional group, bases and more surface area. Our previous study [35] showed that these agricultural waste contains phenols, carboxylic, ethers, Fe, Al and high surface area. Therefore, incorporating of them were suggested to increase the surface area and functional group of the soil leading to an increased glyphosate adsorption. The higher increase in Q_{max} (233.739 mg g^{-1}) obtained from burnt soil was not unexpected because the burning was done in a muffle furnace, thus, is an oxygen-limited condition. This can resulted into charring of the soil organic matter, formation of pyrogenic carbon and more accumulation of aromatic compounds [25] hence, an increase in adsorption capacity of the soil. The higher values of Langmuir constant b obtained from the burnt soil (0.262) indicated higher affinity of glyphosate ions for this soil. This possibly could be due to the lime effect of the resultant ash which can increases the concentration of amorphous Fe and Al oxides [36]. The result of the b values also shows that application of rice husk ash increases glyphosate affinity of the study soil

but, soil applied with cow dung showed a decrease in affinity of this compound.

Linear model is the simplest among the adsorption isotherm model, it assumes constant partitioning of adsorbate in the adsorbent inter facial region, thus, differs straight with the equilibrium concentration of the adsorptive. The model constant K_d denotes a partitioning coefficient of the adsorbate while the constant m was added by Maurya and Mittal, [37] to increase degree of freedom of the model hence its fitness in describing the isotherm data. Based on this model the glyphosate adsorption by the studied soil has the trend of soil applied with rice husk ash > control > soil + cow dung > burnt soil. Both cow dung and rice husk ash are porous in nature with surface area of 9.731 m^2 g^{-1} and 21.5000 m^2 g^{-1}, respectively [35]. Their application was therefore suggested to result into proportionate increase in the soil adsorption surfaces and greater entrapment of glyphosate ions.

Temkin isotherm model assumes that heat of adsorption decreases linearly with increasing surface coverage as a result of adsorbate-adsorbents interaction. It also gives account of the variation of binding energy for adsorbate retention by the adsorbents [38]. The model had two constant A and b which are related to binding energy and heat of adsorption, respectively. According to the model constant A, burnt soil had higher binding energy (987.950 mg g^{-1}) for glyphosate adsorption and it followed by soils amended with rice husk ash (7.664 mg g^{-1}) and cow dung (6.000 mg g^{-1}) while control soil had the least (4.389 mg g^{-1}). The values of b constant are all positive which indicated that, glyphosate adsorption reaction by all adsorbents was endothermic and the range of the values obtained (b = 1.275 – 26.585 j $mole^{-1}$) was an indication of a physical adsorption process [39]. Therefore, based on the present result the mechanisms of glyphosate adsorption by these adsorbents can be suggested to either be an ion entrapment within the inter facial region of the adsorbent, complexation reaction between cations and, phosphonic and carboxylic moieties of glyphosate or through a ligand exchange between these moieties and surface functional groups of the adsorbents.

3.4 Desorption and predicted mobility of glyphosate

Table 6 shows the glyphosate desorption efficiency from the adsorbents. Of the concentration range applied, the percent glyphosate desorption was generally high at low concentration. The desorption decreases with increasing initial concentration except from control and burnt soils where at 300 mg L^{-1} 42.624% and 12.333% of the respective mass was recovered. Amendment with cow dung or rice husk ash was shown to decrease desorption of glyphosate from this soil to the extent that no glyphosate was detected at 300 mg L^{-1} from the soil amended with cow dung. This indicated that incorporation of these agricultural waste can augment the adsorption strength and glyphosate affinity for this soil. The result further revealed that, burning this soil increases it glyphosate desorption efficiency, hence, it possible mobility compared to control and soils amended with cow dung or rice husk ask. The present result concur with the earlier studies which reported glyphosate desorption from sandy and sandy loam soils of Malaysia [40,41].

Table 6. Glyphosate desorption efficiency by the control and amended soils.

Initial concentration (mg l^{-1})	Control soil	Burnt soil	Soil + cow dung	Soil + rice husk ash
0	0	0	0	0
25	13.332	38.289	17.271	10.572
50	14.718	21.402	12.504	12.379
100	9.363	11.734	8.999	8.563
150	6.868	8.737	7.307	7.159
200	5.252	5.214	5.941	6.264
250	4.772	4.329	5.572	6.944
300	42.624	12.333	nd	4.828
Mean	13.847	14.577	9.599	8.101

The Freundlich isotherm model fit well for glyphosate desorption isotherm data (Table 7) from soil added with cow dung (r^2 = 0.996) or rice husk ash (r^2 = 0.951) compared to burnt soil (r^2 = 0.683) and control (r^2 = 0.423). The values of K_F for glyphosate desorption ranged from 244. 682 mg g^{-1} for control soil to as low as 0.022 mg g^{-1} for the burnt soil. Meanwhile the values for 1/n constant for the absorbents were between 0.529 and 4.632. The values of constant K_F and 1/n were greater in glyphosate desorption compared to those for adsorption. Similar trend was reported from the literature [19,36,42,43]. The present study revealed that glyphosate desorption was of two types; reversible and hysteretic. Desorption of glyphosate was strongly hysteretic (HI = 0.248) in control soil but was reversible in the other adsorbents. Thus, glyphosate desorption was highly reversible in the burnt soil (HI = 18.180) then followed by soil added with cow dung (1.037). In contrast, in addition of being reversible technically hysteresis was absent in desorption of glyphosate from the soil added with rice husk ash (HI = 0.732). Mamy and Barriuso [42], reported that no hysteresis can be definite in practice when 0.7 < H < 1.0.

In other to confirm these two types of desorption obtained from the calculated hysteresis index, the glyphosate adsorption and desorption isotherms were compared as shown in Figure 4. Hysteresis tended to be greater in the control soil and this was due to lack of overlap in the adsorption-desorption isotherms. Similarly little hysteretic effect can be seen from glyphosate adsorption-desorption isotherm at high initial concentration in the burnt soil. Meanwhile there was a general overlap, hence, reversible desorption in both burnt soil and soils amended with cow dung or rice husk ash. The greater hysteretic effect observed from the control soil could be due to irreversible binding of

glyphosate ions on the soil native organic matter, mineral content or its Fe and Al oxides. While the little hysteresis shown by glyphosate desorption in the burnt soil at high initial concentration might be attributed to its increased diffusion into the micro pores of this soil [42]. Therefore, since the native organic matter was removed from this soil via burning, this result in predominance soil and/or mineral surface excess which could be of high energy, hence, an increase in intensity for glyphosate especially at higher concentration.

Table 7. Freundlich model constant, correlation coefficients, and hysteresis index for glyphosate desorption from the adsorbents.

Adsorbents	KF (mg g^{-1})	1/n	r^2	HI
Control soil	244.68	0.53	0.423	0.25
Burnt soil	0.02	4.63	0.683	18.18
Soil + cow dung	14.98	1.93	0.996	1.04
Soil + rice husk ash	49.59	1.40	0.951	0.73

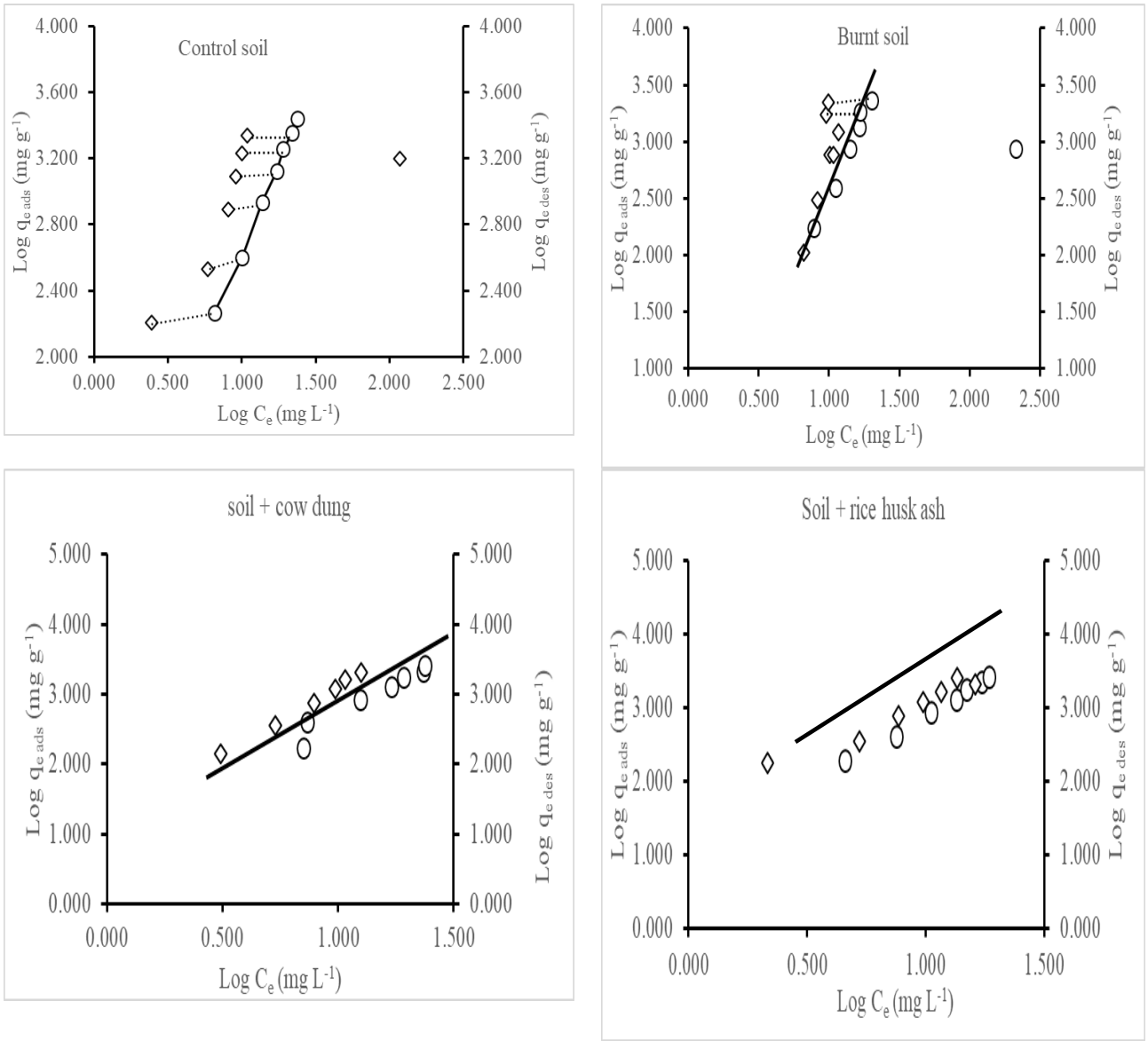


Figure 4. Relationship between glyphosate Freundlich adsorption-desorption isotherms.

Mamy and Barruiso, [42] further opined that it will sometimes be difficult to draw a conclusion that desorption is controlled by sorption mechanisms only with no additional non-sorptive processes. This is due to reliance of desorption parameters on the initial concentration of adsorptive used for the desorption study. These authors used what they termed as general formalism in defining desorption freely of initial concentration of the adsorptive by bringing into used the Equation 8.

$$K_{Fdes} = K_{Fads} C_{e\,ads}^{nads\,(1-H)} \quad (8)$$

The Equation 8 was transformed logarithmically into Equation 9.

$$\text{Log}\,(K_{F\,des}) = \text{Log}\,(K_{F\,ads}) + n_{ads}\,(1\text{-}H)\,\text{Log}\,(C_{e\,ads}) \quad (9)$$

Figure 5 is the result of general formalism in the present study, thus, show the relationship between $K_{F\,des}$ and $C_{e\,ads}$. This relationship from all adsorbents was linear

indicating a control of sorption mechanisms on glyphosate desorption. This result further supports our earlier suggestion that the hysteretic effect observed from control and burnt soils could be due to glyphosate binding on higher energy sorption sites.

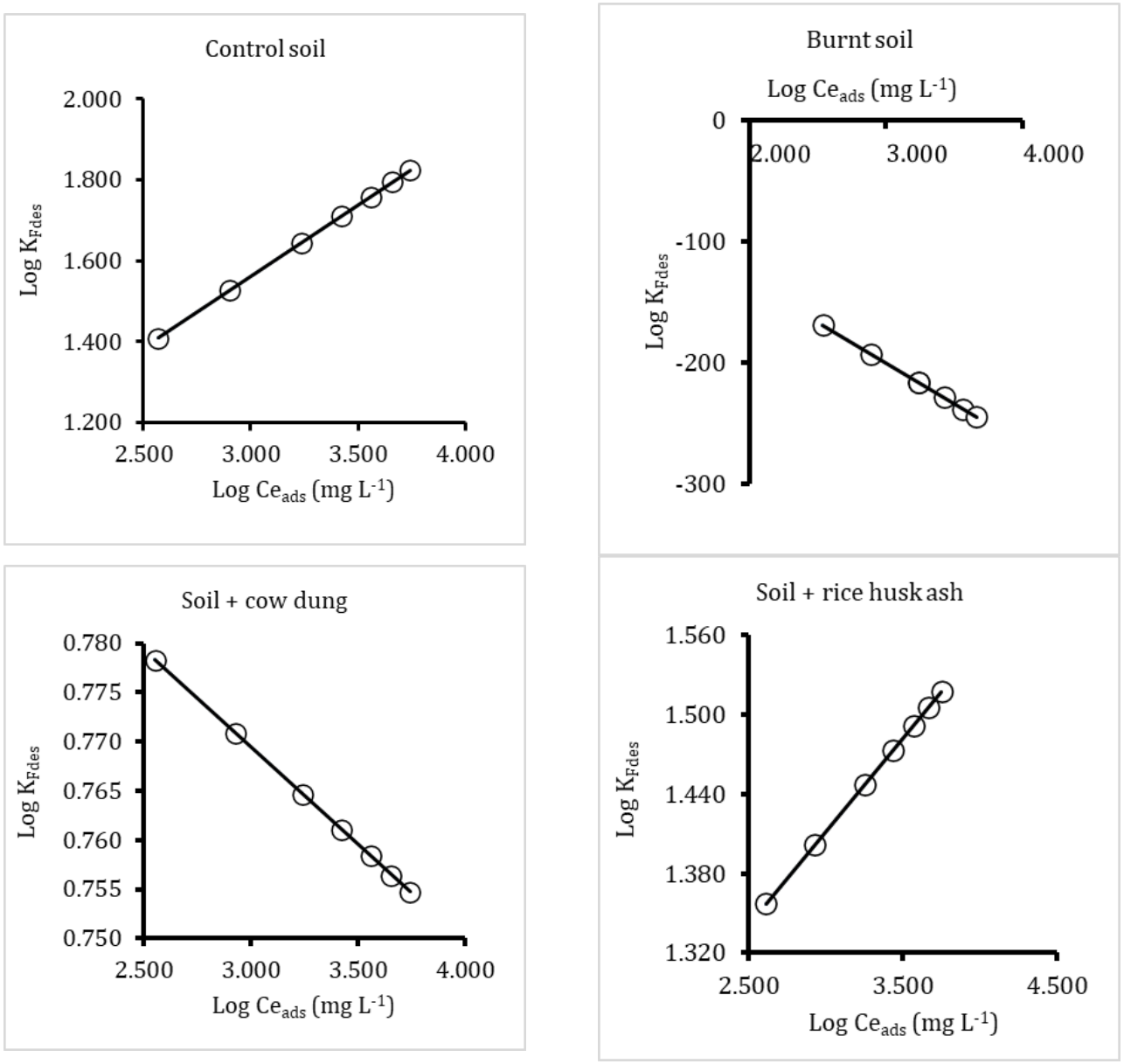


Figure 5. Correlation between the resultant values of K_{Fdes}- Freundlich desorption isotherm constant and Ce_{ads}- concentrations of glyphosate in solution at the commencement of desorption process.

Table 8 depicts the calculated organic carbon normalized adsorption coefficient (K_{oc}) and ground water ubiquity score (GUS) of glyphosate for the adsorbents. Pollutant mobility and persistency are used to assess it leachability in soil and environment [22]. The K_{oc} represent mobility while the persistency is expressed based on the half-life of the pollutant. However, these two indices are both taken into account by GUS to determine the leachability of pollutant in soil and environment. According to this index, Fenoll et al., [44] reports that pollutant with a GUS values of < 1.8 are non-leachers while those with a value range of 1.8 to 2.8 are transitional and those pollutant with values of > 2.8 are potential leachers. Our result showed the calculated GUS for control (2.27) and soil added with cow dung (2.70) or rice husk ash (2.32) to be within the range of 1.8 to 2.8 and thus, glyphosate in these soils was at transitional state. The present result also shows an increase in GUS index owing to addition of cow dung or rice husk ash compared to control. This is suggesting an aging effect on increasing mobility of glyphosate due addition of organic manure. However, due the soil functional biological system these glyphosate ions expected to be released from the adsorption sites will be degraded by microorganisms. The GUS for burnt soil (4.08) was far greater than 2.8 implying the higher possibility of glyphosate mobility and leaching. This is due to killing of soil biota as a result of burning, thus, desorbed glyphosate ions will remain in soil solution which might result in its downward movement and consequent ground water contamination.

Table 8. Organic carbon normalized adsorption coefficient (K_{oc}) and groundwater ubiquity score (GUS) of glyphosate for control and amended soils.

Soil	K_{oc} (μl mg^{-1})	GUS
Control soil	91	2.27
Burnt soil	1	4.08
Soil + cow dung	77	2.70
Soil + rice husk ash	107	2.32

4. Conclusion

Open field burning and manure application are commonly practiced in Malaysia. Benta is one of the Malaysian soil series having sandy texture, high CEC, pH and 2:1 clay mineral, thus, classified as Alfisols according USDA soil taxonomy classification. Burning this soil decreases its uptake of glyphosate but addition of rice husk ash increases its glyphosate removal efficiency. Five different isotherm models were fits into glyphosate batch equilibrium data and based on the MPSD the models were in order of Freundlich > Langmuir > linear > Temkin. Therefore, according to the Freundlich model adsorption of glyphosate was favourable (1/n = 0.255) by burnt soil. Glyphosate desorption was control by adsorption mechanisms and it was inversely proportional (HI = 0.248) to adsorption in control soil, thus, indicating an aging effect of soil organic matter on hysteretic glyphosate desorption. Burning the biota of this soil can be discourage because it increases the mobility (GUS = 4.08) of glyphosate and its potentials in ground water contamination.

Acknowledgement

The authors' appreciation goes to Universiti Putra Malaysia for the provision of funds to complete this study through UPM/GP/IPS/2016-9471900 research grant. Equally, the first author acknowledged the PhD scholarship given to him by tertiary education trust fund (Tetfund) Nigeria through Zamfara State College of Education, Maru, Nigeria.

Author contributions

Jamilu Garba: Investigation, methodology, formal analysis, data curation, writing -original draft.
Samsuri Abd Wahid: Conceptualization, supervision, funding acquisition, editing.
Muhammad Saiful Ahmad Hamdani: Supervision, editing, visualization. validation
Tariq Faruq Sadiq: Formal analysis, visualization, data curation

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Rosmiza, M. Z., Davies, W. P., Rosniza, A. C. R., Mazdi, M., Jabil, M. J., Wan-Toren, W., & CheRosmawati, C. (2014). Farmers' participation in rice straw-utilisation in the MADA region of Kedah, Malaysia. Mediterranean Journal of Social Sciences, 5(23), 229-237. https://doi.org/10.5901/mjss.2014.v5n23p229
2. Rosmiza, M. Z., Davies, W. P., Rosniza, A. C. R., Mazdi, M., & Jabil, M. J. (2014). Farmers' knowledge on potential uses of rice straw: an assessment in MADA and Sekinchan, Malaysia. Geografia, 10(5), 30-43.
3. John, A. (2013). Alternatives to open-field burning on paddy farms. OPTIONS Agricultural and Food Policy Studies Institute, Malaysia, 18, 1-5.
4. de Santana, H., Toni, L. R., Benetoli, L. O. D. B., Zaia, C. T., Rosa Jr, M., & Zaia, D. A. (2006). Effect in glyphosate adsorption on clays and soils heated and characterization by FT–IR spectroscopy. Geoderma, 136(3-4), 738-750. https://doi.org/10.1016/j.geoderma.2006.05.012
5. Kala, D. R., Rosenani, A. B., Fauziah, C. I., Ahmad, S. H., Radziah, O., & Rosazlin, A. (2011). Commercial organic fertilizers and their labeling in Malaysia. Malaysian Journal of Soil Science, 15, 147-157.
6. Kumar, P., Kumar, S., & Joshi, L. (2015). Socioeconomic and environmental implications of agricultural residue burning: A case study of Punjab, India. Springer Nature.
7. Husni, M. H. A., & Samsuri, A. W. (2012). Characterization of local mill rice husk charcoal and its effect on compost properties. Malaysian Journal of Soil Science, 16, 89-102.
8. Tiraieyari, N., Hamzah, A., & Samah, B. A. (2014). Organic farming and sustainable agriculture in Malaysia: organic farmers' challenges towards adoption. Asian Social Science, 10(4), 1-7.
9. Ali, A., & Shaari, N. (2015). Mismanagement of chemical agriculture in Malaysia from legal perspective. Procedia Economics and Finance, 31, 640-650. https://doi.org/10.1016/S2212-5671(15)01152-1
10. Abdul, N. R. (2017). Glyphosate position in Malaysia market. Palma Journal, 16(1), 13-17.
11. Sprankle, P., Meggitt, W. F., & Penner, D. (1975). Rapid inactivation of glyphosate in the soil. Weed Science, 23(3), 224-228. https://doi.org/10.1017/S0043174500052917
12. Coupe, R. H., Kalkhoff, S. J., Capel, P. D., & Gregoire, C. (2012). Fate and transport of glyphosate and aminomethylphosphonic acid in surface waters of agricultural basins. Pest Management Science, 68(1), 16-30. https://doi.org/10.1002/ps.2212

13. Skeff, W., Neumann, C., & Schulz-Bull, D. E. (2015). Glyphosate and AMPA in the estuaries of the Baltic Sea method optimization and field study. Marine Pollution Bulletin, 100(1), 577-585. https://doi.org/10.1016/j.marpolbul.2015.08.015
14. Van Stempvoort, D. R., Roy, J. W., Brown, S. J., & Bickerton, G. (2014). Residues of the herbicide glyphosate in riparian groundwater in urban catchments. Chemosphere, 95, 455-463. https://doi.org/10.1016/j.chemosphere.2013.09.095
15. Franz, J. E., Mao, M. K., & Sikorski, J. A. (1997). Glyphosate: A Unique Global Herbicide. American Chemical Society.
16. Virginia, A., Zamora, M., Barbera, A., Castro-Franco, M., Domenech, M., De Gerónimo, E., & Costa, J. L. (2018). Industrial agriculture and agroecological transition systems: A comparative analysis of productivity results, organic matter and glyphosate in soil. Agricultural Systems, 167, 103-112. https://doi.org/10.1016/j.agsy.2018.09.005
17. Albers, C. N., Banta, G. T., Hansen, P. E., & Jacobsen, O. S. (2009). The influence of organic matter on sorption and fate of glyphosate in soil–Comparing different soils and humic substances. Environmental Pollution, 157(10), 2865-2870. https://doi.org/10.1016/j.envpol.2009.04.004
18. Paramananthan, S. (2000). Malaysian soil taxonomy—a unified Malaysian soil classification system. Soils of Malaysia.
19. Garba, J., Samsuri, A. W., Othman, R., & Ahmad Hamdani, M. S. (2018). Adsorption-desorption and leaching potential of glyphosate and aminomethylphosphonic acid in acidic Malaysian soil amended with cow dung and rice husk ash. Environmental Monitoring and Assessment, 190, 1-15. https://doi.org/10.1007/s10661-018-7034-3
20. Piccolo, A., Celano, G., & Conte, P. (1996). Adsorption of glyphosate by humic substances. Journal of Agricultural and Food Chemistry, 44(8), 2442-2446. https://doi.org/10.1021/jf950620x
21. Garba, J., Othman, R., & Ahmad-hamdani, M. S. (2018). Simplified method for derivatization of extractable glyphosate and aminomethylphosphonic acid and their determination by high performance liquid chromatography. Environmental Research and Technology, 1(2), 19-30.
22. Gustafson, D. I. (1989). Groundwater ubiquity score: a simple method for assessing pesticide leachability. Environmental Toxicology and Chemistry: An International Journal, 8(4), 339-357. https://doi.org/10.1002/etc.5620080411
23. Tessens, E., & Shamshuddin, J. (1982). Characteristics related to charges in Oxisols of Peninsular Malaysia. Pedologie, 32(1), 85-105.
24. Saleh, I. (1997). Pedological Study and Classification of Some Soils Developed on Volcanic Rocks in Jerantut District, Pahang. [Doctoral dissertation, Universiti Putra Malaysia].
25. Ngole-Jeme, V. M. (2019). Fire-induced changes in soil and implications on soil sorption capacity and remediation methods. Applied Sciences, 9(17), 3447. https://doi.org/10.3390/app9173447
26. Reynard-Callanan, J. R., Pope, G. A., Gorring, M. L., & Feng, H. (2010). Effects of high-intensity forest fires on soil clay mineralogy. Physical Geography, 31(5), 407-422. https://doi.org/10.2747/0272-3646.31.5.407
27. Ketterings, Q. M., Bigham, J. M., & Laperche, V. (2000). Changes in soil mineralogy and texture caused by slash-and-burn fires in Sumatra, Indonesia. Soil Science Society of America Journal, 64(3), 1108-1117. https://doi.org/10.2136/sssaj2000.6431108x
28. Araya, S. N., Meding, M., & Berhe, A. A. (2016). Thermal alteration of soil physico-chemical properties: a systematic study to infer response of Sierra Nevada climosequence soils to forest fires. Soil, 2(3), 351-366. https://doi.org/10.5194/soil-2-351-2016
29. Zhelezova, A., Cederlund, H., & Stenström, J. (2017). Effect of biochar amendment and ageing on adsorption and degradation of two herbicides. Water, Air, & Soil Pollution, 228, 1-13. https://doi.org/10.1007/s11270-017-3392-7
30. Garba, J., Abd Wahid, S., Othman, R., Hamdani, M. S. A., & Sadiq, T. F. (2019). Adsorption–desorption of aminomethylphosphonic acid (AMPA) in sandy soil amended with cow dung and rice husk ash. Canadian Journal of Pesticic Management 1(1), 26-36.
31. Sposito, G. (2008). The chemistry of soils. Oxford University Press.
32. Lakshmi, U. R., Srivastava, V. C., Mall, I. D., & Lataye, D. H. (2009). Rice husk ash as an effective adsorbent: Evaluation of adsorptive characteristics for Indigo Carmine dye. Journal of Environmental Management, 90(2), 710-720. https://doi.org/10.1016/j.jenvman.2008.01.002
33. Ayawei, N., Ebelegi, A. N., & Wankasi, D. (2017). Modelling and interpretation of adsorption isotherms. Journal of Chemistry, 2017(1), 3039817. https://doi.org/10.1155/2017/3039817
34. Tévez, H. R., & Afonso, M. D. S. (2015). pH dependence of Glyphosate adsorption on soil horizons. Boletín de la Sociedad Geológica Mexicana, 67(3), 509-516.
35. Garba, J., Samsuri, W. A., Othman, R., & Hamdani, M. S. A. (2019). Evaluation of adsorptive characteristics of cow dung and rice husk ash for removal of aqueous glyphosate and aminomethylphoshonic acid. Scientific Reports, 9(1), 17689. https://doi.org/10.1038/s41598-019-54079-0
36. De Jonge, H., De Jonge, L. W., Jacobsen, O. H., Yamaguchi, T., & Moldrup, P. (2001). Glyphosate sorption in soils of different pH and phosphorus content. Soil Science, 166(4), 230-238.
37. Maurya, N. S., & Mittal, A. K. (2010). Biosorptive color removal: applicability of equilibrium isotherm models. Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management, 14(1), 25-36. https://doi.org/10.1061/(ASCE)1090-025X(2010)14:1(25)
38. Foo, K. Y., & Hameed, B. H. (2010). Insights into the modeling of adsorption isotherm systems. Chemical Engineering Journal, 156(1), 2-10. https://doi.org/10.1016/j.cej.2009.09.013

39. Temkin, M. I. (1940). Kinetics of ammonia synthesis on promoted iron catalysts. Acta Physicochimica URSS, 12, 327-356.
40. Cheah, U. B., Kirkwood, R. C., & Lum, K. Y. (1997). Adsorption, desorption and mobility of four commonly used pesticides in Malaysian agricultural soils. Pesticide Science, 50(1), 53-63. https://doi.org/10.1002/(SICI)1096-9063(199705)50:1<53::AID-PS558>3.0.CO;2-P
41. Ismail, B. S., Zaifah, A. K., Khairiah, J., & Nashriyah, M. (2002). Adsorption-desorption, mobility and degradation of 14 C-glyphosate in two soil series. Jurnal Sains Nuklear Malaysia, 20(1-2), 17-29.
42. Mamy, L., & Barriuso, E. (2007). Desorption and time-dependent sorption of herbicides in soils. European Journal of Soil Science, 58(1), 174-187. https://doi.org/10.1111/j.1365-2389.2006.00822.x
43. Rampoldi, E. A., Hang, S., & Barriuso, E. (2014). Carbon-14-glyphosate behavior in relationship to pedoclimatic conditions and crop sequence. Journal of Environmental Quality, 43(2), 558-567. https://doi.org/10.2134/jeq2013.09.0362
44. Fenoll, J., Vela, N., Navarro, G., Pérez-Lucas, G., & Navarro, S. (2014). Assessment of agro-industrial and composted organic wastes for reducing the potential leaching of triazine herbicide residues through the soil. Science of the Total Environment, 493, 124-132. https://doi.org/10.1016/j.scitotenv.2014.05.098

Turkish Journal of Engineering

https://dergipark.org.tr/en/pub/tuje

e-ISSN 2587-1366

Design of a service for hospital internal transport of urgent pharmaceuticals via drones

Alireza Gholami[*1]

[1] *RL Innovation Inc. MN, 55387, Waconia, 1262 Kinder Drive, United States, ARGholami982@gmail.com*

Cite this study: Gholami, A. (2024). Title of the study. Design of a service for hospital internal transport of urgent pharmaceuticals via drones, 8 (3), 483-497

https://doi.org/10.31127/tuje.1428703

Keywords
Hospital logistics
Drone delivery
Pharmaceutical transportation
Healthcare supply chain
Logistics efficiency

Research Article

Received: 30.01.2024
Revised: 04.03.2024
Accepted: 17.03.2024
Published: 05.07.2024

Abstract

The movement of medical supplies within a hospital heavily depends on people physically carrying these materials. Traditional methods of transporting medical supplies within hospitals often encounter logistical challenges, particularly in densely populated areas like Yalova Merkez in Yalova Province, Turkey. To address these challenges, this study introduces a drone-based delivery system for urgent pharmaceuticals, specifically designed to enhance logistics efficiency and safety within hospital settings. Through a collaborative approach, we developed and validated this service design at CityHospital, a Virtual/Simulated Hospital utilized for our research simulation. Primary user needs were identified through interviews and visual aids, informing the design of the drone service. Feedback from users underscores its potential to significantly improve healthcare logistics. While this system offers notable advantages in efficiency, precautions against risks such as tampering with delivery containers are essential. Proposed strategies include the use of tamper-evident seals and mechatronic locks. Furthermore, this analysis identifies key information for implementing a digital logistics management system, paving the way for future enhancements.

1. Introduction

In modern healthcare systems, the efficient movement of medical supplies within hospitals is not only a logistical concern but a critical determinant of patient care quality [1,2]. Unlike logistics in other industries, managing healthcare logistics entails navigating complex clinical processes, materials handling, and information management, all within the framework of stringent regulations and patient safety protocols [1–3]. Hospital logistics involve a wide range of interconnected activities, each governed by distinct rules, procedures, and pathways, necessitating robust infrastructure and meticulously organized systems to facilitate the seamless flow of diagnostic samples and supplies [1–5]. Traditionally, hospitals have heavily relied on human labor for intra-facility transportation of medical supplies, a method that, while often effective, is not without its inefficiencies and challenges, particularly during times of heightened demand or emergency situations [2]. These challenges include personnel traversing long distances on foot, routing issues resulting in delays, elevator-related bottlenecks, breakdowns in communication, and the heightened risk of hospital-acquired infections stemming from increased human contact with medical supplies during transport [5]. Moreover, the allocation of clinical staff to logistical tasks detracts from the time and attention that could otherwise be dedicated to direct patient care, highlighting the need for more efficient and streamlined logistics solutions [1].

In recent years, researchers have delved into the realm of Key Performance Indicators (KPIs), recognizing their significance as critical metrics utilized by hospitals to optimize internal logistics processes [2]. They found different groups of KPIs related to handling inventory and distributing items in hospitals, which cover areas like quality, time, finances, and productivity. By monitoring and evaluating these KPIs, hospitals can objectively assess the effectiveness of their logistics operations, identify areas for improvement, and implement targeted interventions to enhance supply chain management practices. It is essential to guarantee the safety, affordability, and accessibility of hospital supplies to adequately address the varying requirements of both patients and healthcare providers. Unmanned aircraft systems, often called 'drones,' increasingly demonstrate their potential to revolutionize healthcare [6–9]. Recent research has highlighted the role of drones

in transporting a wide range of essential medical supplies, including blood products, medications, vaccines, and diagnostic samples, particularly in remote and hard-to-reach areas [10–13]. Drones have demonstrated their value in delivering medical supplies to remote communities lacking adequate road infrastructure and in responding rapidly to emergency situations, such as natural disasters or public health crises [7,14,15]. Notable advantages observed in these early experiences include the efficiency of drone operations in time-sensitive scenarios, reduced carbon emissions compared to traditional road transport, and cost savings associated with streamlined logistics operations [14]. While many of these experiments have been conducted in developing countries with more flexible regulatory environments, the integration of drones into urban healthcare logistics presents unique challenges, including technological limitations, public acceptance, and airspace management [16,17].

Since its commencement following the Chicago Convention in 1944, the International Civil Aviation Organization (ICAO) has been instrumental in standardizing international flights for commercial air transport, historically focusing on fixed-wing airplanes [17]. Recently, the ICAO's efforts have extended to include Remotely Piloted Aircraft Systems (RPAS) operating on extended international routes within controlled airspace [18-21]. However, individual ICAO member states are responsible for regulating all other local drone activities, as these flights could pose risks to international commercial aviation, particularly during takeoff and landing. Nevertheless, there are no specific ICAO standards in place for overseeing such domestic drone operations. Small drones, operating below 500 feet AGL, serving various purposes, including short-distance transportation of pharmaceuticals and biological samples. Advancements in aircraft technology have led to the development of gyroplanes capable of road travel and electrically powered multi-rotor aircraft with Vertical Take-Off and Landing (eVTOL) capabilities [19-28]. Urban Air Mobility (UAM) aims to alleviate road congestion using eVTOL aircraft, new aircraft types, and small drones, with a focus on automation and digital technology integration [19,20].

The integration of UAM operations into multimodal transportation and logistic systems underscores the increasing reliance on digital data and information exchange, with the European Aviation Safety Agency (EASA) recognizing the necessity of incorporating UAM into a broader digital ecosystem beyond aviation [19-28]. This paradigm shift not only addresses traditional aviation concerns such as airworthiness and collision prevention but also emphasizes the interconnectedness of aviation with Information Technology (IT) entities in the evolving digital landscape [19-28]. Türkiye has emerged as a leader in establishing comprehensive regulations governing the use of Unmanned Aerial Systems (UAS), Unmanned Traffic Management (UTM), and UAM. Turkiye common rules are 'performance-based,' which, among other things, means widely relying on consensus-based standards developed by industry, keeping the legally binding rules, as much as possible, 'technology-agnostic.' This approach allows the industry to propose new solutions, which, following the established regulatory processes, might be implemented without requiring rules amendment. The common Türkiye rules on the matter are also 'risk-based,' meaning that the approval processes are simpler or non-existent for operations, entailing a lower risk for society. Drones are subject to all applicable Türkiye legislation, e.g., on liability and insurance or privacy and data protection [19-30].

The foundation of the current research is in line with Stakeholder Theory, a conceptual framework pioneered by Freeman [31]. This theoretical perspective is utilized to categorize and prioritize stakeholders integral to the successful deployment of the drone delivery service in healthcare logistics. By adopting the stakeholder theory, a comprehensive approach is included to understand and comprehensively address the needs and interests of diverse groups impacted by the introduction of UAM technologies within the healthcare sector. This theoretical lens guides the exploration of the intricate interplay between various stakeholders, ensuring that their perspectives are duly considered and integrated into the design and implementation processes of UAM-enabled healthcare logistics solutions.

In this research, the critical question of how the utilization of drone technology could be effectively utilized to enhance the internal transport of urgent pharmaceuticals within hospitals, aiming at an enhancement of logistics efficiency while compliance with safety and regulatory standards is maintained, is addressed. The capabilities of UAS are explored, with the design, implementation, and evaluation of a drone-based delivery service, tailored for the healthcare sector, being the focus. The feasibility and effectiveness of UAS in streamlining hospital logistics are assessed, contributing to the broader discourse on innovative logistics solutions in healthcare environments.

Using the preceding sections as a guide, this research aims to create an advanced drone system tailored for urgent pharmaceutical delivery in hospitals. This system will meet the demands of users, safety standards, and regulatory compliance while aligning with the evolving principles of Urban Air Mobility (UAM) outlined by global and Türkiye. Specifically, this study concentrates on swiftly transporting pharmaceuticals from the pharmacy storage area to the Yalova Province within the Virtual/Simulated Hospital (CityHospital), a significant research, healthcare, and Medical Center.

2. Materials and method

A collaborative approach was taken to create a drone service for urgently delivering pharmaceuticals within a hospital, blending input from various stakeholders and users. This method is rooted in the five principles of service design thinking [32], structured resources, and project planning. The authors advocated a process for developing a new service that is (i) centered around users and (ii) co-creative, emphasizing the importance of understanding how consumers or users perceive the service and involving them, along with relevant stakeholders, in the design process. Additionally, they recommended (iii) organizing the service into distinct

moments and sequences and (iv) using visual tools to illustrate it, aiding in a clearer comprehension of the service and soliciting feedback on critical consumer concerns and underlying reasons. Lastly, they advised that the process for designing new services should be (v) holistic, considering various aspects and perspectives of the service's hosting context to gain a comprehensive understanding. Prior to implementing the collaborative approach, foundational information was necessary. Thus, desk research was carried out to gather essential insights about the hospital's context and the stakeholders involved in the internal distribution of drugs using drones. Subsequently, this information was analyzed, leading to (i) a detailed blueprint of the current drug distribution process within the healthcare system (referred to as "As Is") and (ii) the design of a service utilizing Unmanned Aerial Systems (UAS) and the emerging concept of Urban Air Mobility (UAM) (referred to as "Aspirational"). Following this, two interviews were conducted with potential users at CityHospital: one with the pharmacy manager and another with the nursing staff of one of the hospital Operative Units (OU). The primary aim was to validate the service design developed in the prior analysis. The research design of the current study is outlined in Table 1. Subsequent sections elaborate on each phase of the process.

Table 1. Scheme of the research design adopted by the present study.

Phase		Objectives		Activities
Research	a.	Understand the process touchpoints and the users involved	1.	Study of the current hospital drugs distribution process
			2.	Study of UAS technologies used for commercial and medical logistic services
	b.	Understand how UAS technologies are currently used and implemented in similar contexts	3.	Study of UAM concepts
Analysis	a.	Definition of the process of intra-hospital medicine distribution	1.	Definition of the hospital drug distribution process
			2.	Definition of best practices and technologies of the UAS industry applicable to the case study
	b.	Ideate a service including UAM technologies applicable to the case study	3.	Creation of a drone delivery *Aspirational* service process applicable to the hospital case study
Validation	a.	Validate the researchers' understanding of the current process	1.	Pharmacy premises inspection
			2.	Interviews with Pharmacy staff
	b.	Identify unmet needs of prospective users	3.	Interviews with Operative Unit staff
	c.	Validate the designed drone delivery service	4.	Presentation of the analyzed *As Is* process
			5.	Presentation of the *Aspirational* drone delivery service

The methodology employed in this study is based on the principles of service design thinking as delineated by Stickdorn et al. [32], emphasizing the significance of user-centric and collaborative approaches. The effectiveness of this method in enhancing the design of healthcare services is corroborated by previous scholarly literature [33,34]. Through the active involvement of stakeholders and the creation of detailed service blueprints, this approach adheres to established service design best practices [35] while utilizing the capabilities of UAS for medical logistics [36]. This refined strategy addresses all aspects of service design, integrating state-of-the-art UAS technology with the intricate requirements of medical logistics to develop a solution that is both forward-looking and tailored to user needs. This comprehensive integration guarantees the creation of an innovative, user-centered service poised to transform the field of medical logistics.

2.1. Research

The research encompassed three main areas: Firstly, an examination of the current process for distributing drugs within the hospital OU was conducted using internal documentation from CityHospital. This involved studying the workflow phases, touchpoints, and stakeholders involved. Secondly, the focus shifted to investigating unmanned aircraft systems technologies, particularly those used in commercial logistics, specifically in the context of medical services. To understand how UAS technology is employed in healthcare logistics, market research was conducted on various UAS delivery services globally. Finally, this study investigates the analysis of relevant Türkiye regulations, international standards, and industry-specific guidelines to create an outline for the interactions among UAM stakeholders.

2.2. Analysis

The research conducted on hospital documentation extensively examined every phase of the process, the interconnections between users, tools utilized, time estimates for each task, and the essential data framework required to comprehend the service's demands, such as package sizes. This thorough investigation facilitated the development of a service blueprint, which intricately detailed the current internal distribution process of

medications, known as the "As Is" scenario. By crafting this blueprint, it became possible to pinpoint the stakeholders involved in both front and back-office operations. This identification led to creating a stakeholder map, categorizing stakeholders into four tiers based on their impact on the system under review. The classification criteria placed closely involved users in the innermost tier, followed by back-office personnel and those actively supporting the service in the second tier. The third tier encapsulated stakeholders influencing or influenced by the service, while the fourth tier comprised those less directly involved.

Moreover, research into Unmanned Aerial System (UAS) technologies formed the basis for designing a process flow for a UAS delivery service. This comprehensive analysis considered all service touchpoints, components utilized, and the users involved, from order placement to order collection. This facilitated a better understanding of critical factors essential for scaling a UAS delivery service within the scope of the study. A blueprint was developed for an idealized service, and the stakeholder map was updated to encompass new entities involved. Alongside service design, pertinent information necessary for the development and implementation of required technology was included, such as a hierarchy of Key Performance Indicators (KPIs), implementation constraints, and additional prerequisites for deploying the conceptualized digital system. Furthermore, delving into concepts of Urban Air Mobility (UAM) unveiled elements related to pre-service and back-office functions, ensuring a safer integration of routine UAS services. The outcome was the creation of a functional scheme illustrating all main actors within the envisioned UAM ecosystem and their interactions.

In this section, while the manuscript meticulously outlines the development of a service blueprint and the integration of UAS technologies, it could benefit significantly from incorporating a robust mathematical framework to underpin the analysis. Enhancing the methodology with a mathematical modeling approach, such as stochastic modeling or discrete-event simulation, would provide quantitative insights into the system's performance under varying conditions [37]. Stochastic models, for example, could be used to simulate the randomness in delivery times and demand patterns, offering a more detailed understanding of potential service bottlenecks and efficiency under peak loads [38].

Furthermore, the application of optimization algorithms could refine the service design by identifying the most efficient routes and schedules for drone deliveries, taking into account constraints such as flight time limitations and payload capacities. Techniques like Linear Programming and Integer Programming could optimize resource allocation and enhance the service's overall efficiency by minimizing delivery times and operational costs [39]. Incorporating these mathematical techniques would not only strengthen the analysis but also provide a scalable framework for expanding the service in the future. Lastly, the implementation of sensitivity analysis is crucial to understand how variations in key parameters affect the service's performance. This would allow for the identification of critical factors that could impact the reliability and efficiency of the drone delivery system, such as changes in demand or operational disruptions [40]. By integrating these mathematical theories and analytical techniques, the manuscript would offer a more validated and comprehensive framework, enhancing the credibility and applicability of the proposed drone-based pharmaceutical delivery service.

2.3. Validation

The primary aim of this phase of the study encompassed several objectives. Firstly, it sought to confirm the researchers' comprehension of the current process flow. Secondly, it aimed to pinpoint the unfulfilled requirements of potential users. Lastly, it aimed to gather feedback on the planned drone service. The researchers developed tools based on the insights gathered in the analysis phase to accomplish these goals. These tools included a basic service blueprint in the form of a storyboard for the current drug logistics process, one for the proposed UAS drug internal distribution service, and associated stakeholder maps. Additionally, an interview script was formulated to streamline the interview process. The prospective users were divided into two groups: the pharmacy staff and the OU staff, and each interview was divided into two phases. The first phase focused on validating the existing process (As Is), while the second phase concentrated on exploring and validating the desired service (Aspirational). The material used in the initial interviews allowed the researchers to refine the blueprints with a more accurate workflow discussed in the subsequent interviews. The interview script used for the nursing staff closely resembled the one used for the pharmacy staff, but it placed more emphasis on the software used for ordering drugs and probed into the functional and non-functional service requirements. To encourage interaction during the interviews, all the design tools were placed on a table, along with post-its and markers, to create an engaging setting. The interviews began with a presentation of the current internal distribution of pharmaceuticals, starting with the request made by clinical staff and ending with the successful delivery to the OU. At each process stage, questions were posed to identify any missing elements in the workflow. Adjustments and modifications to the process were made collaboratively with the interviewees, and pain points and areas of improvement were defined. After the initial interview phase, the researchers presented a stakeholder map of the existing service. Over the course of the two interviews, the map was reviewed with both the pharmacy and nursing staff, resulting in adjustments to stakeholder relationships and roles based on their input. This led to the discovery of new stakeholders not previously considered, as well as the reevaluation and prioritization of existing ones.

The follow-up interviews centered on unveiling a storyboard for the Aspirational drone delivery service, a prototype conceived during the analysis phase. This storyboard, accompanied by a stakeholder map, was a visual aid to enhance comprehension of the forthcoming service, inviting feedback and insights on each phase and the entire process. Detailed explanations were provided

for each process phase, highlighting variances between the new and existing processes, encompassing fresh stakeholders and interaction points. Engaging with users during these sessions enabled researchers to pinpoint the challenges and constraints within the Aspirational service.

3. Results

The upcoming sections delineate the existing scenario with its constraints and the desired scenario, outlining the envisioned service process. The UAM ecosystem necessary to ensure operational safety and adherence to rules has been defined by examining regulations and global standards. The stakeholder map illustrates the roles of users and the entities involved in this process. The comparison between the 'As Is' and 'Aspirational' scenarios underscores a profound opportunity to enhance the delivery of urgent pharmaceuticals within hospital settings. The observations of this study are in concordance with the findings by Scott and Scott [41], who noted the efficacy of UAS in streamlining medical logistics, particularly highlighting the reduction in delivery times and the potential for improved logistics management. This research contradicts concerns regarding the feasibility of drone deliveries in densely populated areas, as outlined by Yedavalli and Mooberry [42], by presenting a feasible model for the safe and efficient integration of UAS in hospital environments. This model not only addresses safety concerns but also aligns with the progressive trends in UAM, as discussed by Garrow et al. [43]. The demonstration of a viable UAS framework within such critical settings suggests a promising horizon for healthcare logistics, emphasizing the role of innovative technologies in overcoming traditional logistical challenges and enhancing service delivery in healthcare settings. This study contributes to the burgeoning discourse on the potential of UAS in healthcare, suggesting that with appropriate regulatory frameworks and technological advancements, drones can significantly improve the efficiency and responsiveness of pharmaceutical delivery services in hospitals, thereby aligning healthcare logistics more closely with the demands of modern medical care.

3.1. “As Is” Healthcare Logistics Blueprint

The discussions with the pharmacy and OU staff primarily centered on "urgent deliveries," specifically referring to the rapid transport of medications within a relatively short timeframe, ranging from 20 minutes to 2 or 3 hours. These urgent deliveries are necessary when the demand for a medication surpasses the regular delivery schedule or when new inpatients require medications that the OU does not typically stock. This situation commonly arises for specific categories of medicines, including chemotherapeutics, diuretics, special antibiotics, antiblastics, and anti-rejection drugs. Some of these medications are expensive or must be customized for individual patients, making their management critical.

Currently, the responsibility for handling urgent deliveries lies with the hospital pharmacy and the individual OU, who jointly decide on the delivery method on a case-by-case basis. There are three available options for urgent deliveries: (i) using a pneumatic tube system, (ii) employing traditional urgent delivery methods, or (iii) allowing personal pickup by the OU staff.

The pneumatic tube system is the fastest way to transport medicines to an OU, but it has certain limitations. For instance, it may not be suitable for certain types of pharmaceuticals, such as larger packages or fragile compounds. Additionally, it can become overly congested, especially since it is also used for transporting diagnostic samples throughout the hospital. Furthermore, due to its inflexible layout, the recent construction of new buildings to care for COVID-19 patients in CityHospital is not connected to the pneumatic tube system. When the pneumatic tube option is not feasible, the second choice is traditional urgent delivery. In this approach, hospital logistics personnel retrieve medications from the pharmacy and distribute them to all OUs within a hospital building. However, the timing and flow of these urgent deliveries are predetermined, often occurring at a set time of day, like 3 pm. During periods of high demand, there may be delays in delivering the necessary medications, as the delivery staff may not be readily available.

In situations where timely delivery is crucial and other options are not suitable, the third alternative is personal pickup by the OU staff. Nevertheless, this method diverts human resources from clinical duties to retrieve medications from the hospital pharmacy, potentially disrupting patient care. One of the drawbacks highlighted in the interviews is the lack of product traceability, which can result in delivery failures. The pharmacy may receive over 40 urgent delivery requests in a day, with a concentration of these requests between 10 am and 3 pm. After 3 pm, the delivery staffs handle medications ordered before 10 am using the "urgent" option. The delivery method requires a case-by-case negotiation between pharmacy management and the respective OU for medications ordered after that time. Details regarding the weight and dimensions of the cargo are provided in Table 2.

Our findings show that the personal pickup delivery choice proves to be less efficient due to the periodic use of valuable clinical manpower for transporting individual items. Nonetheless, this option becomes crucial during urgent pharmaceutical needs. The scenario referred to as "As Is" described this delivery method and illustrates the process from placing an order to its delivery and filing. Figure 1 illustrates the areas currently under development concerning Innovative Aerial Services.

Table 2. Payload specifications of the pharmaceuticals urgently transported in the hospital.

Minimum Size (cm × cm × cm)	Maximum Size (cm × cm × cm)	Maximum Weight (kg)	Minimum Delivery Frequency (#/h)	Maximum Delivery Frequency (#/h)
4 × 6 × 4	26 × 14 × 14	1.8	3	10

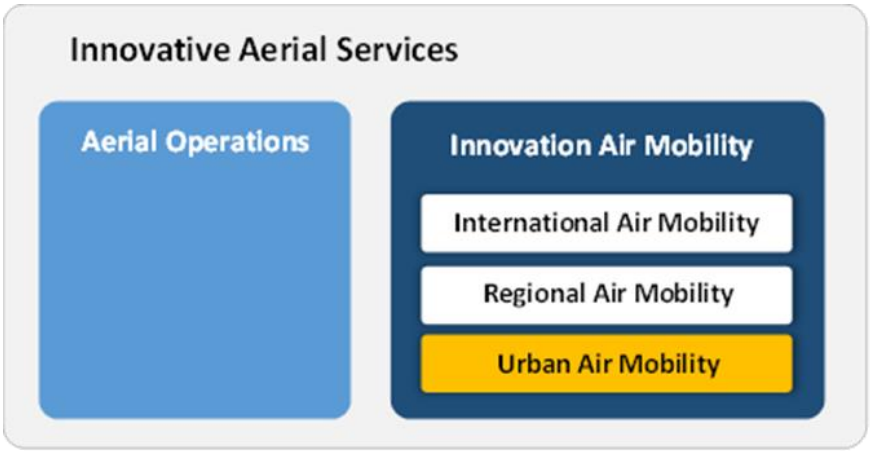


Figure 1. Topics under development regarding Innovative Aerial Services.

3.1.1. Workflow of the *"As Is"* Scenario

The current logistics process begins with the OU coordinator, who has a specific and urgent need for pharmaceuticals outside the regular delivery schedule by the end of the day. To address this requirement, they initiate direct contact with the pharmacy management team for swift negotiations and, if feasible, to place an order. Following successful negotiations and approval from the pharmacy management, the OU coordinator arranges for a personal pickup. Subsequently, the OU coordinator requests one of the nurses or healthcare collaborators to personally visit the pharmacy to obtain the required medications. Concurrently, the pharmacy management team informs the pharmacy coordinator about the placed order and the personal pickup procedure.

While the OU healthcare collaborator is en route to the pharmacy, the OU coordinator proceeds to input the order into the Hospital Management Software. This involves selecting the urgent delivery option and specifying the drugs previously approved by the pharmacy management during their conversation. The order submitted by the OU coordinator is then digitally reviewed and approved by the pharmacy management staff. Once approved, the pharmacy management team forwards the order directly to the pharmacy for processing. Upon arrival at the pharmacy, the pharmacy coordinator generates a delivery note and assigns the task of collecting the drugs to one of their collaborators before the OU collaborator's arrival. The OU collaborator arrives at the pharmacy, collects the prescribed drugs from the pharmacy collaborator, and signs the delivery note. Subsequently, the pharmacy collaborator returns the signed delivery note to the pharmacy coordinator for archiving. Meanwhile, the OU collaborator returns to the OU with the ordered medications.

3.2. 'Aspirational' Healthcare Logistics Blueprint

In the ideal scenario outlined as "Aspirational," the hospital would rely on a drone delivery service. This service could either be managed by an external commercial drone operator, a contracted company responsible for the drones' activities, or it could be an internal asset of the hospital, meaning the hospital itself would operate the drones. Within this scenario, two key roles within the drone operator's organization come into play: the flight manager (FM), responsible for resource allocation (crew and aircraft) for the operations, and the Remote Pilot (RP), in charge of executing the drone flights. The delivery service would utilize highly automated drones equipped with predetermined routes or powered by AI-based software to determine the safest and most efficient flight paths. These flight paths would typically extend from the drone base to a loading point near the pharmacy warehouse. From there, the drones would transport the deliveries to a landing pad located on the relevant floor of the point of care. After completing the delivery, the drones would return to the base. The drone delivery option would be available for OU staff when placing drug orders through the Hospital Management System. Those directly involved with the delivery process would have dedicated software, known as the 'Drone Delivery App,' installed on their workstations or smartphones. This app allows them to interact with the service. Before the operation, the flight manager would use the Drone Delivery App to inform users about drone availability, pickup time, and location. During the operations, users can monitor the delivery's progress using the same software and confirm loading and unloading. This scenario would also necessitate that the regular hospital staff involved in drone operations receive proper training. This training is crucial to ensure that the area is free of people or animals that could be harmed by the drone during loading and unloading. Additionally, it is important to establish a common terminology for effective communication. This is necessary for reporting any issues to the crew or for the crew to instruct hospital staff effectively when needed. The provided flowchart covers the process from order placement to order delivery and filing, specifically focusing on drone operations that have been previously deemed safe, compliant with regulations, and authorized by Türkiye Drone Flight Authorities during the pre-service phase.

3.2.1. Workflow of the Aspirational Scenario

The sequence is initiated by the OU Coordinator in case of an emergency requiring swift delivery of a specific pharmaceutical. Operating from their

workstation, they opt for urgent drug delivery via drone through the Hospital Management Software. The order is then transmitted to the pharmacy management staff, who pass it to the pharmacy. The pharmacy coordinator generates a delivery note and assigns a collaborator to gather the required drugs. Simultaneously, the pharmacy coordinator submits a request for drone delivery using the Drone Delivery App. The flight manager (FM) receives the drone delivery request, checks the availability of both the drone and the Remote Pilot (RP), and organizes the necessary resources for the operation. A notification is sent to confirm drone availability and identification to the pharmacy manager and their collaborator. The collaborator proceeds to the drone base with the drugs and the delivery note, loading them onto the delivery container attached to the drone. Using the Drone Delivery App, they confirm the successful loading of the drone (also here, we can use RFID tags for packs of medicine to confirm them on delivery for both sides). The RP conducts a pre-flight checklist to ensure the drone's operational readiness and safety, verify flight authorization, assess weather conditions, consult available UTM services, ensure the drone parts are undamaged and unobstructed, check battery charge, and confirm clear communication.

Upon completion of the pre-flight checklist, the drone activates its motors, ascends, and flies towards the OU pickup point, following predetermined flight paths under the RP's supervision. The Drone Delivery App notifies the OU collaborators about the drone's approach. Upon landing, the nurse unloads the drone and confirms the unloading via the Drone Delivery App. This confirmation triggers the drone to automatically activate its motors, ascend, and return to the base, notifying involved users that the drugs have been delivered. After landing at the drone base, a notification is sent to the FM about the completed delivery, and the RP conducts a post-flight checklist. In brief, incorporating drone delivery would present an additional choice for OU to obtain medicines. If it demonstrates effectiveness and efficiency, the service could be expanded by incorporating additional drones and resource persons. However, the central link between the drone system and the digital infrastructure would be the FM, potentially requiring around-the-clock staffing in the future.

3.2.2. Hierarchy of KPIs for the urgent intra-hospital transport of medicines

Based on previous research, a framework of Key Performance Indicators (KPIs) for urgent pharmaceutical deliveries within a hospital was formulated [2]. This study further segmented the aspect of 'responsiveness' (timely delivery) into 'preparation time' and 'transport time' to cater to specific scenarios under investigation. 'Responsiveness' here signifies the total duration from the initiation (i.e., order submission intent) to the availability of the requested products for the requesting Operating Unit (OU). The 'preparation time' measures the duration from initiation to cargo loading, while 'transport time' tracks the period taken to move products from the pharmacy to the OU. Moreover, this study introduced 'resolution time' to gauge the interval required to respond to a subsequent request of a similar nature using the same resources after fulfilling the prior order. Although this metric doesn't impact a single delivery, it evaluates the continuity of the ongoing service. The intention behind creating this hierarchy was to steer the development and appropriate scaling of such a service, especially for medium to large-sized hospitals. For instance, this entails determining the right number and types of drones, as well as the necessary charging and delivery container requisites. The primary focus in structuring these KPIs was the prospective advantage for the primary users, while the financial constraints, outlined as 'distribution cost' KPI, were addressed separately. Figure 2 illustrates the KPI hierarchy, with those at higher levels expected to significantly impact users and, consequently, patient care. The KPIs with greater influence from the perspective of Operating Unit staff are detailed on the left side, while those more impactful for pharmacy users are depicted on the right.

The established hierarchy is specific to the use case examined in this study. The planned service is meant to become an integral part of the pharmacy's daily operations, so the key performance indicators (KPIs) focused on the efficiency and accuracy of logistics activities take top priority. Within this study, the primary benefit identified is the reduction in non-clinical tasks performed by clinical staff. Therefore, the measurement of workload distribution, particularly the time clinical staff spend on delivery activities, is of utmost importance, as the implementation of the drone service would significantly impact them. The time taken for preparation significantly affects the workload of pharmacy staff and could potentially influence other Operational Units (OUs) placing orders. KPIs lower in the hierarchy would have a minor to moderate effect; the use case primarily involves medicines needed beyond regular schedules, possibly required by OUs within a few hours. Consequently, responsiveness and delivery frequency were not given high priority. Table 3 displays the time-related KPIs outlined in [2] and introduced in this study, estimated for the Aspirational scenario, and compared with those of the current scenario (As Is).

3.2.3. Constraints for the drone service implementation

The key performance indicator (KPI) of 'distribution cost' was specifically evaluated to address financial constraints regarding the implementation of the proposed service. As the aim wasn't profit generation but rather enhancing the hospital's efficiency and healthcare outcomes—factors challenging to quantify economically—outsourcing this service might not be the most practical choice. The benefits are mainly expressed in terms of innovation, time saved in transport, and improved patient care. Given that the current scenario involves simple drone flights within the hospital's limited airspace, managing this service internally appears feasible for a hospital. This in-house approach would involve initial costs such as certifying the organization as a drone operator, employing one or two staff to manage operations, purchasing the drone fleet, appropriate delivery containers, necessary physical infrastructure

(like landing pads, automated drone recharge stations), and an IT platform (comprising fleet management software and a Drone Delivery App). Operational costs encompass employee training and expenses linked to regulatory compliance. Marginal costs include electricity and internet services, which can be minimized by ensuring proper wireless networking coverage along the flight corridors.

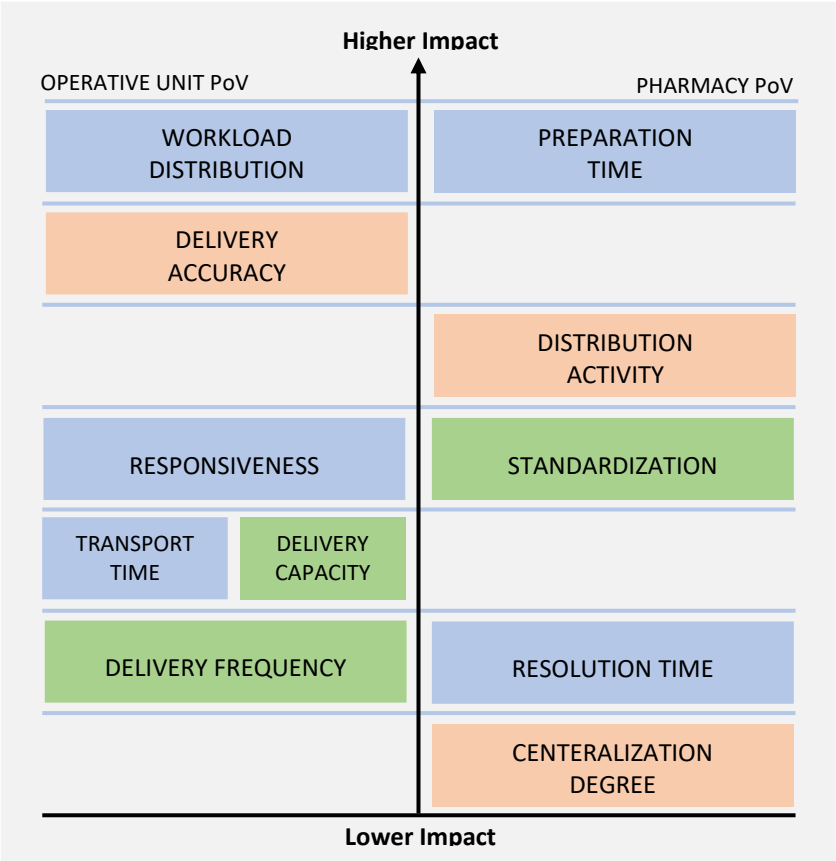


Figure 2. Hierarchy of KPIs for the urgent intra-hospital transport of medicines (time category in green, quality category in blue, productivity category in red).

Table 3. Estimated KPIs in the time category for the investigated scenarios.

Scenario	Responsiveness (min)	Preparation Time (min)	Transport Time (min)	Resolution Time (min)	Workload Distribution (min)
As Is	28–41	15–26	10–17	1–4	18–33
Aspirational	9–20	8–15	3–9	1–2	4–9

For reliable service, it is essential to maintain a continuously charged drone fleet to swiftly respond to new requests. The constraints to consider here involve the frequency of delivery requests, average distance, and cargo weight. These factors help determine the most suitable technical specifications for the drone fleet, including the number of drones, their autonomy, delivery speed, and battery recharge rate. Strategies can range from having charged batteries available when needed to more advanced options like automated drone recharge stations. The researchers used a custom-made drone for their example due to its technical data. They considered a 30-minute drone autonomy (with a maximum loadable weight of 2.7 kg, including the delivery container) and a flight speed of 3 m/s for safety. Estimating an average round trip distance of about 1000 meters at CityHospital, a delivery would require less than 3 minutes of drone activity. The time spent on loading and unloading is not factored in, as the drone would be turned off. A drone could safely conduct at least ten deliveries with a pair of batteries. The drone model's battery recharger can recharge two pairs of batteries in almost 60 minutes, ample time for the activities of a single drone.

The solution's design must also address risks associated with the potential loss of regulated or expensive pharmaceuticals. This study evaluated these risks based on their likelihood and impact levels, identifying mitigations in the service to address them (Table 4). They categorized likelihood as low for rare events (below 0.1%) and medium for occurrences between 0.1% and 1%. The impact was deemed low for minor delivery delays, medium for longer waits (30 minutes to hours), and high for medicine loss or potential damage. The Drone Delivery App could aid risk management by integrating an 'error reporting' feature for all involved users to prompt relevant interventions.

3.2.4. Requirements for the Rollout of the Drone Delivery App

The prioritization of key performance indicators (KPIs) in the current scenario (Figure 2) underscores the

primary emphasis on reducing the time involved in delivery tasks for both the pharmacy and operational unit (OU) staff. Therefore, digital tools enabling precise and timely information exchange play a pivotal role in the ideal scenario. The 'Drone Delivery App' performs various functions in the proposed service, aligning with the logistics plan (Figure 2).

To create an application like the one suggested in this study, the authors recommend utilizing the co-creation approach to gather and confirm requirements. Subsequent development stages must consider the initially identified requirements and functions (Table 5).

Table 4. Identification, evaluation, and mitigation of the risks of loss of pharmaceuticals.

Risk	Likelihood	Impact	Mitigation
Loading failure	Low	Low	Loading confirmation via Drone Delivery App
Unloading failure	Low	Low	Unloading confirmation via Drone Delivery App
Mistaken medicine	Low/Medium	Medium	Delivery note check
Mistaken delivery destination	Low	Medium	Geo-localization feature
Stolen medicine	Low	High	Tamper-evident seal or mechatronic lock
Vehicle crash	Low	High	Crash-proof delivery container; drone operator's emergency response plan

Table 5. Milestones outlined for the development of the Drone Delivery App.

Step	Activities	Tools
Requirements' collection	• Identifying the objectives of the *Aspirational* service • Identifying the users' needs and pain points • Writing user stories regarding the elicited system functionalities	Users' interviews, user journey maps, storyboard, users' stories
Requirements validation	• Identifying the objectives of the *Aspirational* service • Creation of app's mockups • Validation of the proposed solution functions with prospective users • Identifying the objectives of the *Aspirational* service	Users' interviews, user journey maps, low-definition prototypes use case description
Predevelopment planning	• Identifying the objectives of the *Aspirational* service • Selecting where the application will be hosted, on cloud or on-premises servers, depending on legal aspects and data storage security • Selecting the software building language • Data modeling • Starting to set up test cases based on the requirements and defined use cases • Identifying the objectives of the *Aspirational* service	Specification language (software testing)
Quality assurance testing	• Identifying the objectives of the *Aspirational* service • Definition of non-functional metrics • Tests and threshold assessment definition • Validation of the defined metrics • Identifying the objectives of the *Aspirational* service	To be defined depending on the specific metrics
User acceptance testing	• Identifying the objectives of the *Aspirational* service • Definition of usability tests protocol • Recruitment of the target users • Usability test execution • Identifying the objectives of the *Aspirational* service	Usability surveys

3.3. The UAM Ecosystem

The service's design relies on a digital infrastructure that ensures safety and compliance and guarantees the efficiency of the service. Rapid exchange of information among users before, during, and after delivery is crucial. Many of these IT participants are not directly engaged in piloting the drone. The functional diagram in Figure 4 illustrates how information is exchanged among all identified participants and the elements facilitating this communication, known as a UAM ecosystem. This digital ecosystem includes participants from the primary end-user organization (such as the hospital), the UAS operator, and various actors from both aviation and non-aviation backgrounds. These individuals are responsible for coordinating the various activities within the boundaries of their respective establishments. This role represents another component within the digital ecosystem, with no predetermined or restricted count of involved parties.

3.4. Stakeholder map

The pharmaceutical delivery service stakeholders were categorized based on their varying levels of interest. Figure 5 displays the stakeholder map, delineating different tiers. The primary tier encompasses key service users: OU staff, pharmacy staff, and RP, directly engaged in the delivery process. The second tier involves staff whose roles would notably change with the introduction of the new solution, along with those collaborating closely with the main users (such as pharmacy management coordinators, HSR delivery operators, FM, and geo zone managers). The third tier of

stakeholders includes hospital personnel, executives, domain enterprises, and other parties regularly interacting with the most affected users, capable of influencing daily service operations and availability.

In the current 'As Is' scenario, the involved stakeholders include doctors, specialized students, patients, the Health and Safety (H&S) manager overseeing safe movement within the hospital premises, and the surveillance coordinator responsible for implementing safety measures. In the 'Aspirational' scenario, additional stakeholders are introduced. This involves roles like the clinical and research area managers within the hospital, capable of approving or denying airspace use above healthcare and research buildings. Additionally, the fourth tier includes stakeholders like suppliers, regulators, and those not directly linked to the use case but distantly connected to service operations (e.g., pharmacy suppliers and delivery staff. All stakeholders and their role descriptions and levels of interest are detailed in Table 6.

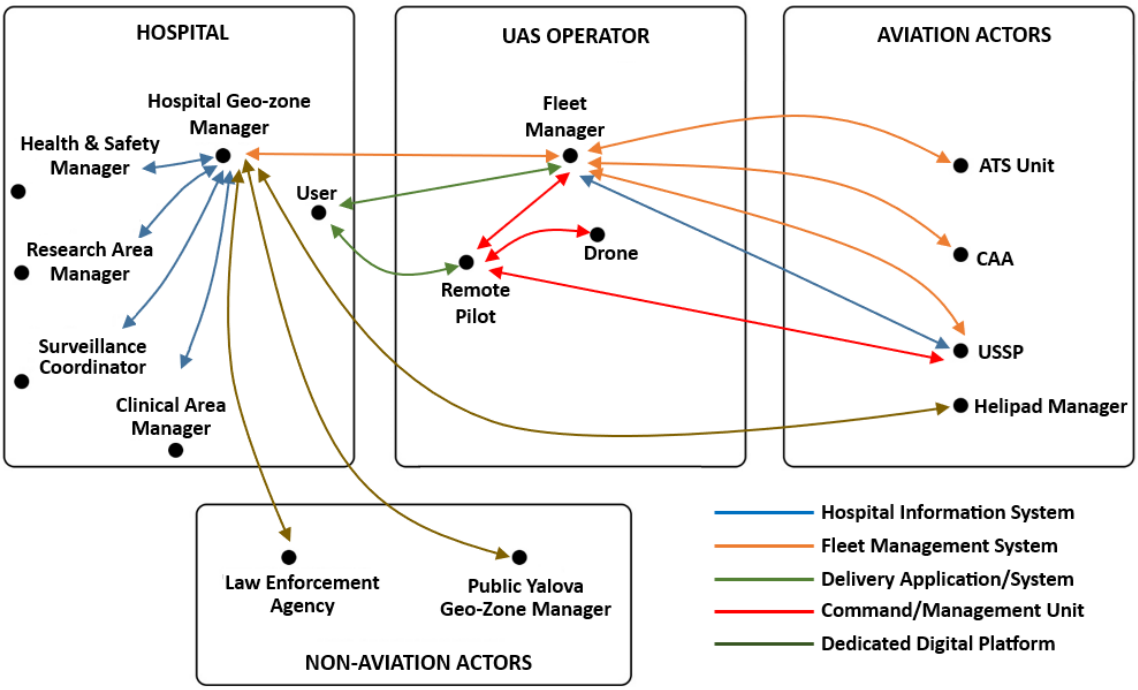


Figure 4. The UAM ecosystem sustains the drone delivery service.

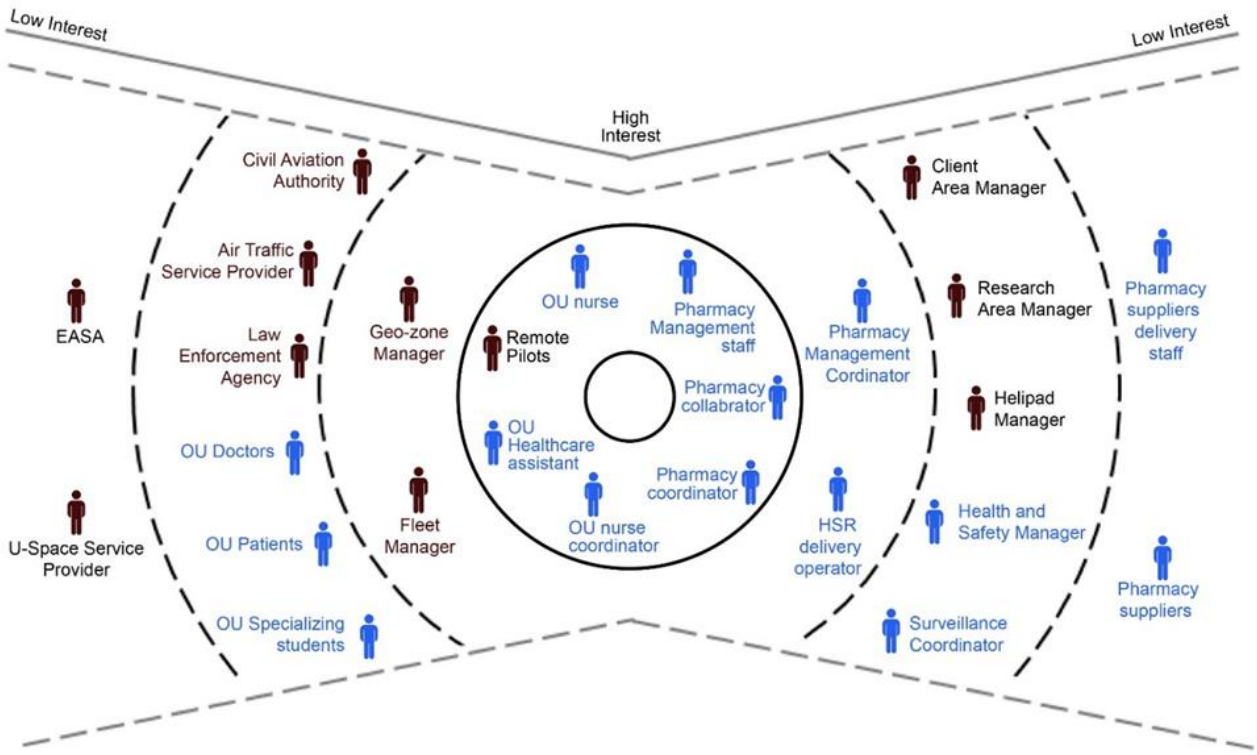


Figure 5. Stakeholder map of the *As Is* (blue) and *Aspirational* (blue and black) scenarios.

Table 6. Summary of the stakeholders' roles and interactions in the investigated scenarios.

Name	Role	Description	Tier
OU Nurse Coordinator	User	Places drug orders, manages the relationships with the Pharmacy Management, manages nurses and healthcare assistants	1
OU Nurse	User	Collects the ordered drugs, unloads the drone	1
OU Healthcare assistant	User	Collects the ordered drugs, unloads the drone	1
Pharmacy Management Staff	Operational support	Manages drugs orders, manages pharmacy relationships	2
Pharmacy Coordinator	User	Manages ordered drugs, manages pharmacy workflow, manages pharmacy personnel, places drone delivery orders	2
Pharmacy collaborator	User	Collects the ordered drugs, loads the drone	2
Remote Pilot	User	Authorizes and supervises the drone delivery	2
Flight Manager	Operational support	Manages the hospital's drone fleet	3
Geo-zone Manager	Regulator	Approves or withdraws permission for airspace use	3
HSR delivery operators	Operational support	Delivers drugs to the operative units in urgent orders with traditional delivery	3
Pharmacy Management Coordinator	Operational support	Manages pharmacy relations	3
Health & Safety Manager	Regulator	Defines the drone service safety rules and limitations in the hospital premises	3
Surveillance Coordinator	Operational support	Implements the drone service safety rules managing the surveillance staff in the hospital	3
Research Area Manager	Regulator	Defines drone service limitations in the airspace above the research department	3
Clinical Area Manager	Regulator	Defines drone service limitations in the airspace above the research department	3
Air Traffic Service Provider	Regulator	Manages the traditional aviation in proximity of the hospital through the relevant ATS unit	3
Civil Aviation Authority	Regulator	Conducts technical regulation and inspection, certification, authorization, coordination and control activities in Italian aviation	3
OU doctors	Other	Define the treatment of patients in the OU	3
OU specializing student	Other	Supports OU doctors	3
OU patients	Other	Trigger the need for pharmaceuticals	3
Law enforcement	Other	Monitors the security of the operations and can enforce the interruption of the drone flights upon the city prefecture's decision	3
Helipad Manager	Other	Is responsible for the area where the Helicopter Emergency Ambulance Service operates	3
Pharmacy supplier	Supplier	Provides drugs for the hospital pharmacy	4
Pharmacy supplier's delivery staff	Supplier	Delivers the ordered drugs to the hospital pharmacy	4
U-Space Service Providers	Supplier	Provide U-Space digital services to the drone operator	4
EASA	Regulator	It carries out certification, regulation and standardization for civil aviation safety in Europe	4

4. Discussion

The Aspirational scenario proposal was formulated to enhance the efficiency of pharmaceutical delivery within hospitals by implementing drone technology. To achieve this goal, the primary end-users were consulted using a collaborative approach. According to the authors, the proposed design would notably enhance all-time key Performance Indicators (KPIs). Specifically, it was anticipated that drones would considerably decrease service response times. Transportation time would be significantly reduced by utilizing drones instead of human walkers, who can be slower and encounter obstacles such as elevator congestion. Furthermore, reducing human interactions between the pharmacy and the OU staff would positively impact preparation time. It is important to note that the maximum efficiency gains wouldn't solely be due to using drones but also to integrating the drone and its operator into a service-focused digital framework. Regarding Quality KPIs, they encompass delivery accuracy, centralization degree, and the disruption of distribution activities. Delivery accuracy measures the successful deliveries as a percentage of the total operations. The centralization degree refers to consolidating all logistic activities under one organizational unit, while the disruption of distribution activities gauges any disturbances caused. In the proposed scenario, it is not expected that there will be significant changes in centralization degree and disruption of distribution activities. This is because the new service isn't meant to replace existing transportation methods but rather to provide an additional, reliable, and swift option for intra-hospital deliveries.

In the Productivity KPI category, three key factors must be considered: delivery capacity, delivery frequency, and standardization [2]. Delivery capacity is a critical aspect of logistics. Based on interviews, it was found that urgent orders typically involve a single item not exceeding 4 liters in volume and with the largest dimension under 26 cm, as indicated in Table 1. This research doesn't delve into delivery containers, assuming that the drone service can handle this capacity and that the weight of ordered items isn't a concern for most urgent requests. The delivery frequency for urgent drug requests in the case study varies from one request every thirty minutes during low demand to approximately eight minutes in high demand (Table 1).

These requests are fulfilled using pneumatic tubes, dedicated hospital staff for intra-hospital delivery, or personal pickups from OUs, depending on the availability of the first two options. The achievable delivery frequency with a drone service depends on the number of drones in the fleet. According to time estimates, one drone could fulfill a delivery request every 10–17 minutes. Thus, it makes sense to consider having at least 2–3 drones to enhance service availability and reliability. Lastly, the degree of standardization is crucial for performance. Interviews revealed that pharmacy management follows specific rules to prioritize different transport modes based on the availability of pneumatic tubes and delivery staff rather than the specific medicine to be transported. However, the digital forms of requests are filled out only after a vocal negotiation between OU staff and pharmacy management, potentially leading to ambiguity and reduced time efficiency compared to a fully digitized system. The current study identified the primary financial limitations of implementing an in-house managed drone service. It was argued that the economic benefits would be challenging to quantify, as the main goal is to enhance hospital efficiency and patient care. Nonetheless, one could argue that introducing a drone service for pharmaceutical delivery might be more cost-effective if the same Unmanned Aerial System (UAS) were utilized for other services during periods of low demand. Potential applications in a hospital include transporting diagnostic samples and medical supplies, monitoring infrastructure integrity, and surveillance. This emphasizes the pivotal role of Facility Management (FM), which may involve multiple drones and Remote Pilots (RPs).

The proposed design offers several additional advantages, including potentially reducing nosocomial infections by minimizing people's movements between clinical and non-clinical areas. Furthermore, it would alleviate the burden on clinical activities as pharmaceuticals would directly reach the point of care. Product traceability is ensured by associating each order with a designated drone identifier and a predefined route for each possible drone. A tamper-evident seal with the order identifier on the container could be applied to enhance traceability and security, a system already used for valuable goods in hospitals. Further security can be achieved by applying a mechatronic lock controlled by users through the Drone Delivery App on the delivery container. It is essential to emphasize that the implementation of drone services should consider public acceptance at all stages.

4.1. Regulatory challenges surrounding automated flight

To summarize, the proposed Aspirational scenario is designed to address the unmet service needs of users and to improve overall efficiency. However, a thorough assessment is essential to implement this service safely and in accordance with regulations. A key safety issue is the high degree of automation suggested for the system. The Joint Authorities for Rulemaking on Unmanned Systems (JARUS) have outlined various potential automation levels for Unmanned Aircraft Systems (UAS) [21, 22, 44].

(i) In a manual Unmanned Aircraft System (UAS), the flight is directly and continuously controlled by a Remote Pilot (RP) who actively manipulates the flight controls.

(ii) There are four advancing levels of automation in UAS. In these, specific operations are automated, yet human oversight remains essential, with the capability for immediate human intervention to modify the aircraft's flight path.

(iii) A fully autonomous UAS characterizes a system that operates independently, managing itself without human intervention.

In this advanced stage of development, the need for human intervention or supervision to change the flight path of the aircraft is no longer essential. However, this does not negate the continued importance of a flight manager (FM) in organizing the operation, initiating commands, and activating emergency protocols as necessary. There are various levels of automation/autonomy, often described using phrases such as 'human in the loop,' 'human on the loop,' or 'human out of the loop.' Autonomous drones are theoretically allowed under the 'specific' category in Turkey. However, the compliance requirements for technical, safety, and operational standards are much stricter, requiring significantly more resources than less autonomous solutions. Currently, international and national bodies focused on integrating these systems into regulated airspace are not considering fully autonomous civilian Unmanned Aircraft Systems (UAS). It is widely agreed that completely autonomous options won't be ready for commercial use in the immediate or medium-term future. Most developers of electric Vertical Take-Off and Landing (eVTOL) vehicles for passenger transport envision a gradual progression:

1. Having a pilot on board;
2. Shifting to a remote pilot, with each pilot controlling only one eVTOL at a time;
3. A remote pilot managing multiple eVTOLs simultaneously;
4. Eventually, replacing the remote pilot with a flight manager responsible for overseeing the operations of an entire eVTOL fleet.

In the scenario described, drone flight operations are automated but overseen by a Remote Pilot (RP), responsible for simultaneously controlling multiple drones. A key safety requirement is the RP's ability to intervene and manually control the drone at any

moment. For operations falling under the 'specific' category, getting approval from the Drone Aviation Authority is necessary. This approval process includes a Specific Operations Risk Assessment (SORA), which assesses risks on the ground, such as the potential for harm to people nearby [21, 22, 44].

To ensure safety, a combination of technological and operational tactics is critical. The exact measures depend on the nature of the operation. For example, strategies might include routing the drone over the tops of buildings (Figure 3), equipping the drone with a parachute for emergency landings, and having a Flight Termination System (FTS) that works independently to prevent common mode failures. In inherently higher-risk situations, these safety measures may need to comply with certain technical standards. The strategy should include procedures that are well-understood by the pilot and tailored to the specific local environment.

4.2. Envisioning the UAM Ecosystem

The growing necessity for Unmanned Traffic Management (UTM) becomes evident when multiple drone operators occupy the same airspace to prevent conflicts among aerial vehicles. While managing a single operator's fleet in a designated area is straightforward, this study explores environments where several operators are permitted to fly drones in hospital airspace. The goal is to develop a scalable solution. This research facilitates assessing the practicality of such a service on a larger scale, such as utilizing the airspace above a large hospital campus for drones controlled by different organizations for various purposes. Establishing a robust digital infrastructure and operational framework is critical for reducing risks. Urban Air Mobility (UAM) entails a systematic sharing of information among various stakeholders to guarantee safety, laying the groundwork for a UAM system that incorporates multiple participants in a 'smart city' setting. An instance of this would be a geo-zone manager implementing temporary airspace restrictions for drone operations within their jurisdiction, coupled with coordination with air ambulance services to avoid interference with emergency helicopters at adjacent helipads.

Fundamentally, it is apparent that multiple parties must work together within the broader ecosystem. For instance, clinical staff could use a software interface to place orders with the pharmacy, while pharmacy staff might use another interface to verify drone availability. Thus, communication between hospital personnel and Fleet Management (FM) is essential. However, it is likely that a single provider might not be able to supply all the varied software and interfaces needed for this system. As a result, the interoperability of technological components is vital for the viability and adaptability of this system to meet various future requirements.

4.3. Limitations

This study has certain limitations that should be considered. It is important to investigate if other medical facilities use similar methods for delivering pharmaceuticals during the final stages. Engaging with representatives from different hospitals to examine the wider relevance of these findings could lead to broader implementation of the study's conclusions. The Key Performance Indicators (KPIs) hierarchy, the identified challenges in service implementation, and the requirements for the Drone Delivery App might be relevant for other institutions with similar needs, such as urgent intra-hospital pharmaceutical delivery, and could possibly extend to other medical supply and diagnostic sample scenarios. Further research involving potential users is necessary to confirm these possibilities.

One aspect that was not addressed in this study is the design of the delivery container. There is currently no universally recognized standard by aviation regulators or standardization bodies for containers used to transport potentially hazardous materials via drones. It is critical to ensure both the product's integrity and bystander safety for an effective design. However, the specifications may differ depending on the type of material being transported. Monitoring acceleration forces, temperature, and humidity within the container could improve the assurance of the product's integrity. As recent studies have highlighted, understanding how various stresses and physical changes during drone flights affect drug stability is crucial. It has been proposed that each pharmaceutical should be evaluated according to pharmacopeia standards to determine its suitability for drone transportation.

5. Conclusion

In conclusion, the adoption of drone technology for the internal transport of urgent pharmaceuticals within hospitals presents a groundbreaking solution to the pressing challenges of efficiency and safety in healthcare logistics. This innovative approach not only promises to streamline the delivery process but also to significantly reduce delivery times, ensuring that critical medications reach patients faster than ever before. By utilizing the speed and agility of drones, hospitals can overcome physical barriers and congestion, enhancing their ability to provide timely care.

The implementation of this drone delivery system also highlights the importance of security measures and regulatory compliance. Through the use of tamper-evident seals and mechatronic locks, the system ensures the integrity and safety of pharmaceuticals during transit. Furthermore, by aligning with Urban Air Mobility principles and engaging with stakeholders, the project sets a precedent for future advancements in hospital logistics and urban transportation. This collaborative approach not only addresses technical and operational challenges but also promotes an environment of innovation and regulatory cooperation.

Looking forward, the successful integration of drones into hospital logistics has the potential to revolutionize not just the healthcare industry but also urban mobility at large. As this technology matures and regulatory frameworks evolve, the possibilities for its application are boundless. This study not only contributes valuable insights into the feasibility and benefits of drone-based pharmaceutical delivery but also opens the door for

further research and development in this exciting field. With continued innovation and collaboration, drone technology can significantly enhance the efficiency, safety, and quality of healthcare services, ultimately leading to better patient outcomes.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Volland, J., Fügener, A., Schoenfelder, J., & Brunner, J. O. (2017). Material logistics in hospitals: A literature review. Omega, 69, 82-101. https://doi.org/10.1016/j.omega.2016.08.004
2. Moons, K., Waeyenbergh, G., & Pintelon, L. (2019). Measuring the logistics performance of internal hospital supply chains–a literature study. Omega, 82, 205-217. https://doi.org/10.1016/j.omega.2018.01.007
3. Vancroonenburg, W., Esprit, E., Smet, P., & Vanden Berghe, G. (2016). Optimizing internal logistic flows in hospitals by dynamic pick-up and delivery models. In Proceedings of the 11th International Conference on the Practice and Theory of Automated Timetabling, Udine, Italy, 371-383.
4. Medical Device Coordination Group. (2019). Guidance on qualification and classification of software in regulation (EU) 2017/745–MDR and regulation (EU) 2017/746–IVDR. European Commission.
5. Zhou, F., Li, J., Lu, M., Ma, L., Pan, Y., Liu, X., ... & Cai, L. (2020). Tracing asymptomatic SARS-CoV-2 carriers among 3674 hospital staff: a cross-sectional survey. EClinicalMedicine, 26, 1-8.
6. Hiebert, B., Nouvet, E., Jeyabalan, V., & Donelle, L. (2020). The application of drones in healthcare and health-related services in north america: A scoping review. Drones, 4(3), 30. https://doi.org/10.3390/drones4030030
7. Ayamga, M., Akaba, S., & Nyaaba, A. A. (2021). Multifaceted applicability of drones: A review. Technological Forecasting and Social Change, 167, 120677. https://doi.org/10.1016/j.techfore.2021.120677
8. Euchi, J. (2021). Do drones have a realistic place in a pandemic fight for delivering medical supplies in healthcare systems problems?. Chinese Journal of Aeronautics, 34(2), 182-190. https://doi.org/10.1016/j.cja.2020.06.006
9. Merkert, R., & Bushell, J. (2020). Managing the drone revolution: A systematic literature review into the current use of airborne drones and future strategic directions for their effective control. Journal of Air Transport Management, 89, 101929. https://doi.org/10.1016/j.jairtraman.2020.101929
10. Turan, V., Avşar, E., Asadi, D., & Aydın, E. A. (2021). Image processing based autonomous landing zone detection for a multi-rotor drone in emergency situations. Turkish Journal of Engineering, 5(4), 193-200. https://doi.org/10.31127/tuje.744954
11. Zeybek, M. (2021). Classification of UAV point clouds by random forest machine learning algorithm. Turkish Journal of Engineering, 5(2), 48-57. https://doi.org/10.31127/tuje.669566
12. Akay, S. S., Ozcan, O., Şanlı, F. B., Bayram, B., & Görüm, T. (2021). Assessing the spatial accuracy of UAV-derived products based on variation of flight altitudes. Turkish Journal of Engineering, 5(1), 35-40. https://doi.org/10.31127/tuje.653631
13. Amukele, T., Ness, P. M., Tobian, A. A., Boyd, J., & Street, J. (2017). Drone transportation of blood products. Transfusion, 57(3), 582-588. https://doi.org/10.1111/trf.13900
14. Nyaaba, A. A., & Ayamga, M. (2021). Intricacies of medical drones in healthcare delivery: Implications for Africa. Technology in Society, 66, 101624. https://doi.org/10.1016/j.techsoc.2021.101624
15. Haidari, L. A., Brown, S. T., Ferguson, M., Bancroft, E., Spiker, M., Wilcox, A., ... & Lee, B. Y. (2016). The economic and operational value of using drones to transport vaccines. Vaccine, 34(34), 4062-4067. https://doi.org/10.1016/j.vaccine.2016.06.022
16. Wang, N. (2021). "As it is Africa, it is ok"? Ethical considerations of development use of drones for delivery in Malawi. IEEE Transactions on Technology and Society, 2(1), 20-30. https://doi.org/10.1109/TTS.2021.3058669
17. Bauranov, A., & Rakas, J. (2021). Designing airspace for urban air mobility: A review of concepts and approaches. Progress in Aerospace Sciences, 125, 100726. https://doi.org/10.1016/j.paerosci.2021.100726
18. International Civil Aviation Organization (1944). Convention on International Civil Aviation (Chicago Convention); International Civil Aviation Organization: Chicago, IL, USA.
19. ISO/DIS 23629-5 (2022). Unmanned Aircraft Systems—UAS Traffic Management (UTM)—Part 5: UTM Functional Structure. International Organization for Standardization: Geneva, Switzerland.
20. ISO/DIS 23629-12 (2021). UAS Traffic Management (UTM)—Part 12: Requirements for UTM Service Providers. International Organization for Standardization: Geneva, Switzerland
21. Jarus FAQ (2021). http://jarus-rpas.org/
22. Joint Authorities for Rulemaking of Unmanned Systems (2019). JARUS Recommendation for Remote Pilot Competency (RPC) for UAS Operations in Category A (Open) and Category B (Specific). JARUS.
23. European Union. (2019). Commission Delegated Regulation (EU) 2019/945 of 12 March 2019 on Unmanned Aircraft Systems and on Third-Country Operators of Unmanned Aircraft Systems; Official Journal C/2019/1821; European Union: Brussels, Belgium
24. European Union (2019). Commission Implementing Regulation (EU) 2019/947 of 24 May 2019 on the Rules and Procedures for the Operation of Unmanned Aircraft; Official Journal C/2019/3824; European Union: Brussels, Belgium.
25. European Union. (2012). Commission Implementing Regulation (EU) 2021/664 of 22 April 2021 on a

regulatory framework for the U-space; Official Journal C/2021/2671; European Union: Brussels, Belgium
26. European Union Aviation Safety Agency (2021). Study on the Societal Acceptance of Urban Air Mobility in Europe; European Union: Brussels, Belgium.
27. European Cockpit Association. Unmanned Aircraft Systems and the Concepts of Automation and Autonomy. (2020). ECA Briefing Paper 2020; ECA: Brussels, Belgium
28. European Union Aviation Safety Agency. (2021). Provisions Applicable to both 'Open' and 'Specific' Category. https://www.easa.europa.eu/the-agency/faqs/drones-uas
29. European Union. Commission (2021). Implementing Regulation (EU) 2021/666 of 22 April 2021 Amending Regulation (EU) No 923/2012 as Regards Requirements for Manned Aviation Operating in U-Space Airspace; Official Journal C/2021/2673; European Union: Brussels, Belgium
30. ISO 21384-3 (2019). Unmanned Aircraft Systems—Part 3: Operational Procedures. International Organization for Standardization: Geneva, Switzerland
31. Freeman, R. E. (2010). Strategic management: A stakeholder approach. Cambridge University Press.
32. Stickdorn, M., Hormess, M. E., Lawrence, A., & Schneider, J. (2018). This is service design doing. O'Reilly Media, Inc.
33. Jones, P. (2013). Design for care: Innovating healthcare experience. Rosenfeld Media.
34. Clack, L. A., & Ellison, R. L. (2019). Innovation in service design thinking. Service design and service thinking in healthcare and hospital management: Theory, concepts, practice, 85-92. https://doi.org/10.1007/978-3-030-00749-2_6
35. Blomkvist, J., Holmlid, S., & Segelström, F. (2010). Service design research: yesterday, today and tomorrow. This Is Service Design Thinking, 308-315.
36. Euchi, J. (2021). Do drones have a realistic place in a pandemic fight for delivering medical supplies in healthcare systems problems?. Chinese Journal of Aeronautics, 34(2), 182-190. https://doi.org/10.1016/j.cja.2020.06.006
37. Law, A. M., Kelton, W. D., & Kelton, W. D. (2007). Simulation modeling and analysis (Vol. 3). New York: Mcgraw-hill.
38. Hillier, F. S., & Lieberman, G. J. (2015). Introduction to operations research. McGraw-Hill.
39. Taha, H. A. (2017). Operations research: an introduction. Pearson Education India.
40. Saltelli, A., Tarantola, S., & Campolongo, F. (2000). Sensitivity analysis as an ingredient of modeling. Statistical science, 377-395.
41. Scott, J., & Scott, C. (2017). Drone delivery models for healthcare. Proceedings of the 50th Hawaii International Conference on System Sciences, 3297-3304.
42. Yedavalli, P., & Mooberry, J. (2019). An assessment of public perception of urban air mobility (UAM). Airbus UTM: Defining Future Skies, 2046738072-1580045281.
43. Garrow, L. A., German, B. J., & Leonard, C. E. (2021). Urban air mobility: A comprehensive review and comparative analysis with autonomous and electric ground transportation for informing future research. Transportation Research Part C: Emerging Technologies, 132, 103377. https://doi.org/10.1016/j.trc.2021.103377
44. Joint Authorities for Rulemaking of Unmanned Systems (2019). JARUS Guidelines on Specific Operations Risk Assessment (SORA). JARUS. http://jarus-rpas.org/sites/jarus-rpas.org/files/jar_doc_06_jarus_sora_v2.0.pdf

Turkish Journal of Engineering

https://dergipark.org.tr/en/pub/tuje

e-ISSN 2587-1366

Encrypted malware detection methodology without decryption using deep learning-based approaches

Abhay Pratap Singh [*1], **Mahendra Singh** [1], **Karamjit Bhatia** [1], **Heman Pathak**[1]

[1] *Gurukula Kangri University, Department of Computer Science, India, rs.abhaypratapsingh@gkv.ac.in, msa@gkv.ac.in, kbhatia@gkv.ac.in, hpathak@gkv.ac.in*

Cite this study: Singh, A. P., Singh, M., Bhatia, K., & Pathak, H. (2024). Encrypted malware detection methodology without decryption using deep learning-based approaches. Turkish Journal of Engineering, 8 (3), 498-509

https://doi.org/10.31127/tuje.1416933

Keywords
Malware
Encrypted traffic
Deep learning
Network security

Research Article

Received: 09.01.2024
Revised: 07.03.2024
Accepted: 08.03.2024
Published: 05.07.2024

Abstract

The encrypted or https traffic on Internet accounts for the safe and secure communication between users and servers. However, cyber attackers are also exploiting https traffic to disguise their malignant activities. Detection of network threats in https traffic is a tiresome task for security experts owing to the convoluted nature of encrypted traffic on the web. Conventional detection techniques decrypt the network content, check it for threats, re-encrypt the network content, and then send it to the server. But this approach jeopardizes the secrecy of data and user. In recent time, deep learning (DL) has emerged as one of the most fruitful AI methods that diminishes the manual resolution of features to enhance classification accuracy. A DL based strategy is suggested for recognition of threat in encrypted communication without using decryption. The three DL algorithms, as used by the proposed approach are, multilayer perceptron (MLP), long short-term memory (LSTM) and 1-D convolutional neural network (1-D CNN), which are experimented on the CTU-13 malware dataset containing flow-based attributes of network traffic. The outcome of the experiment exhibits that MLP based approach performs better in comparison to 1-D CNN and LSTM based ones and other existing approaches. Thus, the secrecy of the data is maintained and the capability of identifying threats in encrypted communication is augmented.

1. Introduction

In the era of digital security, an ever-increasing number of web applications are utilizing security protocols, such as Hypertext Transfer Protocol Secure (HTTPs), Secure Shell (SSH), and Secure Sockets Layer/Transport Layer Security (SSL/TLS), to encipher the content of Internet traffic to protect user's privacy [1]. The most widely utilized protocol for encrypting Internet traffic is the HTTPs. HTTP over SSL/TLS provides a secure and safe communication through a computer network using encryption algorithms [2]. SSL/TLS is the most commonly used security protocols to encipher the content of HTTP. The primary purpose of using encryption across the Internet is to mitigate security threats so that the attacker or third party is unable intercept the information. Alternatively, encryption degraded the ability of network administrators to monitor their infrastructure for malicious traffic identification. The encryption provides a greater advantage to benign users for protecting their privacy, whereas attackers use it to hide their malevolent activities. As the deployment of encryption is on the rise, the diversity of enciphered malware is also evolving simultaneously. Therefore, cyber attackers also started using encryption as a weaponized tool to perform their malicious activity. Since most of the malignant actions are exposed on online platforms [3], cyber attackers are now utilizing other resources such as click jacking, phishing emails, and malvertising campaigns, some of which are exploiting vulnerabilities in genuine web applications or plug-ins to inject a malefic script that switches you to a malefic website [4]. Currently, threat detection techniques like Intrusion Detection Systems (IDSs) or firewalls provide defense against these types of safety risks [5]. The prime objective of using such devices is to distinguish between malignant and benignant traffic by inspecting the running network traffic over the Internet. However, traditional approaches have failed when the network traffic is enciphered. The other

possible solution for detecting malware in HTTPs traffic is to install HTTPs interceptor proxy between the client and the server. This interceptor inspects the HTTPs traffic of the network by installing a special certificate in their systems. Then the enciphered traffic is de-ciphered and checked for malefic content. If the inspection is passed, the network traffic is re-encrypted and sent to the destination server [6], as depicted in Figure 1. But this approach leads to several other problems. The first problem is that it violates the privacy of HTTPs protocol. The second one is that deploying an interceptor proxy is more expensive and computationally slow while dealing with encryption and decryption of the network traffic.

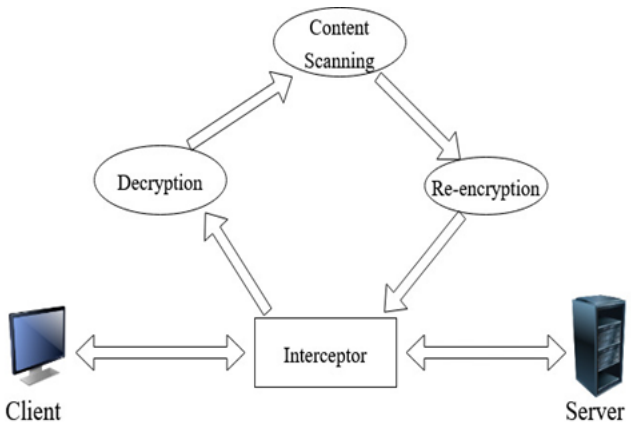


Figure 1. HTTPs Interception.

Therefore, the detection and analysis of enciphered malware remains a continuing challenge for current network security researchers. The application of Machine Learning (ML) based methods to investigate the network traffic measurements have rapidly increased in the recent years. Furthermore, the ML based techniques have shown great permissibility in detecting threats within encrypted traffic. However, the performance of ML based approach is highly reliable on the human-engineered traits and some private traffic information, which can dramatically limit the accuracy and generalizability [7]. The application of DL algorithms in network safety is also growing rapidly. The DL based methodologies facilitate the classification of network traffic by allowing spontaneous extrication and selection of traits by training [8]. The salient point of a DL based modal is that learning ability of it is more than conventional ML models like Support Vector Machine (SVM), Artificial Neural Network (ANN), and k-NN [9]. Hence, these are supposed to learn extremely complex features to achieve greater accuracy with high functionality. DL technology provides a new viewpoint for the applications in which network traffic does not need decryption, but the traits of the network traffic are investigated to handle the malicious traffic [10].

The key contributions of proposed research can be stated as follows:

- Only flow-based features have been extracted from raw encrypted traffic to facilitate traffic classification because the recent adoption of TLS v 1.3 makes classifying malicious traffic challenging.
- The proposed methodology explores different DL models including multi-layer perceptron (MLP), 1-D convolutional neural network (1-D CNN), and long short-term memory (LSTM) neural networks for malware detection in encrypted traffic.
- DL architectures have been tuned and tested while benchmarking with publicly available dataset and performance have been compared with several DL classifiers such as MLP, 1-D CNN, and LSTM.
- A malware detection approach is proposed utilizing DL techniques to achieve higher detection accuracy with a drop in false positive and false negative.

The remaining part of the paper is structured as follows: Section 2 reviews the existing work on malware detection in encrypted traffic and emphasizes the lack of related research. Section 3 explains the phases of the proposed methodology. Architectures of DL techniques are explored in section 4. Section 5 unfolds the experimental details and evaluation indicators. Section 6 contains result analysis. Finally, section 7 concludes the paper.

2. Related work

This section reviews the existing methods being employed for enciphered network traffic classification. We provided a comprehensive review of the ongoing research on malware detection in enciphered traffic with two prime artificial intelligence-based techniques.

2.1 Machine learning based techniques

In [11-12] authors utilized TLS flows, HTTP headers, DNS flow, and SSL/TLS unencrypted metadata information for identifying attacks in enciphered traffic without using decryption. The experimental outcomes displayed a high value for detection accuracy with various ML algorithms. In a similar fashion, Blake Anderson et al. [13] employed many ML algorithms to carry out study and observation of noisy labels and unsteady data. Researchers gathered a number of TLS enciphered flows for one year via a professional malware virtual box and two physically separable large enterprise networks. However, the methodology is dependent largely on human skills to describe the most significant traits. The method exhibited an accuracy of 99% with 0.01 FDR.

An intra-flow data mechanism was proposed in [14], which provided flow-based information to detect and inspect the network's threat. Researchers have also launched a Joy monitoring tool (software) for inspecting the network flows. They gathered an enormous number of malefic flows from threat mesh and normal flows from DMZ of a company network. They studied many features of network traffic, such as series of packet lengths and time, byte organization, and SSL/TLS handshake metadata and then employed these traits for the formation of ML model and classifying enciphered network flow. The detection accuracy of 95.68% was achieved while using flow-based features.

Anish Singh et al. [15] conducted a study on feature analysis of encrypted traffic and divided it into two classes, malefic and benign. The study highlighted feature investigation on the basis of ML models in place of using human skills to explore the appropriate features in enciphered traffic. In addition, authors developed the models using three ML algorithms, namely SVM, XGBoost, and RF, and then carried out feature examination using RFE (recursive feature elimination) method for each one. Among them, the performance of XGBoost was slightly better than RF with the accuracy reaching close to 99%, while SVM yielded a comparatively low accuracy.

A methodology was proposed in [16] based on the intra-flow metadata mechanism. The training and testing data were used from log files created by the Bro IDS tool. The log files used in their work were conn.org, ssl.log, and x509.log. However, in our research work, we used raw PCAP files. They performed with several machine learning algorithms, and experimental results exhibit that XGBoost has the highest detection accuracy of 98.5% among other classifiers. In a similar way, Rui dai et al. [17] proposed a machine learning based model that detected malicious traffic in HTTPs traffic using multi-view features of the network traffic. Authors utilized CTU-13 dataset (log files) and extracted several features such as SSL handshake fields, flow-based statistics and certificates. They investigated and compared four ML algorithms: RF, SVM, DT, and XGBoost. Among these, XGBoost is found to have the highest detection accuracy of 97.71%.

Bryan Scarbrough [18] used only unencrypted portion of the SSL/TLS handshake combined with Open-Source Intelligence (OSINT) data pertaining to IP addresses and domain names. The metadata is then analyzed utilizing three different ML based algorithms: SVM, One-Class SVM (OC-SVM), and an Autoencoder Neural Network. The proposed methodology uses an imbalanced dataset with OC-SVM model achieving high accuracy among all other algorithms for malware detection.

A novel method was proposed in [19] for identifying malignant TLS traffic by using the traits of communication channel. A unique set of attributes for the communication channel was planned with the placement attributes, consistency, and statistical attributes of the TLS network traffic. Then, researchers used a RF algorithm to train and test the detection model. The proposed model achieved an accuracy of 97.44%, which is more conducive for detecting highly disguised malicious traffic.

Luo Ziming et al. [20] introduced a distributed automation mechanism for malware detection in encrypted traffic based on machine learning. They utilized TLS flow of metadata and contextual flow data from network traffic. They applied three ML based algorithms, namely SVM, RF, and XGBoost, for testing the performance of ML model for identification of threats. From experimental results, the performance of RF was found the best among all classifiers. Moreover, this research conducted experiments using the multi-classification model. In addition, Wei Wang et al. [21] implemented an efficient feature extraction methodology based on structural correlation for the identification of threats in TLS malicious enciphered traffic. The extracted features are fed into a RF based ML model. The results generated out of experiment confirmed the accuracy of the model to 99.38%.

Apart from these, a novel unsupervised methodology was presented in [22] for detecting and clustering malicious TLS flows. Researchers built an unsupervised detector that measures the distance to the cluster to determine whether a given flow is malicious or not. They have also evaluated their approach using 972k traces and 35M TLS flows from a commercial sandbox. The proposed unsupervised detector achieves a F1 score of 0.91 and an FDR of 0.032% over the network traffic.

The above ML-based techniques needed the manual design of network traffic features and could not handle the end-to-end enciphered traffic classification. Therefore, the researchers are now utilizing DL networks to design the detection model for enciphered traffic classification to mitigate this issue.

2.2 Deep Learning based techniques

DL methods have been extensively utilized in Image Processing and Natural Language Processing (NLP) but are novel to the field of malicious enciphered traffic identification. Currently, the research on malicious enciphered traffic detection using DL methods has limited studies. This section reviews existing DL-based techniques to deal with threats in enciphered network traffic. Tangda Yu et al. [23] proposed an enciphered malignant network traffic detection method utilizing neural network-based techniques. The authors utilized a self-generated dataset and marked the benign and malicious flow with labels. They used multilayer networks of Autoencoder for feature extraction. Experimental results revealed that the proposed system had high detection accuracy and low loss rate.

Zhihong Zhou et al. [24] suggested a malignant SSL/TLS traffic detection strategy by employing the feature adaptive learning. The proposed detection system consists of three phases; first phase pre-processes the data, during second phase, an unsupervised neural network automatically extracts essential features from the encrypted traffic and optimizes the input data, and in third phase logistic regression classifier is applied which produces classification accuracy of 89.25%. However, the model in unable to predict future malicious traffic pattern and values of FPR and FNR are also not included in the classification result.

Wei Jihong et al. [25] introduced the Hybrid Neural Network Identification Model (HNNIM) for the classification of malicious threats in TLS-based network traffic. The proposed model comprises two layers, the first layer is used for feature extraction, and the second layer uses these features as input for learning. The HNNIM model combined the plain text of the handshake phase of the TLS protocol information and TCP protocol header field information, which integrates a fully connected deep neural network for an effective identification and classification of enciphered traffic. The experimental results showed that the proposed HNNIM

model achieves the average accuracy of 89.28% on the multi-classification task, which was higher than other classifiers. Similar work was carried out by researchers in [26] by proposing a hybrid DL model to classify and detect enciphered network traffic. The CNN and gated recurrent unit (GRU) were used together for rapid feature extraction and learning. This hybrid model provides high accuracy and can be implemented in different networking environments to uniquely classify malicious traffic. Furthermore, the effectiveness of the proposed hybrid DL model was evaluated and tested using three different datasets NSL-KDD, UNSW-NB15, and CICIDS 2017. The proposed work developed a DL based model for identification of malicious activities in enciphered traffic.

A novel method was proposed by Wajdi Bazuhair et al. in [27] where connection features of network traffic are converted into images using the augmentation method. Authors utilized *Perlin Noise* as a carrier function to augment images of network flow features to be used with CNN for training and testing the DL model and successfully classified malignant and benign traffic. The proposed DL model performs better than the traditional machine learning based models by achieving a high detection accuracy of 97%, a low false negative rate of 0.4%, and a relatively higher false positive rate of 5.6%. Another DL based light-weight framework named as Deep-Full Range (DFR) was proposed in [28] for enciphered traffic classification and malicious traffic detection. Three DL algorithms, namely, CNN, LSTM, SAE (stacked auto encoder) are utilized for understanding the raw network traffic. The proposed framework DFR was evaluated on two public datasets and provided a much more robust and precise performance for both enciphered and malicious traffic classification with a minimum storage resource requirement.

3. Proposed methodology

The proposed methodology, which utilizes DL architectures to detect malware in enciphered network traffic, is elaborated in this section. The various phases of the methodology are depicted in Figure 2. Initially, raw network traffic comprising both benign and malicious packet capture files (PCAPs) is collected. Flow based features are then extracted from these PCAPs by the flow generator tool and saved as .CSV file. Next, the preprocessing phase is applied to the .CSV file to clean the data. Subsequently, different DL classifiers are applied to the dataset to train, validate, and test the model to classify the benign and malicious traffic. Finally, several evaluation indicators are used to evaluate the performance of the DL techniques.

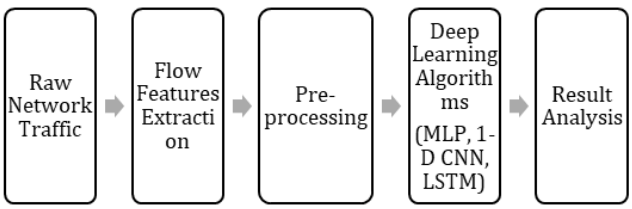


Figure 2. Overview of proposed methodology.

3.1. Dataset description

In our research, we have used the CTU-13 dataset collected from CTU University, Czech Republic available in public domain [29]. CTU-13 dataset contains 13 malicious traffic captures comprising benign and malicious packets. The malicious and benign part of the dataset was collected using a Windows 7 systems running on the virtual machine. The benign traffic utilizes unknown web browsers, and the malware perhaps utilizes its own libraries to communicate with the internet. These network traffic captures are available as PCAP files. We have used these PCAP files for dataset generation. A PCAP file contains the information of network packets and includes the header and payload of every packet. These files are utilized primarily in network analysis and security related tasks.

3.2. Flow features extraction

In this step, various features from benign and malicious PCAPs are extracted through the CICFlowmeter tool [30] and saved into a CSV format file. This tool can pull out more than 80 statistical network flow attributes. It is a network traffic stream creator tool, coded in Java and gives more elasticity when extracting relevant information from PCAP files. It also produces the bidirectional flows (source to destination and destination to source) of captured network flux. A flow in the network is represented by a sequence of packets from a source node to a destination node with similar values for five attributes: Source IP, Destination IP, Source Port, Destination Port, and Protocol. Flow base features can facilitate in-depth network traffic inspection. Table 1 displays a list of flow-based features except Source IP, Destination IP, Source Port, and Protocol because these features are either intrinsically non informative or set up within a simulated environment [31-32].

3.3 Data preprocessing

The original data may have redundancies, errors, and imbalance issues; therefore, it is necessary to remove all such pitfalls from the data before further processing. The features such as Flow Byte/s, Flow Packet/s, and Forward Packet/s that contain infinite and empty values are removed from the dataset using the Pandas library of Python, which relies on a numerical based technique. Dropna() method is utilized to remove the rows which contain infinite and empty values.

3.4 Data normalization

Normalization is an essential step when training a DL model because, without normalization, the model may give false-prone outcomes. For normalization, several techniques are used, literature survey reveals that min max approach, which exhibits good result in terms of scaling and solving outliers. Therefore, a feature scaling technique (min-max scalar) was applied to normalize the values of the dataset for predicting better results and fast

Table 1. The list of flow-based statistical features.

SN	Feature Name	SN	Feature Name	SN	Feature Name
1	Destination Port	27	Bwd IAT Mean	53	Avg Packet Size
2	Flow Duration	28	Bwd IAT Std	54	Avg Fwd Segment Size
3	Total Fwd Packets	29	Bwd IAT Max	55	Avg Bwd Segment Size
4	Total Bwd Packets	30	Bwd IAT Min	56	Fwd Header Length
5	Total Length of Fwd Pkts	31	Fwd PSH Flags	57	Fwd Avg Bytes/bulk
6	Total Length of Bwd Pkts	32	Bwd PSH Flags	58	Fwd Avg Packets /bulk
7	Fwd Packet Length Max	33	Fwd URG Flags	59	Fwd Avg Bulk Rate
8	Fwd Packet Length Min	34	Bwd URG Flags	60	Bwd Avg Bytes/Bulk
9	Fwd Packet Length Mean	35	Fwd Header Length	61	Bwd Avg Packets/Bulk
10	Fwd Packet Length Std	36	Bwd Header Length	62	Bwd Avg Bulk Rate
11	Bwd Packet Length Max	37	Fwd Packets/s	63	Sub flow Fwd Packets
12	Bwd Packet Length Min	38	Bwd Packets/s	64	Sub flow Fwd Bytes
13	Bwd Packet Length Mean	39	Min Packet Length	65	Sub flow Bwd Packets
14	Bwd Packet Length Std	40	Max Packet Length	66	Sub flow Bwd Bytes
15	Flow Bytes/s	41	Packet Length Mean	67	Init-Win-bytes-forward
16	Flow Packets/s	42	Packet Length Std	68	Init-Win-bytes-backward
17	Flow IAT Mean	43	Packet Length Variance	69	act-data-pkt-fwd
18	Flow IAT Std	44	Fin Flag Count	70	min-seg-size-forward
19	Flow IAT Max	45	Syn Flag Count	71	Active Mean
20	Flow IAT Min	46	RST Flag Count	72	Active Std
21	Fwd IAT Total	47	PSH Flag Count	73	Active Max
22	Fwd IAT Mean	48	ACK Flag Count	74	Active Min
23	Fwd IAT Std	49	URG Flag Count	75	Idle Mean
24	Fwd IAT Max	50	CWE Flag Count	76	Idle Std
25	Fwd IAT Min	51	ECE Flag Count	77	Idle Max
26	Bwd IAT Total	52	Down/Up Ratio	78	Idle Min

convergence of the model [33]. Min-Max scaling feature values in the range of [0, 1] using Equation 1. Now, all the features have the same weights and are in one scope.

$$X_{scaled} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

where, $\min(x)$ and $\max(x)$ indicate the minimum and maximum values of feature x, respectively. X_{scaled} is the normalized feature.

4. Deep Learning Classifiers

Our approach aims to utilize the power and efficiency of the DL models to detect malware in HTTPs traffic. A scalable DL model provides an optimal solution with incredible accuracy. The proposed methodology applies three DL classifiers MLP, 1-D CNN, and LSTM for malicious traffic detection. Our dataset consists of time series data and high dimensional input features; where, MLP, 1-D CNN, and LSTM are reported to be more accurate [34]. An overview of these three classifiers along with the summary of used hyper-parameters is outlined hereunder.

4.1 MLP (Multi-layer Perceptron Classifier)

The MLP classifier is primarily used for regression and classification problems [35]. It is a feed-forward neural network that consists of three layers: an input layer, various hidden layers, and an output layer. Every layer has various neurons which are closely connected to the adjacent layers. A neuron uses a weighted sum of its input feature and produces an output that passes through a non-linear activation function. MLP also utilizes back propagation for training the neural network. The input layer contains the feature of a dataset that feeds into the hidden layer, which acts as a computational engine between the input layer and output layer. Finally, the output layer exhibits the result of the given input feature. In our proposed study, the selection of hyper-parameters for the MLP classifier is illustrated in Table 2.

Table 2. Hyper-parameters summary for MLP.

Hyper-parameters	Value
Hidden layer size	(100,100)
Learning rate	0.1
Loss function	Cross-entropy
Optimizer	Stochastic gradient descent (sgd)
Activation function	Logistic
Max iterations	200

4.2 1-D CNN (1-Dimensional convolutional neural network)

The CNN model is primarily utilized in image processing, computer vision, and object detection. This type of architecture consists of an input layer, various hidden layers (convolution layers, pooling, and fully connected), and an output layer. For most image processing applications, 2-D CNN is applied to image data. It is also known as 2 dimensional CNN because the kernel moves the image from left to right and top to bottom, whereas the 1-D CNN kernel slides along one dimension. The only major difference between 1-D CNN and 2-D CNN is the filter and input dimensions of the data [36]. We selected 1D-CNN over 2D-CNN for our study because the structure of dataset format is text or sequential. The architecture of the 1D-CNN classifier is shown in Figure 3.

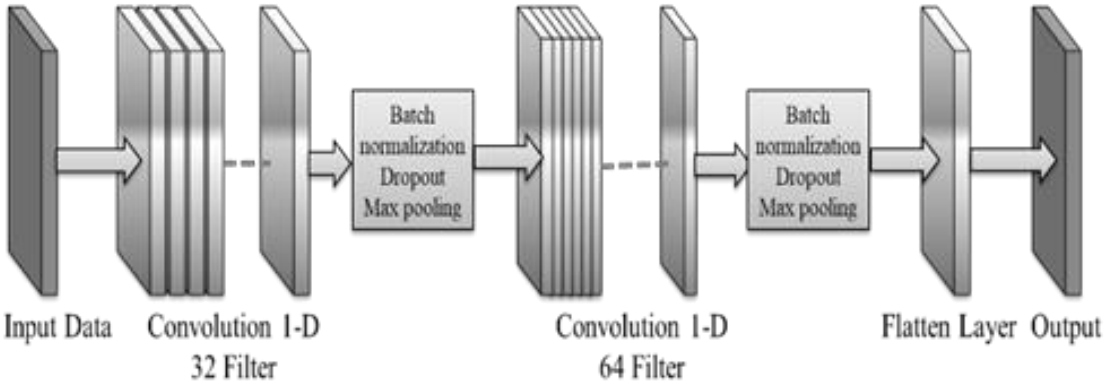


Figure 3. Architecture of 1-D CNN.

The architecture of 1D-CNN consists of two convolution layers, batch normalization, dropout, max pooling, flatten layer, and a dense connected layer with a sigmoid activation function.

The first convolution layer processes the input data with 32 filters, where kernel size is 2 with stride 1 [37]. Each filter moves 1 step after one convolution operation. The results of the convolution layer are forwarded to Rectified Linear Unit (ReLU) activation function. The purpose of the ReLU function is to remove all negative values in the filtered layer and replace them with zeros. The equation for ReLU is defined in Equation 2.

$$f(x) = \max(0, x) \quad (2)$$

Afterward, the results are processed with batch normalization, dropout, and max-pooling layer. Batch normalization is used for improving the speed and performance of neural networks [38]. It makes the training process of the neural networks easy. The key idea is to use the dropout technique to randomly delete units from the neural networks during the training process, keeping the neural network away from being dependent on some specific features leading to better classification results in any circumstances [38]. The main objective of the dropout layer is to prevent over fitting issues. Max-pooling layer is used to reduce the size of convoluted features and computation in the network. The result goes through a second convolution layer, which is similar to the first one. The only difference between these two convolutional layers is that the second convolution layer has 64 filters and different dropout rates. The flatten layer involves in transforming the entire pooled feature map matrix into a single column vector. In the end, data will move to a dense connected layer with 500 neurons, followed by a ReLU activation function and a 50% dropout rate. Finally, the output label is attained by the sigmoid activation function (Equation 3). In the context of the proposed research, the selected hyper-parameters for 1-D CNN are illustrated in Table 3.

Figure 4 shows the model summary, describing each layer and the number of parameters. The total numbers of parameters used in this model are 581,641, where 581,449 are trainable, and 192 are non-trainable.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

Table 3. Hyper-parameters Summary for 1-D CNN.

Hyper-parameters	Value
Number of filters (1D-CNN1)	32
Number of filters (1D- CNN2)	64
Kernel size	2
Max-pooling	Pool size (2,2)
Dropout (1D-CNN1)	0.2
Dropout (1D-CNN2)	0.5
Dense layer	500 = units
Dropout (dense layer)	0.5
Loss function	Binarycross-entropy
Optimizer	Adam
Learning rate	0.001
Batch size	32
Number of epochs	200

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 75, 32)	96
batch_normalization (BatchNo	(None, 75, 32)	128
dropout (Dropout)	(None, 75, 32)	0
max_pooling1d (MaxPooling1D)	(None, 37, 32)	0
conv1d_1 (Conv1D)	(None, 36, 64)	4160
batch_normalization_1 (Batch	(None, 36, 64)	256
dropout_1 (Dropout)	(None, 36, 64)	0
max_pooling1d_1 (MaxPooling1	(None, 18, 64)	0
flatten (Flatten)	(None, 1152)	0
dense (Dense)	(None, 500)	576500
dropout_2 (Dropout)	(None, 500)	0
dense_1 (Dense)	(None, 1)	501

Total params: 581,641
Trainable params: 581,449
Non-trainable params: 192

Figure 4. 1-D CNN Model Summary.

4.3. Long Short Term Memory (LSTM)

Hochreiter et al. [39] presented a model architecture known as the LSTM network. This is a type of neural network which is especially designed to prevent the long term dependency problem that cannot be resolved in a Recurrent Neural Network (RNN). It enables the efficient learning process of both long and short term dependencies by modifying the fundamental processing unit. The LSTM based network also resolves the issues related to gradient vanishing and gradient exploding when the network is too large [40]. Typically, LSTM is a type of network which is designed to deal with time-

related information or sequential data. Figure 5 shows the layered architecture LSTM classifier.

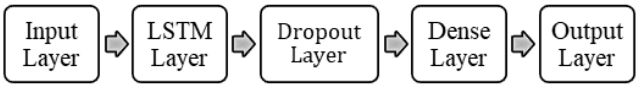


Figure 5. Architecture of LSTM.

LSTM has multi-layered architecture. First, the input data will pass through to the LSTM layer with 170 neurons. In order to prevent overfitting and to achieve better generalization of the model, the dropout layer is used. At last, data will enter to a densely connected layer followed by a sigmoid activation function to predict the output. In the proposed research work, the list of selected hyper-parameters for the LSTM classifier is depicted in Table 4.

Figure 6 all are shows the model summary; the total number of parameters used in this model is 117,131 where 117,131 are trainable.

Table 4. Hyper-parameters Summary for LSTM.

Hyper-parameters	Value
Hidden layer	170 Neurons
Dropout	0.1
Loss function	Binary-cross entropy
Optimizer	Adam
Activation function	Sigmoid
Learning rate	0.01
Max epoch	200
Batch size	32

```
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (32, 170)                 116960
_________________________________________________________________
dropout (Dropout)            (32, 170)                 0
_________________________________________________________________
dense (Dense)                (32, 1)                   171
=================================================================
Total params: 117,131
Trainable params: 117,131
Non-trainable params: 0
```

Figure 6. LSTM Model Summary.

5. Experimental Details

We have implemented three DL architectures with all the layers and activation function built on the flow-based features extracted from raw PCAPs. The dataset is divided into 7:2:1, i.e., 70 % of data is used for training the model, 20 % data is used for validation, and 10 % data is used as test set to evaluate the effectiveness of the proposed methodology. It contains 288,000 packets for training, 72000 packets for validation, and 40000 packets for testing. All the evaluations are performed on the test set data, which the model has never seen during the training phase. For the proposed research study, we acquired the Google Colab Pro platform for efficiently performing all experiments. It enables us faster GPUs, more memory, and faster executions.

5.1 Evaluation Indicators

The performance of the classifiers can be measured using a number of evaluation indicators. The DL classifiers used in the proposed research are evaluated based on indicators of accuracy, precision, recall, F1-score, FPR, and FNR. Achieving the highest accuracy of the classifier is not the only relevant factor in evaluating the classifier's reliability. Therefore, we utilize several evaluation indicators to assess the reliability of the proposed methodology. These evaluation indicators are defined the follows:

TP (True Positive): Encrypted malicious traffic identified as malicious traffic.
FP (False Positive): Encrypted benign traffic identified as malicious traffic.
TN (True Negative): Encrypted benign traffic identified as benign traffic.
FN (False Negative): Encrypted malicious traffic identified as benign traffic.

The values of the indicators are determined as shown in Equation 4-9.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F1 - Score = 2.\frac{precison \times recall}{precison + recall} \quad (7)$$

$$FPR = \frac{FP}{TN + FP} \quad (8)$$

$$FNR = \frac{FN}{TP + FN} \quad (9)$$

6. Results and Analysis

The proposed methodology uses DL classifiers to classify malicious and benign traffic in enciphered network traffic. The following sub-sections elaborate and analyze the results obtained by each classifier used.

6.1 MLP Classifier Results

In order to retrieve the best hyper parameter for MLP based classification technique, we performed a series of experiments with different number of learning rate and the number of iterations (Figure 7).

Table 5. Results of MLP classifier.

Iterations	Training accuracy (%)	Validation accuracy (%)	Testing accuracy (%)
50	97.17	97.20	97.08
100	97.52	97.57	97.60
150	99.06	99.03	99.01
200	99.12	99.13	99.10

This phenomenon is known as hyper-parameter tuning. Table 2 shows the list of hyper parameters used in MLP classification. Initially, the neural network is

trained for a small number of iterations, and it is found that as we increase the number of iterations, the accuracy also improves. Finally, the maximum testing accuracy of 99.10% is achieved at 200 iterations. The accuracy is not changing after 200 iterations because the error rate is significantly decreased. The results obtained from the MLP classifier based on different number of iterations are shown in Table 5.

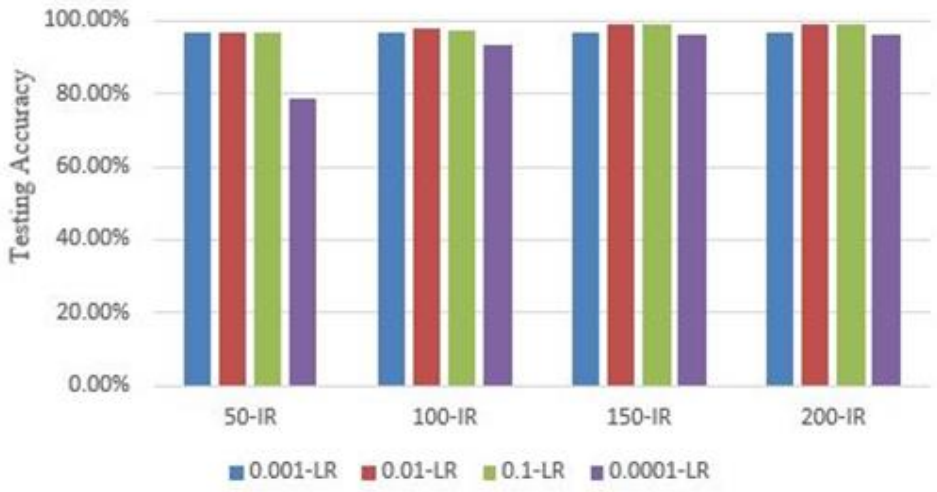


Figure 7. Accuracy changes with different learning rate and iterations.

6.2 1-D CNN Classifier Results

In the 1-D CNN classifier, we also tune the hyper parameter for the classification technique to find a suitable list of hyper parameters for the classification process (Table 3). Learning rate and batch size are important hyper parameters for the DL process. Figure 8 shows the accuracy generated with different numbers of learning rate and batch size for the classification model.

The learning rate determines how fast the parameters are updated, and batch size refers to the number of samples that will be propagated through the neural network before updating the model hyper parameter. The maximum testing accuracy achieved for the 1-D CNN classifier is 98.68% at 200 epochs as shown in Table 6.

The training and validation accuracy, along with training and validation loss for the 1-D CNN model, is shown in Figure 9 and 10 respectively. The purpose of using these curves is to diagnose the overfitting and under fitting problems in the model. It can be observed from Figure 9 and 10 that the gap between training and validation accuracy and that for training and validation loss is relatively small. Thus, it shows no sign of overfitting and underfitting in the model.

Table 6. Results of 1-D CNN classifier.

Epochs	Training accuracy (%)	Validation accuracy (%)	Testing accuracy (%)
50	98.36	98.34	98.21
100	98.66	85.64	98.50
150	98.67	99.14	98.69
200	98.67	98.53	98.68

6.3 LSTM classifier results

Similarly, the hyper parameter is also tuned for this classification approach in order to find a suitable list of hyper parameters for the classifier (Table 4). Figure 11 shows the accuracy regarding the different numbers of learning rate and batch size for the classification technique.

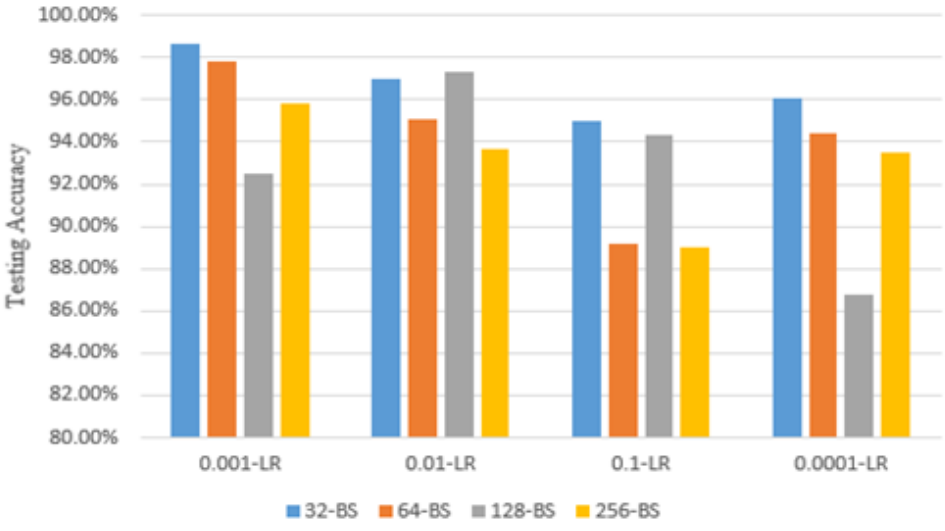


Figure 8. Accuracy changes with different learning rate and batch size.

Figure 9. 1-D CNN training and validation accuracy.

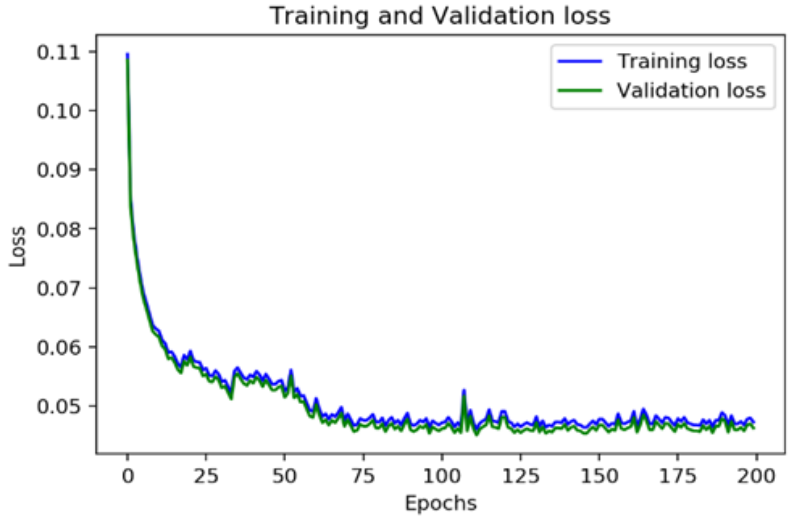


Figure 10. 1-D CNN training and validation loss.

It can be observed from the Table 7 that with each round of epoch, accuracy also increases. The testing accuracy stabilizes at 97.33 % when epoch gets close to 200. The accuracy of the classifier is high, and the technique is relatively reliable, as verified by statistical tests.

Table 7. Results of LSTM classifier.

Epochs	Training accuracy (%)	Validation accuracy (%)	Testing accuracy (%)
50	96.01	94.78	94.92
100	94.94	95.12	95.87
150	97.38	97.43	97.21
200	97.36	97.39	97.33

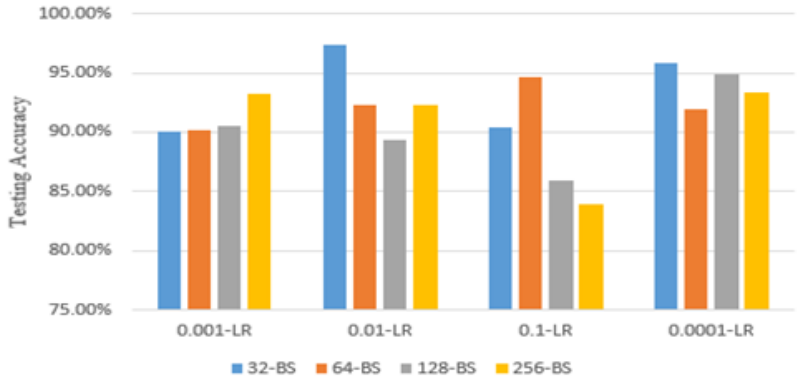


Figure 11. Accuracy changes with different learning rate and batch size.

6.4 Comparative analysis

Results of the proposed classification strategies (MLP, 1-D CNN, and LSTM based), and that of other approaches, are compared and analyzed in this section. The numerical values of evaluation indicators for the proposed strategies and that of other state of arts are shown in Table 8. It is obvious that the accuracy of MLP based

technique is significantly higher than 1-D CNN and LSTM based ones for the classification of encrypted malicious data set containing flow-based features of the traffic. Moreover, it is observed that the proposed DL based methodologies (MLP and 1-D CNN based) perform better than the existing studies [17] and [24] in terms of accuracy, precision, recall, and F1-score. This significant improvement of accuracy, precision, recall, and F1-score can be attributed to the use of selective statistical features based on network flow, in our proposed research study. Another contribution of our proposed research study is to generate the optimum low values of FPR and FNR, which are essential indicators in cyber security in order to assess the performance of a detection strategy. The high values of FPR and FNR may lead to misclassification in any detection model. The existing studies [17] and [24] have not considered FPR and FNR to evaluate their classification strategies.

The experimental results are not only evaluated in terms of evaluation indicators, but also as a confusion matrix. A confusion matrix is a simple table layout used to analyze or to understand the performance of the classifier. It contains the information regarding the actual and predicted classification on the test data. The confusion matrices of all three classifiers are shown in Figure 12, 13, and 14 respectively.

Table 8. Comparison of the proposed techniques with existing works.

Techniques		Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	FPR (%)	FNR (%)
Proposed	MLP	99.10	1.0	98.20	99.09	0.0	0.017
	1-D CNN	98.68	97.40	99.99	98.68	0.026	0.00005
	LSTM	97.33	95.22	99.60	97.36	0.049	0.003
Rui Dai et.al [17]		97.71	98.11	97.49	97.80	-	-
Zhihong et.al [24]		89.25	-	-	-	-	-

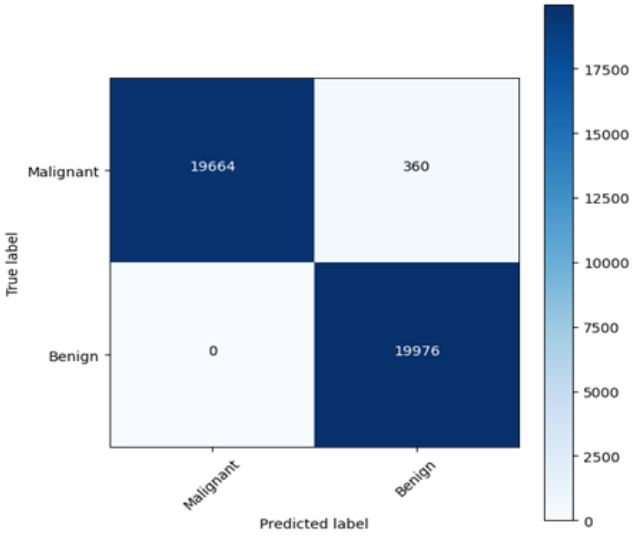


Figure 12. Confusion matrix of MLP model.

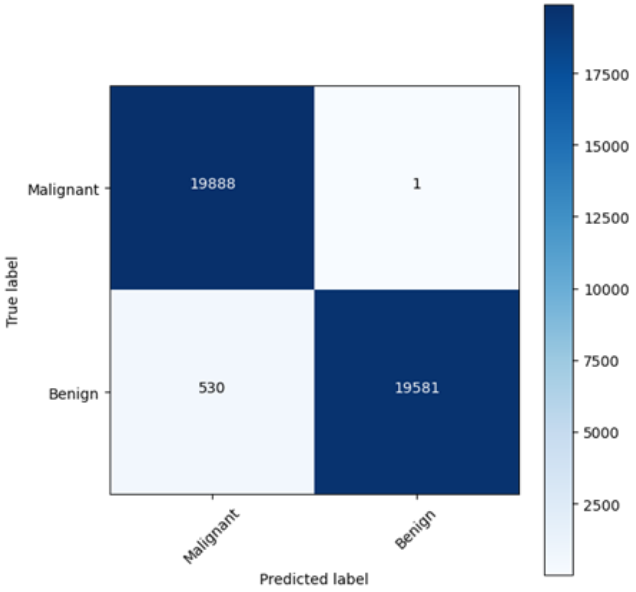


Figure 13. Confusion matrix of 1-D CNN model.

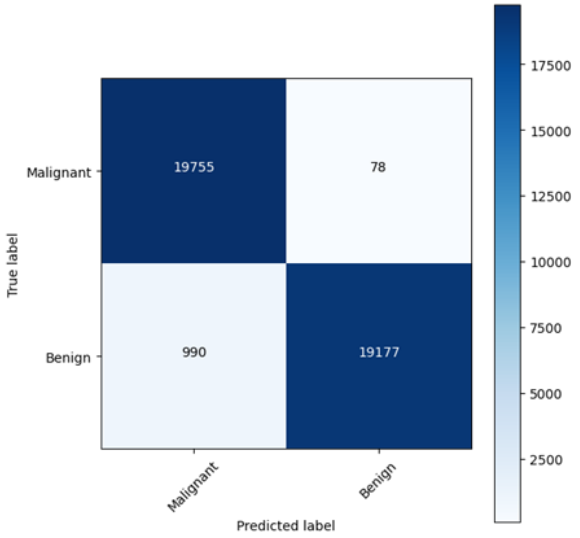


Figure 14. Confusion matrix of LSTM model.

7. Conclusion

As encrypted traffic increases, attackers are also increasingly using encryption to cover up the intention of the attack, which brings considerable value to detection difficulty. Therefore, the importance of encrypted malicious traffic detection has become an increasingly prominent in the prevailing scenario. The researchers are considering the deep learning-based strategies to deal with this critical problem. In this paper, deep learning-based methodology is proposed, which is capable of classifying benign and malicious enciphered traffic with a high accuracy and precision without intercepting the network traffic. The proposed methodology utilizes three DL classifiers MLP, 1-D CNN and LSTM to facilitate classification of network traffic using the CTU-13 dataset. In contrast to other state of arts, the proposed DL-Based approach does not require human intervention for selecting features and private featured details about SSL/TLS metadata and at the same time also maintains the confidentiality of the data. The

results of the experiment reveal that the proposed MLP based approach shows a feasible and scalable solution for malware detection in enciphered network traffic using flow-based features. Furthermore, the proposed approach can be tested on other data sets comprising different types of features. For future investigations, deep learning or hybrid models such as resnet and Bi-LSTM ANN can also be explored for detecting the encrypted malware.

Author contributions

Abhay Pratap Singh: Conceptualization, Proposed Methodology, Writing-original draft preparation.
Mahendra Singh: Methodology, Validation, Writing-original draft preparation.
Karamjit Bhatia: Validation, Writing-Reviewing and Editing.
Heman Pathak: Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Papadogiannaki, E., & Ioannidis, S. (2021). A survey on encrypted network traffic analysis applications, techniques, and countermeasures. ACM Computing Surveys (CSUR), 54(6), 1-35. https://doi.org/10.1145/3457904
2. Singh, A. P., & Singh, M. (2021). A comparative review of malware analysis and detection in HTTPs traffic. International Journal of Computing and Digital Systems, 10(1), 111-123. http://dx.doi.org/10.12785/ijcds/100111
3. Ayas, M. Ş. (2021). A brief review on attack design and detection strategies for networked cyber-physical systems. Turkish Journal of Engineering, 5(1), 1-7. https://doi.org/10.31127/tuje.640282
4. Grier, C., Ballard, L., Caballero, J., Chachra, N., Dietrich, C. J., Levchenko, K., ... & Voelker, G. M. (2012, October). Manufacturing compromise: the emergence of exploit-as-a-service. In Proceedings of the 2012 ACM Conference on Computer and Communications Security, 821-832. https://doi.org/10.1145/2382196.238228
5. Mishra, N., & Pandya, S. (2021). Internet of things applications, security challenges, attacks, intrusion detection, and future visions: A systematic review. IEEE Access, 9, 59353-59377. https://doi.org/10.1109/ACCESS.2021.3073408
6. Zhu, T., Weng, Z., Fu, L., & Ruan, L. (2020). A web shell detection method based on multiview feature fusion. Applied Sciences, 10(18), 6274. https://doi.org/10.3390/app10186274
7. Zhao, Y., Yang, Y., Tian, B., Yang, J., Zhang, T., & Hu, N. (2021). Edge intelligence based identification and classification of encrypted traffic of Internet of Things. IEEE Access, 9, 21895-21903. https://doi.org/10.1109/ACCESS.2021.3056216
8. Wang, P., Ye, F., Chen, X., & Qian, Y. (2018). Datanet: Deep learning based encrypted network traffic classification in sdn home gateway. IEEE Access, 6, 55380-55391. https://doi.org/10.1109/ACCESS.2018.2872430
9. Atli, B. G., Miche, Y., Kalliola, A., Oliver, I., Holtmanns, S., & Lendasse, A. (2018). Anomaly-based intrusion detection using extreme learning machine and aggregation of network traffic statistics in probability space. Cognitive Computation, 10(5), 848-863. https://doi.org/10.1007/s12559-018-9564-y
10. Guo, L., Wu, Q., Liu, S., Duan, M., Li, H., & Sun, J. (2020). Deep learning-based real-time VPN encrypted traffic identification methods. Journal of Real-Time Image Processing, 17(1), 103-114. https://doi.org/10.1007/s11554-019-00930-6
11. Anderson, B., & McGrew, D. (2016, October). Identifying encrypted malware traffic with contextual flow data. In Proceedings of the 2016 ACM workshop on Artificial Intelligence and Security, 35-46. https://doi.org/10.1145/2996758.2996768
12. Anderson, B., Paul, S., & McGrew, D. (2018). Deciphering malware's use of TLS (without decryption). Journal of Computer Virology and Hacking Techniques, 14, 195-211. https://doi.org/10.1007/s11416-017-0306-6
13. Anderson, B., & McGrew, D. (2017, August). Machine learning for encrypted malware traffic classification: accounting for noisy labels and non-stationarity. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1723-1732. https://doi.org/10.1145/3097983.3098163
14. McGrew, D., & Anderson, B. (2016). Enhanced telemetry for encrypted threat analytics. In 2016 IEEE 24th International Conference on Network Protocols (ICNP), 1-6. https://doi.org/10.1109/ICNP.2016.7785325
15. Shekhawat, A. S., Di Troia, F., & Stamp, M. (2019). Feature analysis of encrypted malicious traffic. Expert Systems with Applications, 125, 130-141. https://doi.org/10.1016/j.eswa.2019.01.064
16. Hamad, M., Durad, M. H., & Yousaf, M. (2018). Mitigation of the effect of standard networks attacks in SSL encrypted traffic by encrypted traffic analysis. VFAST Transactions on Mathematics, 6(1), 15-22. https://doi.org/10.21015/vtm.v8i1.578
17. Dai, R., Gao, C., Lang, B., Yang, L., Liu, H., & Chen, S. (2019, November). SSL malicious traffic detection based on multi-view features. In Proceedings of the 2019 9th International Conference on Communication and Network Security, 40-46. https://doi.org/10.1145/3371676.3371697
18. Scarbrough, B. (2021). Malware Detection in Encrypted TLS Traffic Through Machine Learning. Global Information Assurance Certification Paper.
19. Zheng, R., Liu, J., Li, K., Liao, S., & Liu, L. (2020, August). Detecting malicious tls network traffic based on communication channel features. In 2020 IEEE 8th International Conference on Information, Communication and Networks (ICICN), 14-19. https://doi.org/10.1109/ICICN51133.2020.9205087

20. Luo, Z. M., & Xu, S. B. (2020). Scheme for identifying malware traffic with TLS data based on machine learning. Chinese Journal of Network and Information Security, 6(1), 77-83.
21. Wang, W., Sun, C. S., & Ye, J. N. (2021). A method for TLS malicious traffic identification based on machine learning. Advances in Science and Technology, 105, 291-301. https://doi.org/10.4028/www.scientific.net/AST.105.291
22. Gomez, G., Kotzias, P., Dell'Amico, M., Bilge, L., & Caballero, J. (2023). Unsupervised detection and clustering of malicious tls flows. Security and Communication Networks, 2023(1), 3676692. https://doi.org/10.1155/2023/3676692
23. Yu, T., Zou, F., Li, L., & Yi, P. (2019). An encrypted malicious traffic detection system based on neural network. In 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 62-70. https://doi.org/10.1109/CyberC.2019.00020
24. Zhou, Z., Bin, H., Li, J., Yin, Y., Chen, X., Ma, J., & Yao, L. (2022). Malicious encrypted traffic features extraction model based on unsupervised feature adaptive learning. Journal of Computer Virology and Hacking Techniques, 18(4), 453-463. https://doi.org/10.1007/s11416-022-00429-y
25. Jie, F. (2020, September). Research on malicious TLS traffic identification based on hybrid neural network. In 2020 International Conference on Advance in Ambient Computing and Intelligence (ICAACI), 42-46. https://doi.org/10.1109/ICAACI50733.2020.00013
26. Bakhshi, T., & Ghita, B. (2021). Anomaly detection in encrypted internet traffic using hybrid deep learning. Security and Communication Networks, 2021(1), 5363750. https://doi.org/10.1155/2021/5363750
27. Bazuhair, W., & Lee, W. (2020, January). Detecting malign encrypted network traffic using perlin noise and convolutional neural network. In 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), 0200-0206. https://doi.org/10.1109/CCWC47524.2020.9031116
28. Zeng, Y., Gu, H., Wei, W., & Guo, Y. (2019). Deep-Full-Range: A deep learning based network encrypted traffic classification and intrusion detection framework. IEEE Access, 7, 45182-45190. https://doi.org/10.1109/ACCESS.2019.2908225
29. Ctu-13 dataset. https://www.stratosphereips.org/datasets-ctu13
30. CICFlowmeter. https://www.unb.ca/cic/reserach/applications.html
31. Zhang, C., Chen, Y., Meng, Y., Ruan, F., Chen, R., Li, Y., & Yang, Y. (2021). A novel framework design of network intrusion detection based on machine learning techniques. Security and Communication Networks, 2021(1), 6610675. https://doi.org/10.1155/2021/6610675
32. Pontes, C. F., De Souza, M. M., Gondim, J. J., Bishop, M., & Marotta, M. A. (2021). A new method for flow-based network intrusion detection using the inverse Potts model. IEEE Transactions on Network and Service Management, 18(2), 1125-1136. https://doi.org/10.1109/TNSM.2021.3075503
33. Başarslan, M. S., & Kayaalp, F. (2023). Sentiment analysis with ensemble and machine learning methods in multi-domain datasets. Turkish Journal of Engineering, 7(2), 141-148. https://doi.org/10.31127/tuje.1079698
34. Rezaei, S., & Liu, X. (2019). Deep learning for encrypted traffic classification: An overview. IEEE Communications Magazine, 57(5), 76-81. https://doi.org/10.1109/MCOM.2019.1800819
35. Dirik, M. (2023). Machine learning-based lung cancer diagnosis. Turkish Journal of Engineering, 7(4), 322-330. https://doi.org/10.31127/tuje.1180931
36. Sharma, A., Malacaria, P., & Khouzani, M. H. R. (2019, June). Malware detection using 1-dimensional convolutional neural networks. In 2019 IEEE European symposium on security and privacy workshops (EuroS&PW), 247-256. https://doi.org/10.1109/EuroSPW.2019.00034
37. Azizjon, M., Jumabek, A., & Kim, W. (2020, February). 1D CNN based network intrusion detection with normalization on imbalanced data. In 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 218-224. https://doi.org/10.1109/ICAIIC48513.2020.9064976
38. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data, 8, 1-74. https://doi.org/10.1186/s40537-021-00444-8
39. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735
40. Dang, D., Di Troia, F., & Stamp, M. (2021). Malware classification using long short-term memory models. Cryptography and Security,1-16. https://doi.org/10.48550/arXiv.2103.02746

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Effective adsorption of malachite green with silica gel supported iron-zinc bimetallic nanoparticles

Memduha Ergüt [*1], **Ayla Özer** [2]

[1] *Sivas University of Science and Technology, Department of Chemical Engineering, Türkiye, memduha.ergut@gmail.com*
[2] *Mersin University, Department of Chemical Engineering, Türkiye, ayozer@mersin.edu.tr*

Cite this study: Ergüt, M., & Özer, A. (2024). Effective adsorption of malachite green with silica gel supported iron-zinc bimetallic nanoparticles. Turkish Journal of Engineering, 8 (3), 510-523

https://doi.org/10.31127/tuje.1413970

Keywords
Adsorption
Malachite green
Bimetallic nanoparticles
Silica gel
Composite material

Research Article

Received: 03.01.2024
Revised: 19.02.2024
Accepted: 21.02.2024
Published: 05.07.2024

Abstract

In the present study, a composite material consisting of silica gel 60 supported iron-zinc bimetallic nanoparticles (Si/Fe-ZnNPs) was prepared and characterized by SEM, EDX, FTIR, and XRD analysis. The adsorbent properties of the synthesized composite material were evaluated with the removal of Malachite Green (MG). According to characterization results, the cubic structures and agglomerated nano-sized spherical particles (≈30 nm) were formed. The FT-IR spectrum confirms the formation of Fe-Zn NPs through the observation of the Fe-O stretches and metal-metal stretching vibrations of (Zn^{2+} – O^{2-}) adsorption bands. Additionally, the FT-IR revealed the presence of Si-O-Si, Si-O-H stretching, and O-Si-O bending vibrations attributed to silica gel 60. The optimal environmental conditions for adsorption were determined to be a natural pH of 3.3, an adsorption temperature of 50°C, and an adsorbent concentration of 1.0 g/L. An increase in equilibrium uptakes of MG was observed with a linear correlation to initial dye concentrations. Thermodynamic studies indicated that the adsorption process was endothermic, non-spontaneous, and increasing disorder at the solid-solution interface during adsorption with positive ΔH, ΔG, and ΔS values, respectively. The experimental results revealed that the Langmuir isotherm model provided the best fit for the equilibrium data. The maximum monolayer coverage capacity of Si/Fe-ZnNPs was 666.67 mg/g at an optimum temperature of 50°C. Further analysis displayed that the kinetic adsorption data adhere to the pseudo-second-order kinetic model. Additionally, Weber-Morris model results revealed the effectiveness of both the film and intra-particle diffusion in the adsorption.

1. Introduction

Today, the escalating worldwide contamination of natural water from chemicals is a foremost challenge that humanity confronts as it produces considerable and unpredictable short and long-term impacts on aquatic life and human well-being [1]. The most important sources of water pollution are industrial chemicals, pesticides, fertilizers, oil sludge, and radioactive, solid, and commercial wastes. The common source of water pollution is human activity, although some natural and anthropogenic sources can cause it [2]. Many pollutants with adverse health effects such as textile dyestuffs, pesticides, pharmaceutical compounds, plasticizing agents, disinfection by-products, polychlorinated biphenyls, and polycyclic aromatic hydrocarbons have been determined in wastewater discharge [3].

Textile dyestuffs have attracted huge attention because of potential health and environmental hazards. Dyes possess intricate molecular structures that include chromophores and auxochromes. Chromophores are accountable for the color of the dye, while auxochromes comprising -NH_3, -COOH, -HSO_3, or -OH groups control the tint's intensity and make the molecule soluble in water [4]. The yearly global production of dyes surpasses $7x10^5$ tonnes, and industrial pollutants expel 10-15% of them into water sources. Even at a concentration of only 1 mg/L, dyes and their intermediates are the main and most dangerous pollutants found in the effluent produced by the textile industry [5].

Malachite green (MG) is a cationic dye that belongs to the triphenyl methane family. It is one of the most widely used industrial dyes in the manufacturing industries such as textile, paper, and food processing. Its uncontrolled discharge brings about harmful effects on the aquatic environment and human health due to its toxic, cancerogenic, and mutagenic characteristics [6].

Different treatment methods such as adsorption, ion exchange, electrochemical, catalysis, ozonation, reverse osmosis, coagulation, chemical precipitation, membrane filtration, and biological process techniques have been used for discharging dyestuffs from wastewater [7]. Adsorption is still a well-known treatment method owing to its being cost-effective, ease of operation, high removal performance, and environment-friendly nature [8].

As a consequence of rapid developments in the field of nanotechnology, the number of studies on the use of reusable, highly efficient, and low-cost nanoparticles as an adsorbent due to their large specific surface areas are increasing. Nanomaterials have many advantages, such as low diffusion resistance, easy access to the reactants, and a large number of active centers due to their large surface area [9]. Bimetallic nanoparticles prepared by adding other metal-based materials to iron-based nanoparticles to increase the adsorption capacity of them show more effective properties compared to mono-metallic ones. Moreover, their nanostructure and magnetic properties contribute to adsorption performance collectively and synergistically in the treatment of various contaminants such as dyestuffs, phenolics, pharmaceuticals, and heavy metals [10].

Various iron-based bimetallic nanoparticles have been reported as adsorbents. It has been indicated that the addition of a second catalytic metal such as Zn, Ag, Ni, Cu, or Pd may enhance the adsorption rate compared to using iron nanoparticles alone [11]. Additionally, incorporating bimetallic nanoparticles onto solid supports such as SiO_2, activated carbon, multi-walled carbon nanotube, zeolite, alginate, hydroxyapatite, chitosan, and gum can enhance the mechanical robustness and enhance the uptake of adsorbent [12].

Several iron-based materials, including zero-valent iron nanoparticles (nZVI FeNPs) [13–15], iron-containing bimetallic nanoparticles such as Fe/Ni-NPs [16–18], Fe/Zn-NPs [11, 19], and Fe/Cu–NPs [20, 21] as well as porous materials containing iron, like zeolites, clays, activated carbon, carbon nanotubes, graphene-based materials, biochar, mesoporous silica and metal-organic frameworks (MOFs) have been tested as effective adsorbents for the removal of different types of textile dyestuffs [22–24].

Kumar et al. synthesized bimetallic Fe–Zn nanoparticles via a co-precipitation method and used them for the adsorptive removal of carcinogenic dye malachite green and Congo red. The Langmuir maximum adsorption capacity for MG and CR was found to be 21.74 and 28.56 mg/g, respectively [11].

Silica comes from a variety of sources including fumed silica, sodium metasilicate, silica gel, and tetraethyl orthosilicate [25]. Silica gel is a type of amorphous synthetic silica. It is a hydrophilic mesoporous material and is mainly composed of SiO_2. It is a rigid, continuous network of colloidal silica connected to very small grains of hydrated SiO_2. It is non-hazardous, low cost, and has a specific surface area of about 100-1000 m^2/g. Porous silica gels are widely used as adsorbent materials in various applications such as adsorption, cooling, dehumidification, gas separation, and desalination [26]. Moreover, mesoporous silica gel has wide applications as a catalyst and catalytic support due to its outstanding surface properties and porosities, which are higher than those of zeolites and clays [27]. Additionally, silica gel has been used as an adsorbent for treating wastewater due to its stability, large pore channels, possible reuse, lower price, excellent chemical stability, high surface area, and high mechanical resistance [28, 29]. The surface of silica gel exhibits high chemical reactivity due to the presence of silanol groups (Si-OH). Additionally, the large pore channels enable selective adsorption of specific pollutants [30]. Silica materials are considered effective adsorbents for cationic dye removal owing to hydrogen bonding with silanol groups allowing cationic and non-ionic surfactants to adsorb onto the silica surface [27]. Many studies have reported that styryl pyridinium dyes are adsorbed on the silica gel surface [31].

Several studies have been reported on the synthesis of silica-based adsorbents and the adsorption of MG. Mansa et al. carried out the preparation of modified silica NPs for MG adsorption [32]. In the study of Hasan-Zadeh et al. acid modified silica nanoparticles were employed for the treatment of MG in wastewater. The material has a diameter of approximately 20.12 nm and a pore volume of around 1.04 cm^3/g. The adsorption capacity was determined to be 116.3 mg/g at optimum pH 6.5, 10 mg/L initial dye concentration, and 60 min equilibrium time. The adsorption performance indicated that the process of MG adsorption may occur on a multilayer porous material [33].

Abderrahman [25] synthesized different zeolite nanostructures from waste aluminum cans by hydrothermal method using different silicon sources. The zeolite nanostructures were synthesized using fumed silica (ZF), sodium metasilicate (ZM), silica gel (ZS), and tetraethyl orthosilicate (ZT). The prepared zeolite nanostructures were utilized in the removal of MG dye from aqueous media. The maximum adsorption capacities (mg/g) of MG dye on (ZF, ZM, ZS, and ZT) adsorbents were reported as 226.757, 239.234, 29.744, and 25.221, respectively.

Samiey and Toosi [31] investigated the adsorption of MG on silica gel under varying conditions, including temperature (308-328 K), pH, NaCl concentration, and binary mixtures of 2-propanol with water. The maximum adsorption capacity was found to be 39.7 mg/g.

To our knowledge, no study has been reported so far on the synthesis of silica gel supported iron-zinc bimetallic nanoparticles (Si/Fe-Zn NPs) composite material and its use as an adsorbent in MG removal from aqueous media. The study aims to synthesize a useful composite material with a high adsorption capacity for MG through a combination of silica gel and Fe-ZnNPs.

Therefore, the purpose of this study was to synthesize and characterize a composite material consisting of silica gel 60 supported Fe-Zn bimetallic nanoparticles, and to investigate its effectiveness as an adsorbent in treatment of MG. The characterizations of the synthesized Si/Fe-ZnNPs were carried out using XRD, FTIR, SEM, and EDX analysis methods. Subsequently, the Si/Fe-ZnNPs were tested as an adsorbent for MG treatment in a batch adsorption system. The study examined the effects of environmental conditions such as adsorbent concentration, initial dye concentration, and

temperature. The research also investigated the adsorption thermodynamics, and modeling studies of equilibrium, kinetic, and mass transfer for MG adsorption.

2. Method

2.1. Materials

$FeCl_3.6H_2O$ (≥98.0%), $ZnCl_2.4H_2O$ (≥98.0%), Malachite Green oxalate (C.I. 42000), and NaOH (pellets, anhydrous, ≥99%) were provided from Merck. Silica gel (Silica gel 60 (SiO_2), for column chromatography, mean pore diameter 60 Å, 63–200 µm, spherical) was acquired from Sigma Aldrich. MG was obtained in its commercial pure form. It is a cationic type with a molecular weight of 365 g/mol and a maximum absorption wavelength (λmax) of 618 nm. Its chemical structure is depicted in Figure 1, [34].

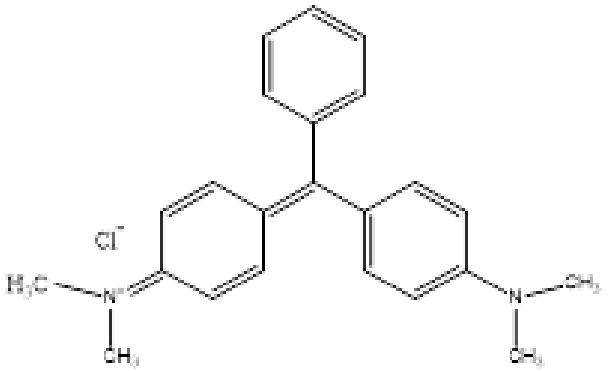


Figure 1. The molecular structure of Malachite Green.

2.2. Synthesis of composite material

Silica gel supported iron-zinc bimetallic nanoparticles composite material was synthesized by chemical co-precipitation and impregnation methods. Firstly, Fe-Zn bimetallic nanoparticles were produced through the chemical co-precipitation technique and silica gel 60 was mixed with the Fe-Zn bimetallic nanoparticles (Fe-ZnNPs) using the impregnation method. Fe-ZnNPs were prepared according to the procedure reported previously by Kaya et al. At room temperature, 25 mL of 0.4 M $FeCl_3.6H_2O$ and 25 mL of 0.2 M $ZnCl_2.4H_2O$ were mixed for 15 minutes. The solution's pH of 11 was adjusted by the inclusion of 25 mL of 3.0 M NaOH. The color of the resulting solution turned from transparent to deep black, indicating the formation of Fe-ZnNPs [35]. Then, silica gel 60 was mixed with Fe-ZnNPs using a simple impregnation method [36]. 0.5 grams of silica gel 60 were introduced into the bimetallic nanoparticles solution and the mixture was agitated for three hours at 25°C. The resulting composite material (Si/Fe-ZnNPs) was separated through centrifugation washed with deionized water and dried in an oven at 105°C over 12 hours.

2.3. Characterization studies

Functional groups of Si/Fe-ZnNPs were identified using Perkin Elmer model Fourier Transform Infrared (FT-IR) spectrometer in the range of 4000–400 cm^{-1} using a resolution of 4 cm^{-1} and 32 scans. The crystal phase was determined using X-ray powder diffraction (XRD) analysis using nickel-filtered Cu Kα radiation in a Philips XPert MPD apparatus operated at 40 kV and 30 mA, in the 2θ range of 10°–90°. The morphology was determined through Scanning Electron Microscope (SEM) analysis utilizing the Zeiss/Supra 55 SEM apparatus. The mean diameter of the Fe-Zn bimetallic nanoparticles was determined through Image J software. The recognition of elements and quantitative analysis of composite materials was carried out using the Energy Dispersive X-ray (EDX) technique using the Zeiss/Supra 55 EDX device. The pH at the point of zero charge (pH_{pzc}) of Si/Fe-ZnNPs was obtained by measuring the ζ (zeta) potential at different ranges of pH (2.0–7.0) with Zetasizer (Malvern Zetasizer Nano ZS model). The pH values of the Si/Fe-ZnNPs suspensions were adjusted from 2.0 to 7.0 by adding 0.1 M HCl or 0.1 M NaOH solutions as required.

2.4. Adsorption studies

The adsorption experiments were conducted using a shaking water bath (Memmert WB 29 model) and 250 mL Erlenmayers with a volume of 100 mL of MG dye solution. MG solutions were prepared with a known initial dye concentration at the natural pH (≈ 3.3) of MG.

Samples were taken before the addition of adsorbent for measuring initial MG dye concentrations. Afterward, 0.1 g of adsorbent, except for adsorbent concentration experiments, was contacted with 100 mL of dye solution at the known initial dye concentration and natural pH of MG. Then, the flasks were agitated at a constant temperature and shaking rate for 180 min which is more than ample time for adsorption equilibrium. Samples (3 mL) were withdrawn from the adsorption solution at pre-determined time intervals (2, 5, 10, 15, 30, 45, 60, 90, 120, and 180 min) and subjected to centrifugation at 3500 rpm for 5 min. The resultant supernatant was analyzed using a UV-vis spectrophotometer (Specord 210 Plus, Analytic Jena, Germany) with a wavelength of 618 nm employed to measure the concentration of unadsorbed dye present within the solution.

According to the difference between the initial and residual concentration of MG solution before and after adsorption, the adsorption capacity at time t [q_t; (mg/g)], equilibrium adsorption capacity [q_e; (mg/g)], and the percentage of adsorption [Adsorption (%)] at equilibrium were calculated with Equations (1 – 3).

$$q_t\ (mg/g) = \frac{C_0 - C_t}{m} \cdot V \quad (1)$$

$$q_e\ (mg/g) = \frac{C_0 - C_e}{m} \cdot V \quad (2)$$

$$\text{Adsorption}\ (\%) = \left[\frac{(C_0 - C_e)}{C_0}\right] \cdot 100 \quad (3)$$

Where C_o, C_t, and C_e represent the initial, time t, and equilibrium concentrations of MG solution, respectively, in units of mg/L. Additionally, m denotes the mass of the adsorbent in grams and V is the volume of the liquid phase in liters.

Experiments were carried out with different initial MG dye concentrations (100–500 mg/L), adsorbent concentrations (0.5–3 g/L), and temperatures (25–50°C) values at the natural pH (≈ 3.3) of MG solutions to determine the optimum adsorption conditions. The MG dye solutions ranging from 100–500 mg/L were prepared by diluting the 1.0 g/L stock solution of MG.

3. Results and Discussion

3.1. Characterization of composite material

3.1.1. SEM analysis

The morphology of Si/Fe ZnNPs was characterized by SEM analysis. In Figure 2 (a, b) and (c, d) the SEM images of the Si/Fe ZnNPs before and after adsorption are shown at different magnifications.

The SEM images (Figure 2 (a, b)) demonstrated that the agglomerated spherical nanosized (≈30 nm) particles and the cubic structures were formed. After MG adsorption, as shown in Figure 2 (c, d), it was observed that the morphological structure changed, the spherical nanoparticles were more agglomerated and the structures in the cubic forms disappeared.

3.1.2. EDX analysis

The EDX spectra before and after adsorption are shown in Figure 3 (a), and (b), and the elemental and quantitative weight composition (wt%) of Si/Fe-ZnNPs is shown in the inset of Figure 3 (a).

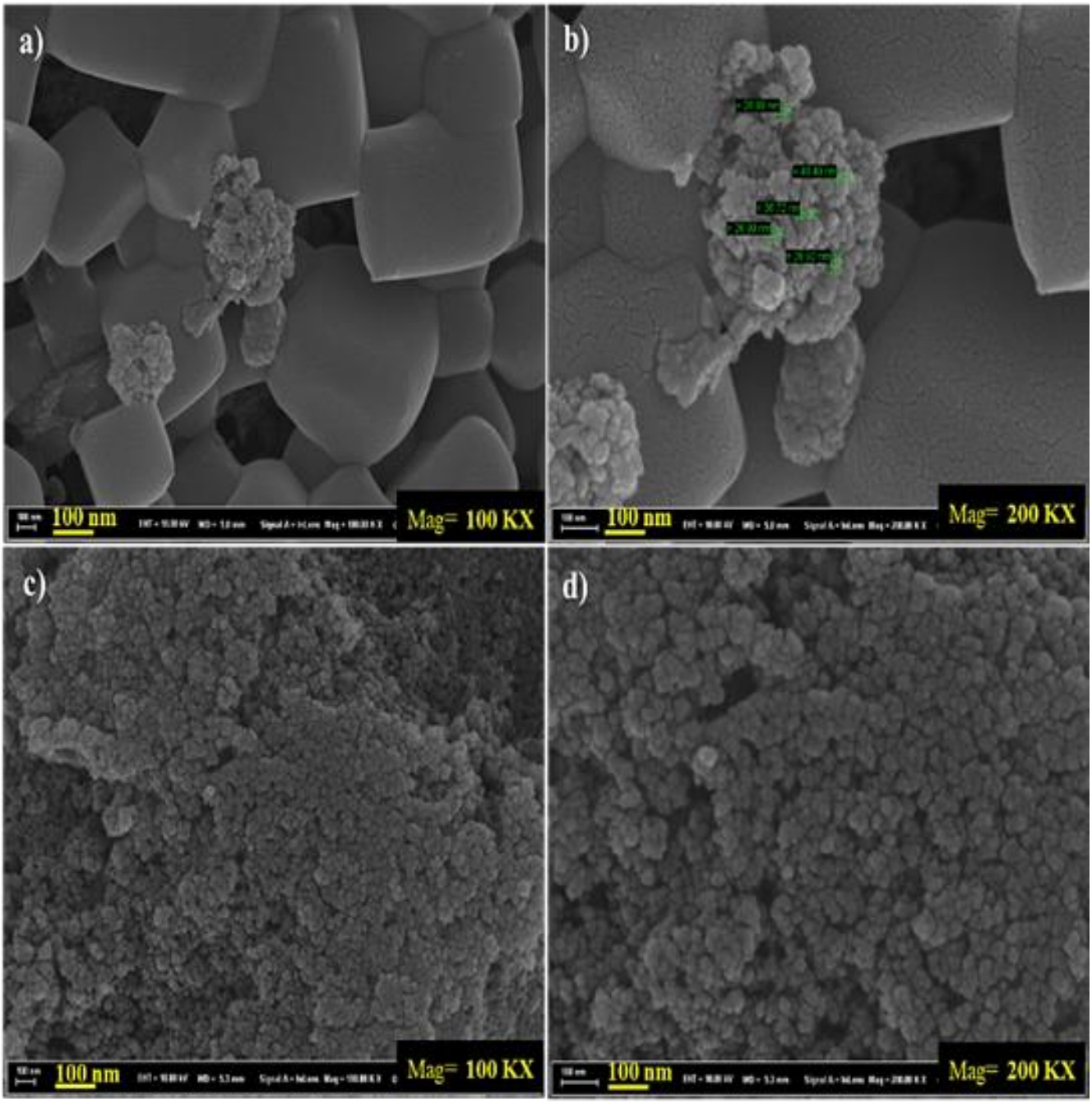


Figure 2. SEM images of Si/Fe-ZnNPs a) – 100 nm, 100 KX, b) – 100 nm, 200 KX before adsorption, c) – 100 nm, 100 KX, d) – 100 nm, 200 KX; after adsorption.

According to EDX analysis results, the weight composition (wt%) of Si/Fe-ZnNPs was determined as 6.78% Fe, 44.07% Zn, 2.44%Si, 8.56% O, and 38.14% Cl. It is worth mentioning that the data are the points on the adsorbent surface and are not the average content [37]. The formation of Si/Fe-ZnNPs was approved by the signals in Fe, Zn, O, and Si regions. Furthermore, following the adsorption (Figure 3(b)), the signals of C and N elements were also detected. This result confirmed that the dye molecules were adsorbed onto the surface of the adsorbent.

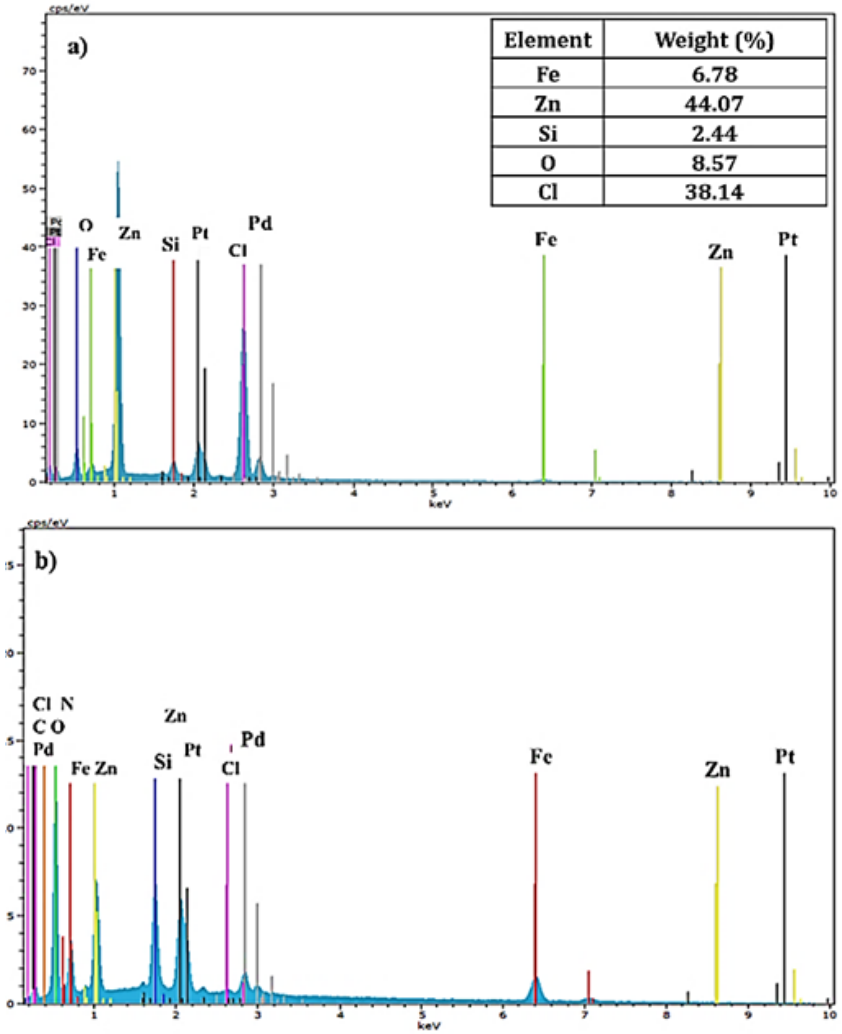


Figure 3. EDX spectra of composite material a) before adsorption and b) after adsorption.

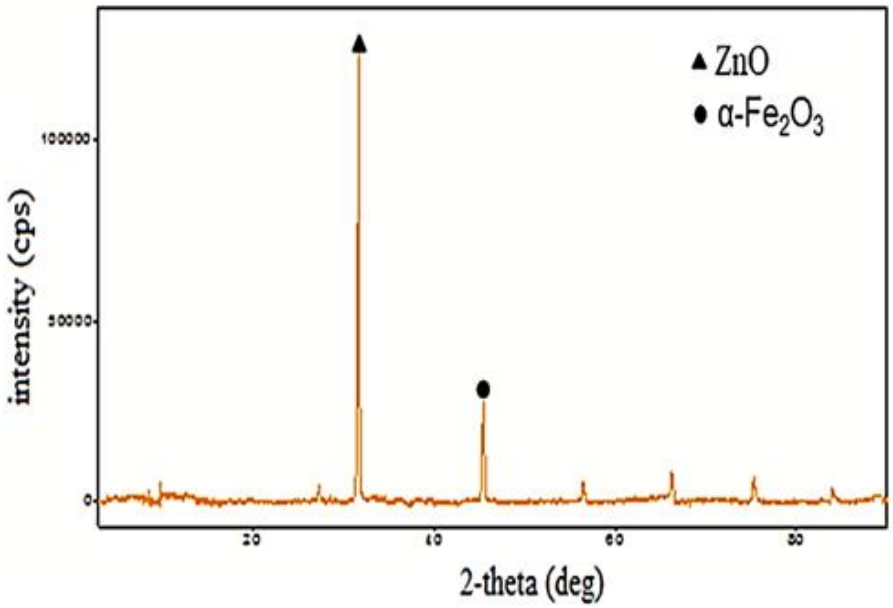


Figure 4. XRD pattern of synthesized Si/Fe-ZnNPs.

3.1.3. XRD analysis

The XRD pattern of synthesized Si/Fe-ZnNPs is given in Figure 4.

The peaks at 2θ = 34° and 45° correspond to the structures of ZnO and maghemite (α-Fe_2O_3), respectively Also, the weak peaks at 56°, and 66° corresponded to zinc oxide [38, 39]. The characteristic diffuse peak of silica is about 2θ = 21.6°. However, the characteristic peak of silica gel was not obtained in the XRD pattern.

This could demonstrate that a high percentage of structure is amorphous [40]. Moreover, this may be due to the low content and relatively low diffraction intensity of SiO_2 [41]. After adsorption, the XRD diagram showed that the material was completely amorphous. This was due to the change in crystallinity and phase.

3.1.4. FT-IR analysis

The FT-IR spectra of the Si/Fe-ZnNPs pre and post-adsorption can be seen in Figure 5, while feasible assignments are found in Table 1. The FT-IR spectra of Si/Fe-ZnNPs before and after adsorption did not show any significant changes as shown in Figure 5. Accordingly, the bands at 3650 cm^{-1}, 1618 cm^{-1}, and 954 cm^{-1} corresponded to O-H stretching vibration related to hydrogen bond, strain of υ(C-N) bond, and C – O stretching vibration, respectively. Furthermore, the absorption peaks at 522 cm^{-1} and 467 cm^{-1} correspond to Fe-O stretches of α-Fe_2O_3 and metal-metal stretching vibrations of (Zn^{2+} – O^{2-}) in the tetrahedral region, respectively, suggesting the formation of Fe-ZnNPs. At the absorption bands of 1067.8 cm^{-1} and 490.96 cm^{-1}, Si-O-H, Si-O-Si stretching, and O-Si-O bending vibrations were detected [42–45].

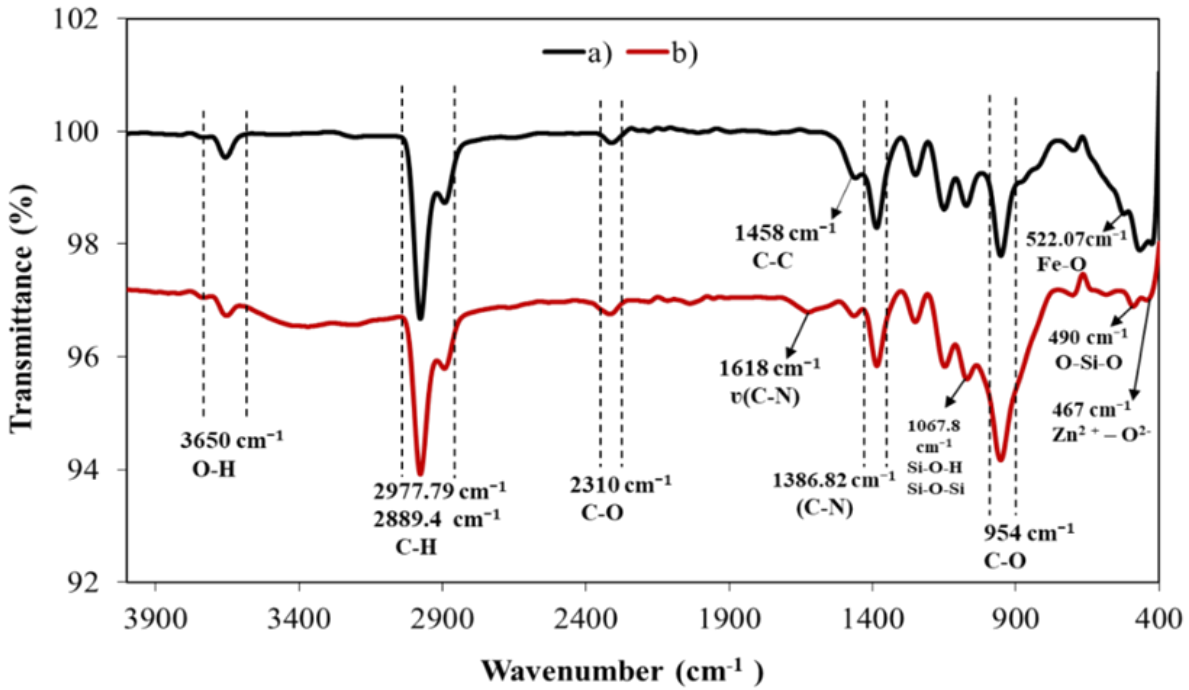


Figure 5. FT-IR spectrum of Si/Fe ZnNPs a) before b) after adsorption.

Table 1. FTIR spectra with suggested assignments.

Frequency (cm^{-1})	Assignment
3650	O-H stretching vibration related to hydrogen bond
2977.79	C-H stretching vibration
2889.4	C-H stretching vibration
2310.1	C-O bending
1618	Strain of υ(C-N) bond
1458	C–C stretching vibration
1386.82	Aromatic amine (C-N) stretching
1067.8	Si-O-H and Si-O-Si stretching vibrations
954	C – O stretching vibration
522	Fe-O stretches of α-Fe_2O_3
490.96	O-Si-O bending vibration
467	Metal-metal stretching vibration of (Zn^{2+} – O^{2-}) in tetrahedral region

3.1.5. Determination of pH of zero point of charge (pH_{zpc})

The isoelectric point (point of zero charge ((pH_{pzc})) of Si/Fe-ZnNPs was determined by zeta potential measurement at different pH (2.0 – 7.0) values and the change of the zeta potential with pH is presented in Figure 6.

The point of zero charge (pH_{pzc}) of Si/Fe-ZnNPs was determined to be 4.30, as shown in Figure 6. The point of

zero charge (pH_{pzc}) is the pH of the solution where the charge of the positive surface sites equals that of the negative sites, meaning the adsorbent surface charge is zero. The surface charge is negative at $pH>pH_{pzc}$ and positive at $pH<pH_{pzc}$ [46].

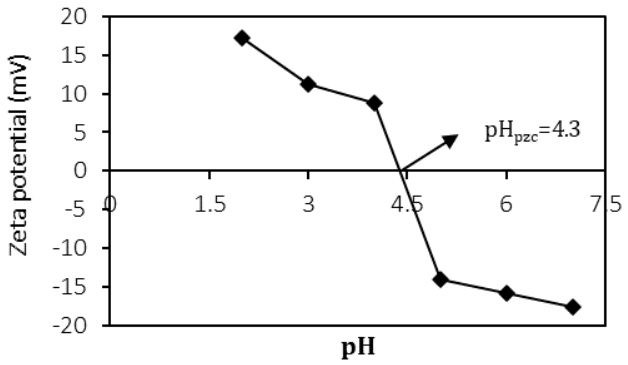


Figure 6. The isoelectric point of Si/Fe-ZnNPs.

3.2. Influence of environmental conditions

The effect of process parameters on adsorption was examined in a batch system, considering the initial pH of the solution, initial dye concentration, adsorbent concentration, and temperature. The results are presented in the following sub-sections.

3.2.1. Influence of initial pH

The initial pH of the solution is an important factor that affects the adsorption. The initial pH of the solution can alter the surface charge of the adsorbent, the degree to which the contaminants are ionized, and the structure of the dye molecules [47].

The determination of the point of zero charge (pH_{pzc}) of the adsorbent is necessary to gain a better understanding of the pH effect. The pH_{pzc} of Si/Fe-ZnNPs was determined to be 4.30, as indicated in Figure 6. This indicates that when adsorption occurs at pH values above the pH_{pzc} of the adsorbent, the surface of the adsorbent becomes negatively charged and the adsorption capacity increases due to the strong electrostatic attraction between the negatively charged surface of the adsorbent and the cationic dye. However, no pH adjustment was made in this study since; it was observed that dyestuff molecules collapsed when the adsorbent-free solutions of MG were adjusted to basic pH values over the pH_{pzc} of the adsorbent. The attitude of obtaining high adsorption capacity at a natural pH of 3.3, can be explained by the natural pH of the solution being near the pH_{pzc} of the adsorbent.

3.2.2. Influence of initial dye concentration

The influence of initial dye concentration on adsorption was investigated within the range of 100 – 500 mg/L MG dye concentration.

The initial dye concentration effect on adsorption is illustrated in Figure 7.

As seen from Figure 7, the equilibrium adsorption capacities increased linearly ($q_e=0.9745*C_o$, $R^2=0.999$) in the studied dye concentrations as a result of the increase in the driving force (ΔC) to get over mass transfer resistances of the dye molecules between the aqueous and solid phases [48]. Moreover, obtaining high equilibrium uptakes allows working in the wide dyestuff concentration ranges.

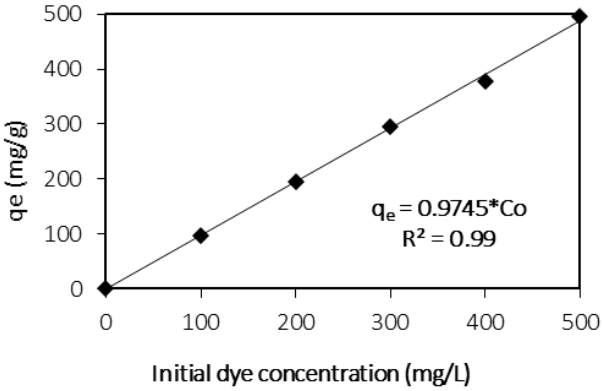


Figure 7. The effect of initial dye concentration (initial pH= natural pH≈ 3.3, X_0 =1 g/L, T= 50°C, t=180 min).

3.2.3. Influence of adsorbent concentration

The effect of adsorbent concentration was investigated in the range of 0.5 – 3.0 g/L adsorbent concentration. The adsorbent concentration effect on the equilibrium adsorption capacity and the percentage of adsorption was demonstrated in Figure 8.

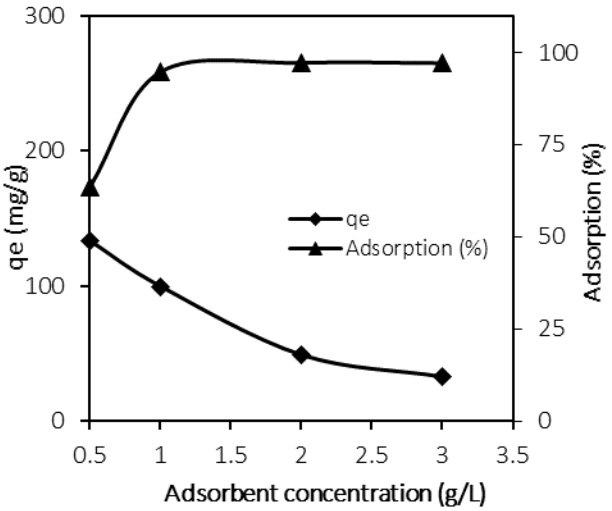


Figure 8. The impact of adsorbent concentration (initial pH= natural pH≈ 3.3, C_0 =100 mg/L, T=50°C, t=180 min).

Figure 8 illustrates the equilibrium adsorption capacities and adsorption percentages obtained for the adsorbent concentrations in the range from 0.5 g/L to 3 g/L. According to Figure 8, the equilibrium adsorption capacities decreased as the concentration of the adsorbents increased from 0.5 g/L to 3 g/L. Nevertheless, there was a marked increase in the adsorption percentage up to an adsorbent concentration of 1.0 g/L, which then became relatively constant with a further increase in adsorbent concentration. The decline in adsorption capacities as the adsorbent concentration increases could result from the agglomeration of

adsorbent particles caused by high adsorbent concentration. The agglomeration can cause a reduction in the active surface area of the adsorbent and an increase in the length of the diffusional pathway [48]. Thus, the reason for achieving lower adsorption capacities at 2 g/L and 3 g/L of adsorbent concentrations could be agglomeration. As a consequence of the experiment, the optimum adsorbent concentration was determined to be 1.0 g/L for the MG adsorption.

3.2.4. Influence of temperature

The effect of temperature was studied at three different values: 25°C, 35°C and 50°C values. The impact of the temperature is shown in Figure 9 below.

Figure 9 illustrates that a higher adsorption capacity was achieved with a temperature increase from 25°C to 50°C leading to the determination of the optimum temperature for MG adsorption as 50°C. The results indicate that the studied adsorption process is endothermic, given the high operation temperature.

The influence of environmental conditions such as initial pH, initial dye concentration, adsorbent concentration, time, and reaction temperature are summarized in Table 2. It presents a summary of the various adsorbents documented in scientific papers for the adsorption of MG. From Table 2, we can see that the value obtained in our study is significantly higher than the adsorption capacity observed in studies using silica gel and Fe-ZnNPs separately as adsorbents for MG removal.

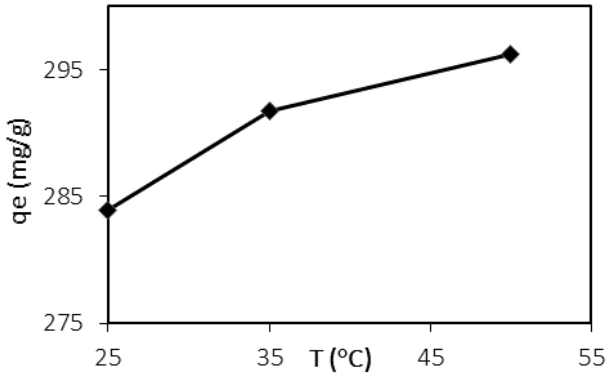


Figure 9. The effect of temperature (initial pH= natural pH≈ 3.3, C_0 =300 mg/L, X_0 =1 g/L, t=180 min).

Table 2. Effect of adsorption parameters of various adsorbents for MG removal in the literature.

Adsorbent	Initial pH	Initial Dye Conc. (mg/L)	Adsorbent Conc. (g/L)	Time (min)	Temp. (°C)	Adsorption Capacity (mg/g)	Removal Eff. (%)	Reference
Acid-functionalized silica nanoparticles	6.5	10	2.0	60	22	116.3	–	[33]
Mesoporous magnetic corn straw-derived biochar supported nZVI composite	6.0	100	0.25	20	25	515.77	99.9	[50]
Cellulose nanofibers and silver nanoparticles composites	8.0	100	0.5	100	30	142	78	[51]
Polymer hybrid bio-nanocomposite	7.0	150	1.0	40	50	384.615	99.79	[52]
GO@ZnO-$NiFe_2O_4$-αAl_2O_3 nanocomposites	7.0	1500	2.0	90	25	607.66	82	[53]
Calcium silicate nanopowders from waste materials	6.5	10	0.7	60	30	59.95	100	[54]
Multi-walled Carbon nanotube modified with Carboxylate group MWCNT-COOH	9.0	50	1.0	10	55	11.73	91.36	[55]
SDS/CTAB@Mt composite	7.0	1200	1.0	140	25	1021.45	99.32	[56]
Carbon dots/silica nano aggregates	6.0	200	0.05	2880	35	4091	–	[57]
Silica gel 60	2.3	29.2	3.5	1200	35	39.7	–	[31]
Fe-ZnNPs	9.0	25	3.0	60	30	28.56	92	[11]
Si/Fe-ZnNPs	No pH adjustment	100	1.0	180	50	666.67	96.81	This work

3.2.5. Thermodynamic studies

The effect of temperature was confirmed through thermodynamic parameters, including the change in Gibbs' free energy (ΔG), the change in enthalpy (ΔH), and the change in entropy (ΔS), which were calculated using the Van't Hoff equation. The Equations 4-5 were employed to determine ΔG, ΔH, and ΔS [49].

ΔH and ΔS were obtained by determining the slope and intercept of the linear plot of ln K_c against 1/T, by

Van't Hoff Equation 4. Furthermore, ΔG was computed following Equation 5.

The equilibrium constant (K_c) values were calculated utilizing Equation 6.

$$\ln K_c = \left(\frac{\Delta S}{R}\right) - \left(\frac{\Delta H}{R}\right) \cdot \frac{1}{T} \quad (4)$$

$$\Delta G = -RT \ln Kc = \Delta H - T\Delta S \quad (5)$$

$$K_c = \frac{C_{ad,e}}{C_e} \quad (6)$$

Where $C_{ad,e}$, and C_e represent the equilibrium concentrations (mg/L) of adsorbed and unadsorbed dyestuff ions, respectively. $C_{ad,e}$ is equal to q_e at the studied temperature when the adsorbent concentration is 1.0 g/L. In this study, we found the linear form of the Van't Hoff equation for adsorption to be $\ln K_c$ = -2348.9*1/T + 3.8537, with a regression coefficient of 0.999 (data not shown). Additionally, the calculated corresponding thermodynamic parameters are presented in Table 3.

As seen from Table 3, the positive ΔH value indicated the adsorption system was endothermic. The positive ΔG values showed that adsorption was non-spontaneous. Furthermore, a decrease in ΔG values with an increase in temperature showed that adsorption was more favorable at higher temperatures. The positive ΔS value reflected the irreversibility of adsorption and the randomness of the dye species adsorbed at the solid/solution interface.

Table 3. Thermodynamic parameters of adsorption.

T(K)	ΔG (J/mole)	ΔH (kJ/mole)	ΔS (J/mole.K)
298	9968.582		
308	9681.758	19.530	32.039
323	9170.201		

3.3. Modelling studies

3.3.1. Modelling adsorption equilibrium

The experimental equilibrium data for MG adsorption at the investigated temperatures were fitted with the renowned linearized forms of the Langmuir and Freundlich isotherm models.

The Langmuir isotherm presumes that the species adsorbed are in equilibrium with the adsorbent. The model relies on primary presumptions i) Adsorption occurs in a monolayer. ii) No interaction exists between the adsorbed molecules. iii) The surface is homogeneous, with identical binding sites [58].

The Freundlich isotherm model is a modification of the Langmuir isotherm that describes multilayer adsorption on heterogeneous surfaces [59].

The linearized forms of models are given in Equation 7:

$$\text{Langmuir: } \frac{1}{q_e} = \frac{1}{Q^\circ b} \cdot \frac{1}{C_e} + \frac{1}{Q^\circ} \quad (7)$$

$$\text{Freundlich: } \ln q_e = \ln K_F + \left(\frac{1}{n}\right) \ln C_e \quad (8)$$

where:

q_e is the adsorbed amount per unit mass of adsorbent (mg/g);

C_e is the unadsorbed dyestuff concentration at equilibrium (mg/L);

Q^o is the maximum monolayer coverage capacity of adsorbent (mg/g);

b is a Langmuir constant associated with binding site affinity (L/mg);

K_F is the adsorption capacity indicated by the Freundlich constant ((mg/g)/(L/mg)$^{1/n}$);

1/n is the Freundlich constant indicative of adsorption intensity.

The isothermal constants and their corresponding regression coefficients (R^2) were presented in Table 4 and the comparison of experimental and modeled isotherms at 50°C was given in Figure 10. Table 4 and Figure 10 demonstrate that the Langmuir isotherm model is in good agreement with the experimental equilibrium data, as indicated by the higher regression coefficients. The results show that a monolayer coating of MG was formed on the surface of Si/Fe-ZnNPs, indicating that the process took place at specific homogeneous sites within the adsorbent.

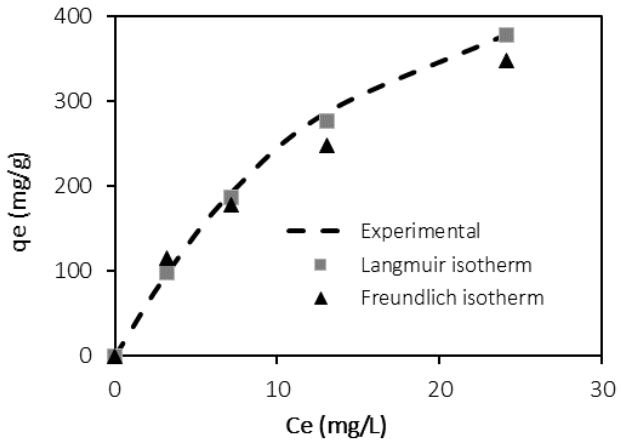


Figure 10. The comparison of the experimental and the predicted isotherms.

The maximum monolayer coverage capacity (Q°) of the adsorbent was determined to be 666.67 mg/g at 50 °C, which is the optimum temperature. As shown in Table 4, the maximum monolayer coverage capacity values of the adsorbent for MG increased with increasing temperature because of the adsorption process is endothermic. The maximum adsorption capacity (Q°) of Si/Fe ZnNPs was contrasted with that of Q° values of different adsorbents reported for MG adsorption. The comparison of Q° values is given in Table 5. Table 5 shows that the adsorption capacity of Si/Fe-ZnNPs was good when compared with other studies.

3.3.2. Kinetic modelling

Adsorption kinetics were explained by correlating adsorption kinetic data using linear forms of Lagergen's pseudo-first-order (PFO) [71] and pseudo-second order kinetic (PSO) models [72]. Equations 9 and 10 give the linearized forms of models:

Table 4. The constants of the isotherm models with the values of the regression coefficient (R^2).

T (°C)	Langmuir isotherm model [$q_e = Q^o b C_e / (1+bC_e)$]			Freundlich isotherm model [$q_e = K_F C_e^{1/n}$]		
	Q^o	b	R^2	K_F	1/n	R^2
25	476.19	0.1468	0.999	60.327	0.549	0.942
35	625.00	0.0994	0.999	80.356	0.544	0.944
50	666.67	0.05395	0.998	98.612	0.373	0.909

Table 5. Maximum adsorption capacity (Q°) of Si/Fe-ZnNPs with other reported adsorbents in the literature for MG dye adsorption.

Adsorbent	Langmuir isotherm model Q°(mg/g)	References
Modified rice husk	996.97	[60]
Fe_3O_4/β-CD/GO	740.74	[61]
nZVI magnetic composite supported by straw biochar (nZVI/BC)	515.77	[50]
Magnetic $CuFe_2O_4$ nano-adsorbent	197	[62]
ZnS: Cu nanoparticle-loaded activated carbon	168.1	[63]
Graphene oxide/aminated lignin aerogels	113.5	[64]
Natural zeolite	98.04	[65]
Rattan sawdust	62.71	[66]
Multi-walled carbon nanotubes functionalized with carboxylates	49.45	[67]
Sodium alginate-coated superparamagnetic Fe_3O_4 nanoparticles	47.84	[68]
Montmorillonite/PVDF/PEO microporous membranes	33.4	[69]
Zeolite synthesized using silica gel (ZS)	29.744	[25]
ZnO nanorod-loaded activated carbon	20	[70]
(Si/Fe-Zn NPs)	666.67	This study

$$\text{PFO: } \log(q_e - q_t) = \log(q_e) - \frac{k_1 t}{2.303} \quad (9)$$

$$\text{PSO: } \frac{t}{q_t} = \left(\frac{1}{q_e^2 k_2}\right) + \frac{t}{q_e} \quad (10)$$

where:

q_e is the adsorbed amount per unit mass of adsorbent (mg/g);

q_t is the adsorbed amount per unit mass of adsorbent at any time (mg/g);

k_1 is the pseudo-first-order kinetic rate constant (1/min);

k_2 is the sseudo-second-order kinetic rate constant (g/mg.min)

The parameters with regression coefficients of the PFO and PSO kinetic models are shown in Table 6.

The nomenclatures shown in Table 6 are as follows:

$q_{e,exp}$ is the experimental adsorbed amount per unit mass of adsorbent (mg/g);

q_{e,cal_1} is the calculated amount of adsorbed substance per unit mass of adsorbent from the pseudo-first-order kinetic model (mg/g);

q_{e,cal_2} is the calculated amount of adsorbed substance per unit mass of adsorbent from the pseudo-second-order kinetic model (mg/g).

From Table 6, an excellent fit of the PSO kinetics was shown by the higher R^2 values and the agreement between experimental and calculated equilibrium adsorption capacities. If the adsorption data conforms to the PSO kinetic model, chemisorption via electron exchange or sharing between functional groups on the sorbent and dyestuff ions predominantly occurs [72].

Table 6. Model parameters with R^2 values for kinetics and mass transfer.

C_o (mg/L)	$q_{e,exp}$ (mg/g)	PFO			PSO			Weber-Morris Model		
		k_1 (min^{-1})	q_{e,cal_1} (mg/g)	R^2	k_2 (g/mg.min)	q_{e,cal_2} (mg/g)	R^2	K_i ($mg/g.min^{0.5}$)	Intercept	R^2
100.96	97.75	0.0304	10.73	0.649	0.01553	97.75	0.999	0.1593	95.778	0.999
199.89	194.49	0.0278	24.67	0.575	0.01000	193.95	0.999	0.2026	189.95	0.998
300.53	292.75	0.0232	53.80	0.725	0.00031	275.99	0.999	4.7475	253.06	0.997
402.57	378.43	0.0315	138.36	0.913	0.00111	373.48	0.999	13.096	272.14	0.997
509.00	496.21	0.0039	361.83	0.910	0.00023	379.11	0.994	14.400	323.75	0.996

3.3.3. Mass transfer modelling

The Weber-Morris model was used to examine the interfacial layer and intraparticle diffusion mechanism between dye and adsorbent, and the model parameters and regression coefficients were summarized in Table 6. The Weber-Morris model is defined by Equation (11) [73].

$$q_t = K_i . t^{0.5} + I \quad (11)$$

In the Weber-Morris equation, I is the cut-off point associated with the interfacial effect, and Ki is the rate constant of intra-particle diffusion ((mg/g.min1/2)). The Ki value can be obtained by determining the slope of the linear graph of qt versus t0.5. This model suggests that if a linear plot of qe versus t0.5 lies through the origin, it suggests that intra-particle diffusion is the only rate-controlling step. Nevertheless, the line may not always cross the origin, as the diffusion kinetics of adsorption can be controlled simultaneously by film diffusion and pore diffusion. In this case, multilinear plots can be

obtained with an intercept value. Based on the results, a multilinear plot was observed in this study. This suggests that pore and film diffusion play a role in the adsorption process.

4. Conclusion

A composite material of Si/Fe-ZnNPs was synthesized and subsequently used in this study for MG adsorption. The material was characterized by means of SEM, EDX, FTIR, and XRD analyses. Adsorption studies were performed to determine the influence of operating parameters over the adsorption of MG. The study revealed that an adsorption percentage of 96.81% was obtained in 180 minutes, utilizing natural pH of solution (≈ 3.3), an adsorbent concentration of 1.0 g/L, and a temperature of 50°C with a 100 mg/L of MG. The Langmuir isotherm model was employed to model the adsorption equilibrium. This model assumes of monolayer adsorption on a homogeneous surface. According to the thermodynamic parameters, positive values of ΔG, ΔH, and ΔS indicate that the process was non-spontaneous and endothermic and that the disorder at the solid/solution interface increased during the process. The kinetic data are consistent with the PSO kinetic model, and mass transfer modeling studies showed that both intraparticle and film diffusion appear to be effective.

Consequently, the research findings demonstrate that the Si/Fe-ZnNPs composite material synthesized is an effective adsorbent for eliminating MG dye from aqueous solutions.

Author contributions

Memduha Ergüt: Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing – original draft, Visualization. **Ayla Özer:** Review & editing, Supervision, Funding acquisition

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Rathi, B. S., & Kumar, P. S. (2021). Application of adsorption process for effective removal of emerging contaminants from water and wastewater. Environmental pollution (Barking, Essex: 1987), 280, 116995. https://doi.org/10.1016/j.envpol.2021.116995
2. Ahmad, H. R., Aziz, T., Zia-ur-Rehman, M., Sabir, M., & Khalid, H. (2016). Sources and composition of waste water: Threats to plants and soil health. In Soil Science: Agricultural and Environmental Prospectives (349–370). Springer International Publishing. https://doi.org/10.1007/978-3-319-34451-5_16
3. Senthil Kumar, P., Joshiba, G. J., Femina, C. C., Varshini, P., Priyadharshini, S., Arun Karthick, M. S., & Jothirani, R. (2019). A critical review on recent developments in the low-cost adsorption of dyes from wastewater. Desalination And Water Treatment, 172, 395–416. https://doi.org/10.5004/dwt.2019.24613
4. Ahmouda, K., Boudiaf, M., & Benhaoua, B. (2022). A novel study on the preferential attachment of chromophore and auxochrome groups in azo dye adsorption on different greenly synthesized magnetite nanoparticles: investigation of the influence of the mediating plant extract's acidity. Nanoscale Advances, 4(15),3250–3271. https://doi.org/10.1039/d2na00302c
5. Oruç, Z., Ergüt, M., Uzunoğlu, D., & Özer, A. (2019). Green synthesis of biomass-derived activated carbon/Fe-Zn bimetallic nanoparticles from lemon (Citrus limon (L.) Burm. f.) wastes for heterogeneous Fenton-like decolorization of Reactive Red 2. Journal of Environmental Chemical Engineering, 7(4), 103231. https://doi.org/10.1016/j.jece.2019.103231
6. Swan, N. B., & Zaini, M. A. A. (2019). Adsorption of Malachite Green and Congo Red Dyes from Water: Recent Progress and Future Outlook. Ecological Chemistry and Engineering S, 26(1), 119–132. https://doi.org/10.1515/eces-2019-0009
7. M. Nahiun, K., Sarker, B., N. Keya, K., I. Mahir, F., Shahida, S., & A. Khan, R. (2021). A Review on the Methods of Industrial Waste Water Treatment. Scientific Review, 7(73), 20–31. https://doi.org/10.32861/sr.73.20.31
8. Tulun, Ş., Bahadir, T., Şimşek, İ., & Karataş, M. (2019). The removal of nickel ions with walnut shell. Turkish Journal of Engineering, 3(2), 102-105. https://doi.org/10.31127/tuje.456741.
9. Sharma, G., Kumar, A., Sharma, S., Naushad, M., Prakash Dwivedi, R., ALOthman, Z. A., & Mola, G. T. (2019). Novel development of nanoparticles to bimetallic nanoparticles and their composites: A review. Journal of King Saud University - Science, 31(2), 257–269. https://doi.org/10.1016/j.jksus.2017.06.012
10. Scaria, J., Nidheesh, P. V, & Kumar, M. S. (2020). Synthesis and applications of various bimetallic nanomaterials in water and wastewater treatment. Journal of Environmental Management, 259,110011. https://doi.org/10.1016/j.jenvman.2019.110011
11. Gautam, R. K., Rawat, V., Banerjee, S., Sanroman, M. A., Soni, S., Singh, S. K., & Chattopadhyaya, M. C. (2015). Synthesis of bimetallic Fe-Zn nanoparticles and its application towards adsorptive removal of carcinogenic dye malachite green and Congo red in water. Journal of Molecular Liquids, 212, 227–236. https://doi.org/10.1016/j.molliq.2015.09.006
12. Kharissova, O. V, Dias, R., & Kharisov, B. I. (2015). Magnetic adsorbents on the basis of micro- and nanostructurized materials. RSC Advances, 5, 6695–6719.
13. Chen, Z. X., Cheng, Y., Chen, Z., Megharaj, M., & Naidu, R. (2014). Kaolin-supported nanoscale zero-valent iron for removing cationic dye–crystal violet in aqueous solution. In Nanotechnology for Sustainable Development, First Edition, 189–196. https://doi.org/10.1007/978-3-319-05041-6_15

14. Hamdy, A., Mostafa, M. K., & Nasr, M. (2018). Zero-valent iron nanoparticles for methylene blue removal from aqueous solutions and textile wastewater treatment, with cost estimation. Water Science and Technology, 78(2), 367–378. https://doi.org/10.2166/wst.2018.306
15. Sahu, N., Rawat, S., Singh, J., Karri, R. R., Lee, S., Choi, J.-S., & Koduru, J. R. (2019). Process Optimization and Modeling of Methylene Blue Adsorption Using Zero-Valent Iron Nanoparticles Synthesized from Sweet Lime Pulp. Applied Sciences, 9(23), 5112. https://doi.org/10.3390/app9235112
16. Naser, R., & Shahwan, T. (2019). Comparative assessment of the decolorization of aqueous bromophenol blue using Fe nanoparticles and Fe-Ni bimetallic nanoparticles. Desalination and Water Treatment, 159, 346-355. https://doi.org/10.5004/dwt.2019.24136
17. Alruqi, S. S., AL-Thabaiti, S. A., & Khan, Z. (2019). Iron-nickel bimetallic nanoparticles: Surfactant assisted synthesis and their catalytic activities. Journal of Molecular Liquids, 282, 448–455. https://doi.org/10.1016/j.molliq.2019.03.021
18. Bokare, A. D., Chikate, R. C., Rode, C. V., & Paknikar, K. M. (2008). Iron-nickel bimetallic nanoparticles for reductive degradation of azo dye Orange G in aqueous solution. Applied Catalysis B: Environmental, 79(3), 270–278. https://doi.org/10.1016/j.apcatb.2007.10.033
19. Tahir, H., Saad, M., Attala, O. A., El-Saoud, W. A., Attia, K. A., Jabeen, S., & Zeb, J. (2023). Sustainable Synthesis of Iron–Zinc Nanocomposites by Azadirachta indica Leaves Extract for RSM-Optimized Sono-Adsorptive Removal of Crystal Violet Dye. Materials, 16(3), 1023. https://doi.org/10.3390/ma16031023
20. Ulucan-Altuntas, K., & Kuzu, S. L. (2019). Modelling and optimization of dye removal by Fe/Cu bimetallic nanoparticles coated with different Cu ratios. Materials Research Express, 6(11), 1150a4. https://doi.org/10.1088/2053-1591/ab4bb5
21. Mahmoud, A. S., Mostafa, M. K., & Peters, R. W. (2021). A prototype of textile wastewater treatment using coagulation and adsorption by Fe/Cu nanoparticles: Techno-economic and scaling-up studies. Nanomaterials and Nanotechnology, 11, 18479804211041181.https://doi.org/10.1177/18479804211041181
22. Raman, C. D., & Kanmani, S. (2016). Textile dye degradation using nano zero valent iron: A review. Journal of Environmental Management, 177, 341–355. https://doi.org/10.1016/j.jenvman.2016.04.034
23. Bhattacharya, S., Saha, I., Mukhopadhyay, A., Chattopadhyay, D., & Chand, U. (2013). Role of nanotechnology in water treatment and purification: Potential applications and implications. International Journal of Chemical Science and Technology, 3(3), 59–64.
24. Kharisov, B. I., Dias, H. R., Kharissova, O. V., Jiménez-Pérez, V. M., Pérez, B. O., & Flores, B. M. (2012). Iron-containing nanomaterials: synthesis, properties, and environmental applications. Rsc Advances, 2(25), 9325-9358.
25. Abdelrahman, E. A. (2018). Synthesis of zeolite nanostructures from waste aluminum cans for efficient removal of malachite green dye from aqueous media. Journal of Molecular Liquids, 253, 72–82. https://doi.org/10.1016/j.molliq.2018.01.038
26. Maher, H., Rupam, T. H., Rocky, K. A., Bassiouny, R., & Saha, B. B. (2022). Silica gel-MIL 100(Fe) composite adsorbents for ultra-low heat-driven atmospheric water harvester. Energy, 238, 121741. https://doi.org/10.1016/j.energy.2021.121741
27. Boukoussa, B., Mokhtar, A., El Guerdaoui, A., Hachemaoui, M., Ouachtak, H., Abdelkrim, S., Addi Ait, A., Babou, S., Boudina, B., Bengueddach, A., Hamacha, R. (2021). Adsorption behavior of cationic dye on mesoporous silica SBA-15 carried by calcium alginate beads: Experimental and molecular dynamics study. Journal of Molecular Liquids, 333, 115976. https://doi.org/10.1016/j.molliq.2021.115976
28. Volikov, A. B., Ponomarenko, S. A., Konstantinov, A. I., Hatfield, K., & Perminova, I. V. (2016). Nature-like solution for removal of direct brown 1 azo dye from aqueous phase using humics-modified silica gel. Chemosphere, 145, 83–88. https://doi.org/10.1016/j.chemosphere.2015.11.070
29. Patra, A. S., Ghorai, S., Sarkar, D., Das, R., Sarkar, S., & Pal, S. (2017). Anionically functionalized guar gum embedded with silica nanoparticles: An efficient nanocomposite adsorbent for rapid adsorptive removal of toxic cationic dyes and metal ions. Bioresource Technology, 225, 367–376. https://doi.org/10.1016/j.biortech.2016.11.093
30. Zhang, Y., Xia, K., Liu, X., Chen, Z., Du, H., & Zhang, X. (2019). Synthesis of cationic-modified silica gel and its adsorption properties for anionic dyes. Journal of the Taiwan Institute of Chemical Engineers, 102, 1–8. https://doi.org/10.1016/j.jtice.2019.05.005
31. Samiey, B., & Toosi, A. R. (2010). Adsorption of malachite green on silica gel: effects of NaCl, pH and 2-propanol. Journal of hazardous materials, 184(1–3), 739–745. https://doi.org/10.1016/j.jhazmat.2010.08.101
32. Mansa, R. F., Sipaut, C. S., Rahman, I. A., Yusof, N. S. M., & Jafarzadeh, M. (2016). Preparation of glycine–modified silica nanoparticles for the adsorption of malachite green dye. Journal of Porous Materials, 23(1), 35–46. https://doi.org/10.1007/s10934-015-0053-3
33. Hassan-Zadeh, B., Rahmanian, R., Salmani, M. H., & Salmani, M. J. (2021). Functionalization of Synthesized Nanoporous Silica and Its Application in Malachite Green Removal from Contaminated Water. Journal of Environmental Health and Sustainable Development, 6(2), 1311–1320. https://doi.org/10.18502/jehsd.v6i2.6542
34. Hossain, M. A., Hossain, M. L., & Hassan, T. Al. (2016). Equilibrium, Thermodynamic and Mechanism Studies of Malachite Green Adsorption on Used Black Tea Leaves from Acidic Solution. International Letters of Chemistry, Physics and Astronomy, 64(2018), 77–88. https://doi.org/10.56431/p-c20qfs
35. Kaya, M., Zahmakiran, M., Özkar, S., & Volkan, M. (2012). Copper(0) nanoparticles supported on silica-

coated cobalt ferrite magnetic particles: Cost effective catalyst in the hydrolysis of ammonia-borane with an exceptional reusability performance. ACS Applied Materials and Interfaces, 4(8), 3866–3873. https://doi.org/10.1021/am3005994

36. Wang, J., Liu, C., Hussain, I., Li, C., Li, J., Sun, X., Shen, J., Han, W., Wang, L. (2016). Iron-copper bimetallic nanoparticles supported on hollow mesoporous silica spheres: The effect of Fe/Cu ratio on heterogeneous Fenton degradation of a dye. RSC Advances, 6(59), 54623–54635. https://doi.org/10.1039/c6ra08501f
37. Huo, Y., Ding, W., Huang, X., Xu, J., & Zhao, M. (2011). Fluoride removal by lanthanum alginate bead: Adsorbent characterization and adsorption mechanism. Chinese Journal of Chemical Engineering, 19(3), 365–370. https://doi.org/10.1016/S1004-9541(09)60222-6
38. Kumar, A., Rana, A., Sharma, G., Naushad, M., Dhiman, P., Kumari, A., & Stadler, F. J. (2019). Recent advances in nano-Fenton catalytic degradation of emerging pharmaceutical contaminants. Journal of Molecular Liquids, 290, 111177. https://doi.org/10.1016/j.molliq.2019.111177
39. Sun, S., Yang, X., Zhang, Y., Zhang, F., Ding, J., Bao, J., & Gao, C. (2012). Enhanced photocatalytic activity of sponge-like ZnFe2O4 synthesized by solution combustion method. Progress in Natural Science: Materials International, 22(6), 639–643. https://doi.org/10.1016/j.pnsc.2012.11.008
40. Ullah, R., Deb, B. K., Yousuf, M., & Mollah, A. (2014). Synthesis and Characterization of Silica Coated Iron-Oxide Composites of Different Ratios. International Journal of Composite Materials, 4(2), 135–145. https://doi.org/10.5923/j.cmaterials.20140402.13
41. Yu, J., Jin, J., Cheng, B., & Jaroniec, M. (2014). A noble metal-free reduced graphene oxide-cds nanorod composite for the enhanced visible-light photocatalytic reduction of CO2 to solar fuel. Journal of Materials Chemistry A, 2(10), 3407–3416. https://doi.org/10.1039/c3ta14493c
42. Muhammad, W., Ullah, N., Haroon, M., & Abbasi, B. H. (2019). Optical, morphological and biological analysis of zinc oxide nanoparticles (ZnO NPs) using: Papaver somniferum L. RSC Advances, 9(51), 29541–29548. https://doi.org/10.1039/c9ra04424h
43. Yedurkar, S., Maurya, C., & Mahanwar, P. (2016). Biosynthesis of Zinc Oxide Nanoparticles Using Ixora Coccinea Leaf Extract—A Green Approach. Open Journal of Synthesis Theory and Applications, 05(01), 1–14. https://doi.org/10.4236/ojsta.2016.51001
44. Wang, P., Wang, X., Yu, S., Zou, Y., Wang, J., Chen, Z., Alharbi, S. N., Alsaedi, A., Hayat, T., Chen, Y., & Wang, X. (2016). Silica coated Fe 3 O 4 magnetic nanospheres for high removal of organic pollutants from wastewater. Chemical Engineering Journal, 306, 280–288. https://doi.org/10.1016/j.cej.2016.07.068
45. Konicki, W., Sibera, D., Mijowska, E., Lendzion-Bieluń, Z., & Narkiewicz, U. (2013). Equilibrium and kinetic studies on acid dye Acid Red 88 adsorption by magnetic ZnFe2O4 spinel ferrite nanoparticles. Journal of colloid and interface science, 398, 152–60. https://doi.org/10.1016/j.jcis.2013.02.021
46. Kragović, M., Stojmenović, M., Petrović, J., Loredo, J., Pašalić, S., Nedeljković, A., & Ristović, I. (2019). Influence of Alginate Encapsulation on Point of Zero Charge (pHpzc) and Thermodynamic Properties of the Natural and Fe(III) - Modified Zeolite. Procedia Manufacturing, 32, 286–293. https://doi.org/10.1016/j.promfg.2019.02.216
47. Banerjee, S., & Chattopadhyaya, M. C. (2017). Adsorption characteristics for the removal of a toxic dye, tartrazine from aqueous solutions by a low cost agricultural by-product. Arabian Journal of Chemistry, 10, S1629–S1638. https://doi.org/10.1016/j.arabjc.2013.06.005
48. Musa, M., Hasan, H., Malkoç, H., Ergüt, M., Uzunoğlu, D., & Özer, A. (2020). Effective adsorption of tetracycline with Co_3O_4/Fe_3O_4 bimetallic nanoparticles. Turkish Journal of Engineering, 4(4), 209-217. https://doi.org/10.31127/tuje.648882
49. Meena, A. K., Kadirvelu, K., Mishraa, G. K., Rajagopal, C., & Nagar, P. N. (2008). Adsorption of Pb(II) and Cd(II) metal ions from aqueous solutions by mustard husk. Journal of Hazardous Materials, 150(3), 619–625. https://doi.org/10.1016/j.jhazmat.2007.05.011
50. Eltaweil, A. S., Ali Mohamed, H., Abd El-Monaem, E. M., & El-Subruiti, G. M. (2020). Mesoporous magnetic biochar composite for enhanced adsorption of malachite green dye: Characterization, adsorption kinetics, thermodynamics and isotherms. Advanced Powder Technology, 31(3), 1253–1263. https://doi.org/10.1016/j.apt.2020.01.005
51. Chinthalapudi, N., Kommaraju, V. V. D., Kannan, M. K., Nalluri, C. B., & Varanasi, S. (2021). Composites of cellulose nanofibers and silver nanoparticles for malachite green dye removal from water. Carbohydrate Polymer Technologies and Applications, 2, 100098. https://doi.org/10.1016/j.carpta.2021.100098
52. Sarojini, G., Venkatesh Babu, S., Rajamohan, N., & Rajasimman, M. (2022). Performance evaluation of polymer-marine biomass based bionanocomposite for the adsorptive removal of malachite green from synthetic wastewater. Environmental Research, 204, 112132. https://doi.org/10.1016/j.envres.2021.112132
53. Hojjati-Najafabadi, A., Nasr Esfahani, P., Davar, F., Aminabhavi, T. M., & Vasseghian, Y. (2023). Adsorptive removal of malachite green using novel GO@ZnO-NiFe2O4-αAl2O3 nanocomposites. Chemical Engineering Journal, 471(May), 144485. https://doi.org/10.1016/j.cej.2023.144485
54. Hashem, A. A., Mahmoud, S. A., Geioushy, R. A., & Fouad, O. A. (2023). Adsorption of malachite green dye over synthesized calcium silicate nanopowders from waste materials. Materials Science and Engineering: B, 295(December 2022), 116605. https://doi.org/10.1016/j.mseb.2023.116605
55. Rajabi, M., Mirza, B., Mahanpoor, K., Mirjalili, M., Najafi, F., Moradi, O., Sadegh, H., Shahryari-ghoshekandi, H. R., Asif, M., Tyagi, I., Agarwal, S., Gupta, V. K. (2016). Adsorption of malachite green from aqueous solution by carboxylate group functionalized multi-walled carbon nanotubes: Determination of equilibrium and kinetics

parameters. Journal of Industrial and Engineering Chemistry, 34, 130–138. https://doi.org/10.1016/j.jiec.2015.11.001

56. Haounati, R., Ouachtak, H., El Haouti, R., Akhouairi, S., Largo, F., Akbal, F., Benlhachemi, A., Jada, A., Addi, A. A. (2021). Elaboration and properties of a new SDS/CTAB@Montmorillonite organoclay composite as a superb adsorbent for the removal of malachite green from aqueous solutions. Separation and Purification Technology, 255(2020), 117335. https://doi.org/10.1016/j.seppur.2020.117335
57. Li, X. Y., Wang, W. R., Xue, R. C., Chen, P. Y., Wang, Y., & Yu, L. P. (2023). Carbon dots/silica nanoaggregates for highly efficient adsorption of alizarin red S and malachite green dyes. New Journal of Chemistry, 47(18), 8965–8973. https://doi.org/10.1039/d3nj01273e
58. Langmuir, I. (1918). The adsorption of gases on plane surfaces of glass, mica and platinum. Journal of the American Chemical Society, 40(9), 1361–1403. https://doi.org/10.1021/ja02242a004
59. Freundlich, H. (1907). Über die Adsorption in Lösungen. Zeitschrift für Physikalische Chemie, 57U(1), 385–470. https://doi.org/10.1515/zpch-1907-5723
60. You, X., Zhou, R., Zhu, Y., Bu, D., & Cheng, D. (2022). Adsorption of dyes methyl violet and malachite green from aqueous solution on multi-step modified rice husk powder in single and binary systems: Characterization, adsorption behavior and physical interpretations. Journal of Hazardous Materials, 430, 128445. https://doi.org/10.1016/j.jhazmat.2022.128445
61. Wang, D., Liu, L., Jiang, X., Yu, J., & Chen, X. (2015). Adsorption and removal of malachite green from aqueous solution using magnetic β-cyclodextrin-graphene oxide nanocomposites as adsorbents. Colloids and Surfaces A: Physicochemical and Engineering Aspects, 466, 166–173. https://doi.org/10.1016/j.colsurfa.2014.11.021
62. Vergis, B. R., Hari Krishna, R., Kottam, N., Nagabhushana, B. M., Sharath, R., & Darukaprasad, B. (2018). Removal of malachite green from aqueous solution by magnetic CuFe2O4 nano-adsorbent synthesized by one pot solution combustion method. Journal of Nanostructure in Chemistry, 8(1), 1–12. https://doi.org/10.1007/s40097-017-0249-y
63. Dastkhoon, M., Ghaedi, M., Asfaram, A., Goudarzi, A., Langroodi, S. M., Tyagi, I., Agarwall, S., & Gupta, V. K. (2015). Ultrasound assisted adsorption of malachite green dye onto ZnS:Cu-NP-AC: Equilibrium isotherms and kinetic studies - Response surface optimization. Separation and Purification Technology, 156, 780–788. https://doi.org/10.1016/j.seppur.2015.11.001
64. Chen, H., Liu, T., Meng, Y., Cheng, Y., Lu, J., & Wang, H. (2020). Novel graphene oxide/aminated lignin aerogels for enhanced adsorption of malachite green in wastewater. Colloids and Surfaces A: Physicochemical and Engineering Aspects, 603, 125281. https://doi.org/10.1016/j.colsurfa.2020.125281
65. Tanyol, M. (2017). Rapid malachite green removal from aqueous solution by natural zeolite: Process optimization by response surface methodology. Desalination and Water Treatment, 65, 294–303. https://doi.org/10.5004/dwt.2017.20185
66. Hameed, B. H., & El-Khaiary, M. I. (2008). Malachite green adsorption by rattan sawdust: Isotherm, kinetic and mechanism modeling. Journal of Hazardous Materials, 159(2–3), 574–579. https://doi.org/10.1016/j.jhazmat.2008.02.054
67. Sadegh, H., Shahryari-Ghoshekandi, R., Agarwal, S., Tyagi, I., Asif, M., & Gupta, V. K. (2015). Microwave-assisted removal of malachite green by carboxylate functionalized multi-walled carbon nanotubes: Kinetics and equilibrium study. Journal of Molecular Liquids, 206, 151–158. https://doi.org/10.1016/j.molliq.2015.02.007
68. Mohammadi, A., Daemi, H., & Barikani, M. (2014). Fast removal of malachite green dye using novel superparamagnetic sodium alginate-coated Fe_3O_4 nanoparticles. International Journal of Biological Macromolecules, 69, 447–455. https://doi.org/10.1016/j.ijbiomac.2014.05.042
69. Ali, H., & Ismail, A. M. (2021). Developing montmorillonite/PVDF/PEO microporous membranes for removal of malachite green: adsorption, isotherms, and kinetics. Journal of Polymer Research, 28(11), 429. https://doi.org/10.1007/s10965-021-02789-3
70. Ghaedi, M., Azad, F. N., Dashtian, K., Hajati, S., Goudarzi, A., & Soylak, M. (2016). Central composite design and genetic algorithm applied for the optimization of ultrasonic-assisted removal of malachite green by ZnO Nanorod-loaded activated carbon. Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy, 167, 157–164. https://doi.org/10.1016/j.saa.2016.05.025
71. Lagergren, S. K. (1898). About the Theory of So-called Adsorption of Soluble Substances. Sven. Vetenskapsakad. Handingarl, 24, 1–39.
72. Ho, Y. S., & McKay, G. (2004). Sorption of copper(II) from aqueous solution by Peat. Water, Air, and Soil Pollution, 158(1), 77–97. https://doi.org/10.1023/B:WATE.0000044830.63767.a3
73. Weber, W. J., & Morris, J. C. (1963). Closure to "Kinetics of Adsorption on Carbon from Solution". Journal of the Sanitary Engineering Division, 89(6), 53–55. https://doi.org/10.1061/jsedai.0000467

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Wind speed prediction using LSTM and ARIMA time series analysis models: A case study of Gelibolu

Adem Demirtop [*1], **Onur Sevli** [1]

[1] *Burdur Mehmet Akif Ersoy University, Department of Computer Engineering, Türkiye, ademdemirtop1987@gmail.com, onursevli@mehmetakif.edu.tr*

Cite this study: Demirtop, A., & Sevli, O. (2024). Wind speed prediction using LSTM and ARIMA time series analysis models: A case study of Gelibolu. Turkish Journal of Engineering, 8 (3), 524-536

https://doi.org/10.31127/tuje.1431629

Keywords
Wind speed prediction
Wind energy
LSTM
ARIMA
Deep learning

Research Article

Received: 04.02.2024
Revised: 11.03.2024
Accepted: 12.03.2024
Published: 08.07.2024

Abstract

Wind energy stands out as a prominent renewable energy source, characterized by its high efficiency, feasibility, and wide applicability. Nonetheless, the integration of wind energy into the electrical system encounters significant obstacles due to the unpredictability and variability of wind speed. Accurate wind speed prediction is essential for estimating the short-, medium-, and long-term power output of wind turbines. Various methodologies and models exist for wind speed time series prediction. This research paper proposes a combination of two approaches to enhance forecasting accuracy: deep learning, particularly Long Short-Term Memory (LSTM), and the Autoregressive Integrated Moving Average (ARIMA) model. LSTM, by retaining patterns over longer periods, improves prediction rates. Meanwhile, the ARIMA model enhances the likelihood of staying within predefined boundaries. The study utilizes daily average wind speed data from the Gelibolu district of Çanakkale province spanning 2014 to 2021. Evaluation using the root mean square error (RMSE) shows the superior forecast accuracy of the LSTM model compared to ARIMA. The LSTM model achieved an RMSE of 6.3% and a mean absolute error of 16.67%. These results indicate the potential utility of the proposed approach in wind speed forecasting, offering performance comparable to or exceeding other studies in the literature.

1. Introduction

In our modern era, energy consumption spans various sectors, predominantly reliant on non-renewable resources, which are depleting steadily [1]. Traditional methods for energy generation predominantly rely on non-renewable resources, exacerbating environmental concerns and energy security issues [2]. The effectiveness of wind energy systems is directly contingent upon wind speed, making accurate wind speed prediction imperative for efficient wind power generation [3]. While statistical methods have traditionally been employed for wind speed estimation, their adequacy may be compromised due to the inherently chaotic nature of wind patterns [4]. Consequently, artificial intelligence algorithms have emerged as viable alternatives for wind speed prediction [5]. Artificial intelligence (AI) technologies offer promising solutions [6] to address these challenges by optimizing energy generation and consumption, particularly in the context of renewable energy sources such as wind power [7]. Numerous studies in the literature highlight the potential of AI algorithms in optimizing wind energy systems, improving efficiency, and overcoming the limitations of traditional statistical methods.

Akbulut and Kemal [8] conducted a study on the effectiveness of deep learning and machine learning models in financial market forecasting. Their research revealed that the Long Short-Term Memory (LSTM) model outperforms the Instance-Based Learning k-Nearest Neighbors method in terms of error rate. The investigation encompassed an analysis of the correlation between commodity and exchange rates as well as stock market indices of developing countries. The findings suggest that the LSTM model exhibits efficiency as a predictive tool. Consequently, it is expected that this model could offer valuable assistance to investors in anticipating market trends.

Eşsiz [9] delves into short-term power prediction by utilizing daily wind data from the Belen region. The study employs analyses employing a radial-basis regressor method and the harmony search algorithm. Results indicate that predictions generated with the harmony

search algorithm exhibit fewer features and errors, leading to a notable 7% enhancement in RMSE. These findings highlight the efficacy of the harmony search algorithm in wind power prediction.

Balcı et al. [10] introduce a hybrid technique for estimating hourly wind speed data collected above 50 m in Balıkesir city. This method combines multilayer perceptions (MLP) with complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and ensemble empirical mode decomposition (EEMD).

According to Balti et al. [11], in their quest to understand the causes of drought, the study explored the predictive abilities of three different methods—ARIMA, Prophet, and LSTM—using meteorological variables such as the standardized precipitation evapotranspiration index (SPEI). The study's results indicated that although the ARIMA model performed better than the Prophet model, the LSTM model outperformed both. This emphasizes the reliability and accuracy of the LSTM model in identifying the underlying causes of drought.

Baykal et al. [12] undertook a study to forecast meteorological drought in Isparta province over the next decade employing the LSTM method. Their research unveiled a parallel declining trend in precipitation and DEM series, with severe droughts noted between 1982 and 2011. The study proposes that extending the time interval of precipitation data generated by the LSTM method could aid in long-term water resource planning and the execution of essential measures.

Canıtez and Savaş [13] conducted a study comparing feature-based LSTM and ARIMA methods for predicting the market value of cryptocurrency, focusing specifically on Bitcoin. The study utilized 10,309 real-time data points. Both ARIMA and LSTM techniques were employed to generate predictions. The results revealed that the MAPE values of both approaches fell within the "very good" range. However, upon comparison, it was observed that the ARIMA method produced superior outcomes, suggesting that the behavior of Bitcoin prices is more accurately captured by the ARIMA approach.

Dave et al. [14] conducted a study focusing on forecasting Indonesia exports using a hybrid ARIMA-LSTM model. Within this study, they assessed a hybrid model (LSTM-ARIMA) to establish an integrated machine learning model for wind speed prediction. The findings of their research revealed that the hybrid model exhibited the lowest error metrics compared to all other models examined. This underscores the hybrid model's superior accuracy and reliability in wind speed prediction.

Demirtop and Işık [15] introduced a novel methodology utilizing artificial neural networks (ANNs) to enhance wind energy efficiency. They utilized a dataset comprising temperature, pressure, humidity, and wind speed data collected from Bozcaada, Çanakkale. ANN models were trained using both WEKA and MATLAB platforms. Among the methods evaluated, the Levenberg-Marquardt algorithm demonstrated superior accuracy, with MATLAB exhibiting better performance than WEKA. The authors advocate for further research utilizing larger datasets and diverse ANN architectures to validate the applicability of ANNs in forecasting wind energy efficiency.

Devi et al. [16] utilized the extended Long Short-Term Memory network-enhanced Forgetting Gate network (LSTM-EFG) model for wind energy prediction. Their study involved training the model on sub-series data obtained through ensemble empirical mode decomposition (EEMD) and refining it with the cuckoo search optimization technique (CSO). The research demonstrated enhanced prediction accuracy, surpassing traditional forecasting approaches.

Elsaraiti and Merabet [17] conducted a comparative analysis of Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), Autoregressive Integrated Moving Average (ARIMA), and Long Short-Term Memory (LSTM) - a variant of RNN - to determine the optimal method for time series forecasting. The findings revealed that the LSTM approach outperformed the ARIMA method in forecast accuracy. These results underscore the superior predictive capabilities of the LSTM method, which yields more precise forecasts with fewer errors.

Erden [18] conducted a comparison between ARIMA and deep learning models for forecasting Borsa Istanbul's EREGL stock, taking into account the nonlinear and complex nature of financial time series data. Through data preprocessing, feature extraction, and analysis of different time periods, prediction performance was enhanced. The Recurrent Neural Network (RNN) algorithm exhibited an impressive accuracy rate of 93% in this context.

Ji et al. [19], a model named ARIMA-CNN-LSTM was developed to predict the price of carbon futures. In this model, long-term relationships in the data are captured by LSTM, hierarchical data structures are captured by CNN, and linear characteristics are captured by ARIMA. The findings of the study demonstrate that the ARIMA-CNN-LSTM model outperforms the benchmark model in terms of prediction accuracy. This outcome illustrates that the ARIMA-CNN-LSTM model provides a more reliable and accurate method for predicting the price of carbon futures.

Kamber et al. [20] conducted an analysis of hourly electricity data within an LSTM-based artificial neural network (ANN) framework using Spain's electricity data for the years 2015-2016, and compared it with ARIMA results. Through this comparison, both forecasting models exhibited similar performance. These findings indicate that both LSTM and ARIMA models are equally effective for hourly electricity forecasting.

Liu et al. [21] proposed a Seasonal Auto Regression Integrated Moving Average (SARIMA) model for forecasting hourly observed wind speeds in the onshore/offshore area of Scotland. The model was trained using three wind speed time series obtained from various heights of a coastal measuring mast designed for servicing an offshore wind turbine. Test results indicated that, compared to the GRU and LSTM models, the SARIMA model produced more reliable and accurate predictions. This outcome underscores the SARIMA model's utility as a valuable tool for forecasting offshore wind speed time series.

Othman [22] utilized Bayesian optimization to optimize various parameters, including the number of biLSTM layers and units. Among the models evaluated, the one utilizing SGDM exhibited the highest

performance, as determined by Spearman's Rank Correlation (r). The study employed three distinct algorithms—SGDM, ADAM, and RMSprop—to train deep CNN-biLSTM models. Ultimately, the study highlighted the effectiveness of the CNN model with LSTM in replicating time series data.

Sevinç and Buket [23] utilized temperature data from the Solhan district of Bingöl province to evaluate the forecasting capabilities of LSTM and ARIMA models. The findings indicated that the LSTM model achieved a mean error (MAE) of 0.73 °C, while the ARIMA model's MAE was 0.76 °C. These results suggest that both models produce forecasts closely aligned with the actual values, indicating their comparable performance. The comparison underscores the excellent accuracy of both models, as evidenced by the achieved MAE values.

Shao et al. [24] focused on the development of an LSTM neural network model and the optimization of hyperparameters for wind speed prediction, which is a significant topic of interest. The study's findings indicated that the FWA-optimized LSTM model outperformed alternative regression techniques commonly employed for wind speed prediction.

Wang and Wang [25] proposed a mixed model incorporating Empirical Mode Decomposition (EMD), Long Short-Term Memory (LSTM), and Autoregressive Integrated Moving Average (ARIMA) for forecasting monthly precipitation. The study demonstrated the superior forecasting performance of this model compared to various combination models and single models, including EMD-LSTM, EEMD-LSTM, EEMD-ARIMA, among others. Additionally, the model exhibited a high level of confidence in the predicted precipitation results.

Zhang et al. [26] aimed to identify the most effective model for predicting the prevalence of hand, foot, and mouth disease (HFMD) in Ningbo. The study evaluated two forecasting models, ARIMA and LSTM. According to the findings, the multivariate LSTM model provided the best fit for the daily incidence of HFMD in Ningbo among the four models tested. This multivariate LSTM model incorporates factors such as precipitation, humidity, and air temperature to enhance the accuracy of HFMD incidence prediction.

Zhang et al. [27] conducted a study yielding significant results demonstrating the efficacy of LSTM-based approaches in time series forecasting. The findings indicate that LSTM-based techniques outperform hybrid and CNN-based techniques, attributed to their superior ability to capture long-term dependencies in time series data. Notably, the study's second case study achieves longer-term forecasts, with deep learning-based techniques outperforming the ARIMA method and exhibiting similar performance among themselves.

Zhao et al. [28] employed ARIMA, LSTM, and various machine learning models for demand forecasting using sales data from a company in the retail sector. The CRISP-DM methodology was adopted, encompassing data preprocessing, time series analysis, and model evaluation stages. According to the findings, machine learning models exhibited superior performance compared to ARIMA and LSTM models. Specifically, within the ARIMA models, the SARIMAX model outperformed both ARIMA and SARIMA models, attributed to its utilization of independent variables.

This study contributes to the literature by comparing the effectiveness of deep learning models such as LSTM with statistical analysis-based ARIMA models in predicting future wind speeds using current meteorological data from a specific region in Türkiye. Particularly, it sheds light on the performance of ARIMA alongside LSTM, emphasizing the role of both methodologies in wind speed prediction. The findings underscore the potential of deep learning techniques, represented by LSTM, in outperforming traditional statistical approaches like ARIMA in wind speed forecasting. Moreover, by evaluating various performance metrics of different model approaches, this research aims to provide valuable insights into their prediction accuracy and implications for practical applications. Additionally, the utilization of ARIMA in conjunction with LSTM offers a comprehensive understanding of the strengths and limitations of both approaches, contributing to the advancement of knowledge in the field of renewable energy forecasting.

2. Method

2.1. Dataset

Our research area is located in the Gelibolu district, situated between the Dardanelles and Saroz Gulf in the Marmara Region of northwestern Türkiye. This area exhibits seasonal transition characteristics, reflecting the typical features of the Mediterranean climate. Due to its northern latitude, winter temperatures are relatively lower, with August experiencing a maximum temperature of +35.8 °C and February reaching a minimum of -4.2 °C. Throughout the year, the average temperature and humidity are 14.7°C and 72.6%, respectively. Additionally, the region is characterized by consistent windy conditions throughout the year, distinguishing it from other locations [29].

In this study, we employed a dataset encompassing daily average wind speed records spanning from the years 2014 to 2021 for the Gelibolu district of Çanakkale province [30]. The dataset comprises 2908 data points, each representing a single day, with the wind speed measurements ranging from a minimum of 2.88 km/h to a maximum of 60.48 km/h. The graph presented in Figure 1 illustrates the distribution of these data points, with wind speed (measured in km/h) plotted on the vertical axis and the total number of days (data points) on the horizontal axis.

The average wind speed across the dataset was calculated to be 15.51 km/h. This comprehensive dataset allows for a detailed analysis of the wind speed patterns and variations observed in the Gelibolu district over the specified time period. Such insights are crucial for understanding the local wind climate dynamics, which in turn can inform various applications, including renewable energy resource assessment, environmental monitoring, and infrastructure planning.

Analyzing the graph for the region by years in Figure 2, it is evident that the average wind speed exhibits variations over the years. Specifically, the data indicates

that the average wind speed was at its lowest in 2014 and reached its peak in 2018. This observation suggests temporal fluctuations in wind speed patterns within the Gelibolu district over the specified time period. Such insights into the interannual variability of wind speeds are valuable for understanding long-term trends and can inform decision-making processes in various sectors, including energy, agriculture, and infrastructure planning [5].

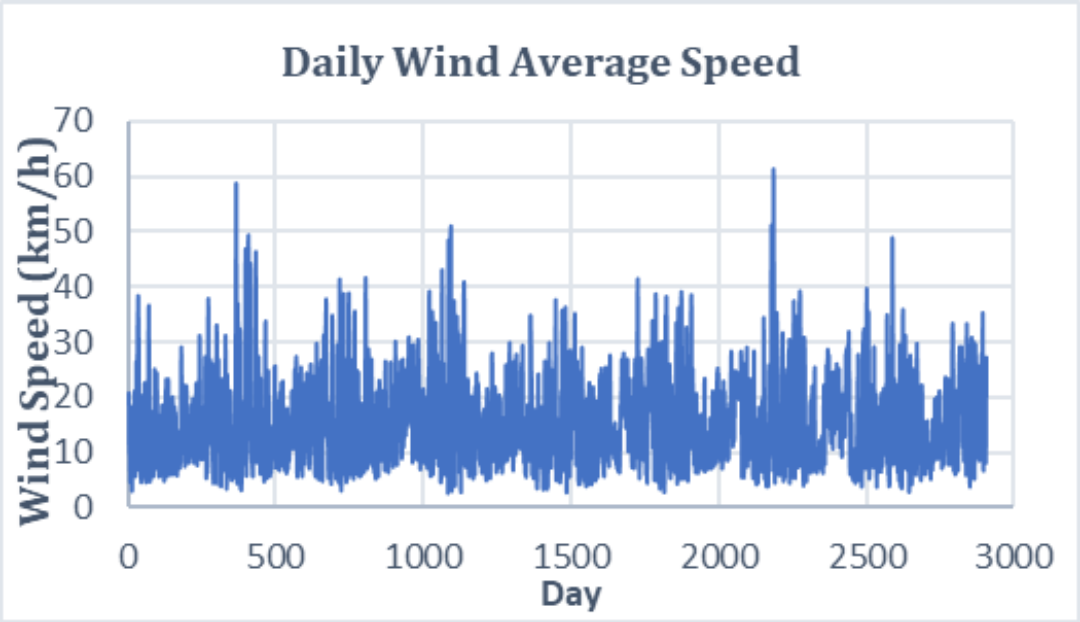


Figure 1. Daily average wind speed of the region (km/h).

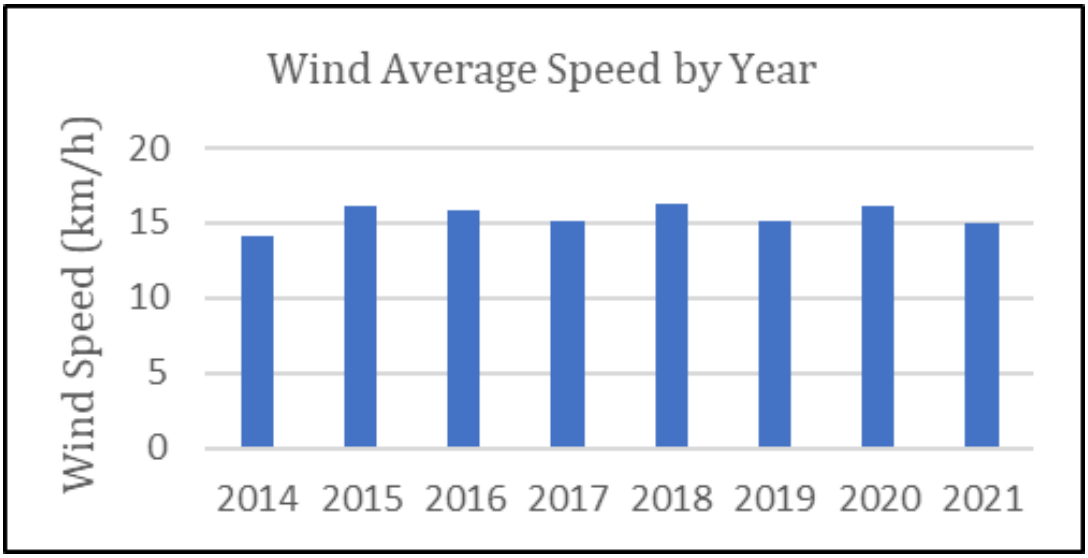


Figure 2. Wind mean speed according to years in the region.

2.2. Analysis models used

2.2.1. Long-short term memory (LSTM)

Recurrent neural networks, including long short-term memory models (LSTMs), are well-suited for processing temporal data due to their ability to capture long-term dependencies [31]. LSTMs are particularly advantageous in applications such as speech recognition and natural language processing, where recognizing patterns over extended sequences is essential [32]. Key to the effectiveness of LSTMs is their utilization of cell states to store information across time [33]. The cell state serves as a memory unit within the LSTM, regulating the flow of information into and out of the network at each time step through controlled updates [34]. This mechanism enables LSTMs to effectively retain and utilize contextual information over extended sequences, facilitating accurate predictions and analysis of temporal data.

Recurrent neural networks, such as long short-term memory models (LSTMs), are adept at handling temporal data due to their capability to discern long-term patterns [35]. LSTMs are particularly favored in applications like speech recognition and natural language processing owing to their ability to recognize extended sequences effectively [36]. LSTMs utilize cell states to retain information across time, regulating the flow of data throughout the network. This mechanism enables LSTMs to maintain and utilize contextual information over

prolonged sequences, thereby facilitating accurate predictions and analysis of temporal data [37].

LSTM networks address the vanishing/exploding gradient problem by employing gates and a well-defined memory cell. These gates regulate the flow of information into and out of the cell. The input gate determines the amount of data from the previous layer that is stored in the cell. The output gate controls how much information about the cell's state is passed to the next layer. The forget gate manages the retention of data stored in the cell. This mechanism enables LSTM networks to effectively learn over multiple time steps and capture long-term dependencies (Figure 3).

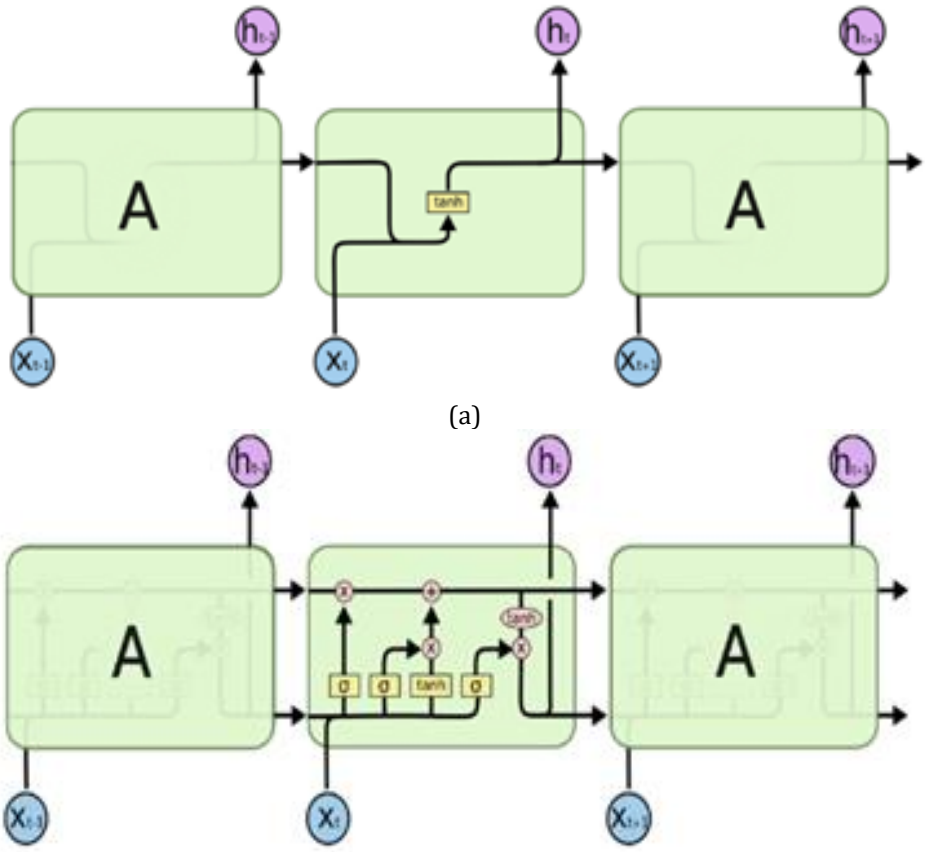


Figure 3. LSTM and RNN recurrent modules. (a) an RNN's recurrent module has one component. This is the neural network stack. (b) An LSTM's recurrent module has four interacting layers [38].

The design depicts the LSTM with three layers stacked, as illustrated in Figure 4. LSTM is a type of artificial neural network utilized for capturing long and short-term dependencies. The three-layered structure indicates a deep architecture, enabling the network to learn more complex relationships.

Figure 4. Long short-term memory neural network (LSTM) design with three layers stacked.

In Table 1, the LSTM model can be trained and predicted in three steps as outlined below:

1. Data Preprocessing: The data undergoes preprocessing where it is rescaled and normalized to the range of 0 to 1. This step is crucial as LSTM models are sensitive to the scale of the input data.

2. Model Parameter Determination: The univariate and multivariate LSTM time steps are adjusted to predict the wind speed for the following day using data from the preceding 7/30/60/180 days. Each LSTM layer in the three-layer stacked LSTM structure consists of a hidden layer that is tailored for the LSTM model. Alternative optimization functions such as Stochastic Gradient Descent (SGD), Adaptive Moment Estimation (Adam), and Root Mean Square Prop (RMSProp) are considered.

3. Model Training: Training is conducted over 200, 250, 500, and 1000 epochs for each learning procedure. The least Root Mean Square Error (RMSE) of each batch of 125 epochs is utilized to determine the best-suited model. The starting learning rate is set at 0.005.

Table 1. Parameters used in the LSTM structure.

Time Stages	7/30/60/180 days
Neurons	4/8/16/32/64/72/128/256
Optimization Functions	Adam/SGD/RMSProp
Number of Data Iterations (Epoch)	200/250/500/1000
First Learning Rate	0.005
Study the Rate Schedule:	fragmented
Learn Rate Decline Time:	125
Factor for Learning Rate Drop:	0.2

2.2.2. Autoregressive integrated moving average (ARIMA)

Autoregressive Integrated Moving Average, abbreviated as ARIMA, is a statistical model commonly employed for time series data forecasting. ARIMA models are based on the premise that data can be represented as a linear combination of historical values, errors, and moving averages. These models are widely utilized for predicting variables such as wind speeds, solar radiation levels, stock prices, interest rates, and inflation. ARIMA modeling enables analysts to make accurate forecasts and projections based on historical patterns and trends observed in the data [39].

ARIMA models are employed for predicting time series data using a methodology known as the Box-Jenkins method [40], which consists of four stages:

1. Identification: In this stage, the order of the ARIMA model is determined. The order is defined by the number of autoregressive terms (p), the number of differences to be taken (d), and the number of moving average terms (q) [40].
2. Estimation: Model parameters are estimated using the method of maximum likelihood estimation [40].
3. Diagnostic check: A diagnostic check is performed on the ARIMA model to ensure that the residuals exhibit white noise behavior and that the model adequately captures the data [40].
4. Forecasting: Future values of the time series are predicted using the estimated parameters of the ARIMA model [40].

ARIMA models are particularly effective for predicting time series data and are commonly utilized in short-term forecasting tasks [40]. The notation ARIMA(p,d,q) is used to denote an ARIMA model, where:

- p represents the autoregressive degree (AR),
- d represents the degree of differencing (I),
- q represents the moving average (MA).

In ARIMA models, the autoregressive parameters represent the lags of the differenced time series, while the moving average terms account for the forecast error delays [41]. If the time series is non-stationary (or seasonal), differencing is applied to make it stationary. The resulting integrated series is then modeled using ARIMA (p, d, q). In this notation, p, d, and q denote the quantities of autoregressive terms, lagged forecast errors, and non-seasonal differences, respectively. Equation 1 provides the general ARIMA formula.

$$y_t = c + \sum_{i=1}^{p} \phi_m y_{t-i} + \sum_{j=0}^{q} \theta_n e_{t-j} \quad (1)$$

In Equation 1 ϕ_m autoregression coefficient, ϕ_m y_(t-i) autoregression lags (at degree p), θ_n moving average parameter, e_(t-q) moving average errors (of order q) and c is the constant term [40].

The linear connection between the time series' lag values and the error term determines the degree of autoregression [40]. The AR(p) model here is as in Equation 2.

$$y_i = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + c \quad (2)$$

The moving average degree relies on a weighted moving average of error values as outlined in the literature [40]. Equation 3 provides a general representation of the MA(q) model.

$$y_i = \theta_0 e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \quad (3)$$

The fusion of autoregressive (AR) and moving average (MA) methodologies constitutes the basis of the ARMA (p,q) approach [40]. Equation 4 presents the autoregressive moving average (ARMA) model.

$$y_i = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \theta_0 e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \quad (4)$$

3. Experimental study and findings

3.1. Data analysis

The training dataset comprised data spanning from January to December 2016, while the test dataset was selected from December 2020 onwards [30]. To forecast daily wind speed data, ARIMA and LSTM models were trained using the training dataset, both with and without incorporating external meteorological factors [16]. Three index metrics were identified to evaluate the models' performance. The primary performance metric for comparing predicted values with actual values is the RMSE [42]. It is calculated using the Equation 5.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_i - X_i')^2}{n}} \quad (5)$$

The mean absolute error (MAE) [42], which is the second performance metric, is defined as Equation 6.

$$MAE = \frac{\sum_{i=1}^{n} | X_i - X_i' |}{n} \quad (6)$$

The third performance metric is relative overall conformance, quantified by the mean absolute percentage error (MAPE) [42]. The formula for calculating this metric is presented in Equation 7.

$$MAPE = \frac{\sum_{i=1}^{n} \frac{|X_i - X_i'|}{X_i} \times 100}{n} \quad (7)$$

The RMSE, MAE, and MAPE measures are commonly employed to assess the performance of models in time series forecasting, including both ARIMA and LSTM models [16]. RMSE is particularly suitable for evaluating

the accuracy of the forecast, as it provides a measure of the average magnitude of the errors between predicted and actual values [43]. On the other hand, MAE is preferred for evaluating the consistency of the forecast, as it calculates the average absolute errors between predicted and actual values. Additionally, MAPE can be utilized to evaluate both accuracy and consistency, as it expresses the average percentage difference between predicted and actual values, providing insights into the relative performance of the models across different time series [42].

3.2. Forecast verification

The real-world environment is characterized by instability and frequent sudden changes, which can significantly impact time series data such as wind speed. A forecasting model for wind speed must be capable of adapting rapidly to these dynamic changes in order to remain effective [44]. Therefore, it is imperative for a wind speed forecasting model to be flexible and responsive to sudden fluctuations in environmental conditions. Only by incorporating adaptability into the forecasting model can it effectively account for the unpredictable nature of the environment and provide accurate forecasts even in the face of rapid changes [3].

In this study, a rolling prediction scenario, also known as forward walking model validation, is utilized. In this scenario, the test dataset's time steps are advanced incrementally. At each step, the observed value from the test set is used to forecast the subsequent time step using the model. This process simulates real-world conditions, where fresh daily wind speed data is continuously collected and utilized to predict wind speed for the following day. By employing this approach, the model's performance can be evaluated in a dynamic and evolving context, mirroring the conditions under which it would be utilized in practice [45].

The two parameters commonly used to forecast wind speed are the MAE and the RMSE, which assess the accuracy and precision of the models, respectively. The MAE measures the average absolute difference between the expected and actual values. On the other hand, the RMSE computes the square root of the mean square difference between the actual and predicted values, providing a measure of the overall deviation between them [42]. These metrics are essential for evaluating the performance of wind speed forecasting models and assessing their effectiveness in providing accurate predictions.

In this study, both an LSTM model and several sequential ARIMA models are employed for wind speed prediction. Through the utilization of a rolling forecast scenario, the LSTM model is demonstrated to outperform the other models in terms of accuracy and flexibility. This superiority can be attributed to the LSTM model's ability to retain and leverage past data over prolonged periods, allowing it to make more informed forecasts for future time steps.

RMSE represents the square root of the mean square of all errors, providing a comprehensive measure of error. Widely acknowledged as a superior error metric for numerical predictions, RMSE is commonly employed in regression tasks across both statistical and machine learning domains [42].

The RMSE quantifies the difference between expected and actual values, with a larger RMSE indicating greater deviations. A notable property of the RMSE is that squaring the errors assigns significantly more weight to larger errors. Therefore, a mistake with a value of 10 is considered 100 times more impactful than a mistake with a value of 1 in the context of RMSE calculation. This weighting mechanism emphasizes the significance of larger errors in the overall evaluation of predictive accuracy. While it depends on size, RMSE is a useful metric for assessing accuracy. As a result, it can only be applied to compare prediction errors within a variable, not across variables, between models or model configurations [42].

The MAE measures the average discrepancy between expected and actual values, serving as a metric for error assessment. It represents the mean absolute difference between these values. The MAE provides insight into the average magnitude of errors expected from the forecast. Unlike RMSE, the inaccuracy scales linearly with MAE, meaning that each error contributes equally to the overall measure. Consequently, a deviation of 10 is ten times more significant than a deviation of 1 when considering MAE calculation [42].

4. Results

4.1. Prediction model ARIMA

The research period aimed at selecting the ARIMA model structure encompasses data gathered by the General Directorate of Meteorology spanning from 2014 to 2021 [30]. Out of a total of 2500 daily average time series wind speed data points, the initial 2400 points were employed in constructing the models. Subsequently, prediction and performance evaluation were conducted using the remaining 100 data points. The determination of the P and q orders was facilitated through an analysis of the autocorrelation function (ACF) and partial autocorrelation function (PACF) graphs.

In this study, the conventional ARIMA modeling approach is applied to forecast wind speed, involving the derivation of an appropriate model structure and parameters based on the collected data.

Initially, the stationarity of the time series data is evaluated utilizing the autocorrelation function (ACF) and running order charts. These visual aids assist in scrutinizing trends within the data and validating the assumption of constant variance. Sequential differencing of the data series is conducted based on the observed characteristics in the ACF and PACF plots until stationarity is confirmed [46].

Subsequently, the autoregressive (AR) and moving average (MA) terms are determined using the ACF and PACF plots. These graphical representations aid in the selection of the AR and MA variables for the model [47].

Initially, the stationarity of the time series data is evaluated utilizing the autocorrelation function (ACF) charts. The ACF chart in Figure 5 displays values ranging from a maximum of 1.0 to a minimum of -0.2 for lag values between 0 and 4, indicating a moderate level of

autocorrelation. Sequential differencing of the data series is conducted based on the observed characteristics in the ACF plots until stationarity is confirmed.

Subsequently, the autoregressive (AR) and moving average (MA) terms are determined using the ACF charts. These graphical representations aid in the selection of the AR and MA variables for the model. This process ensures the accurate modeling and prediction of wind speed [47].

The series depicted in Figure 6 exhibits clear and regularly recurring cyclical activity, suggesting potential regularity within the underlying processes of interest. Understanding these processes could be facilitated by examining the speed or frequency of the oscillations observed in the main series.

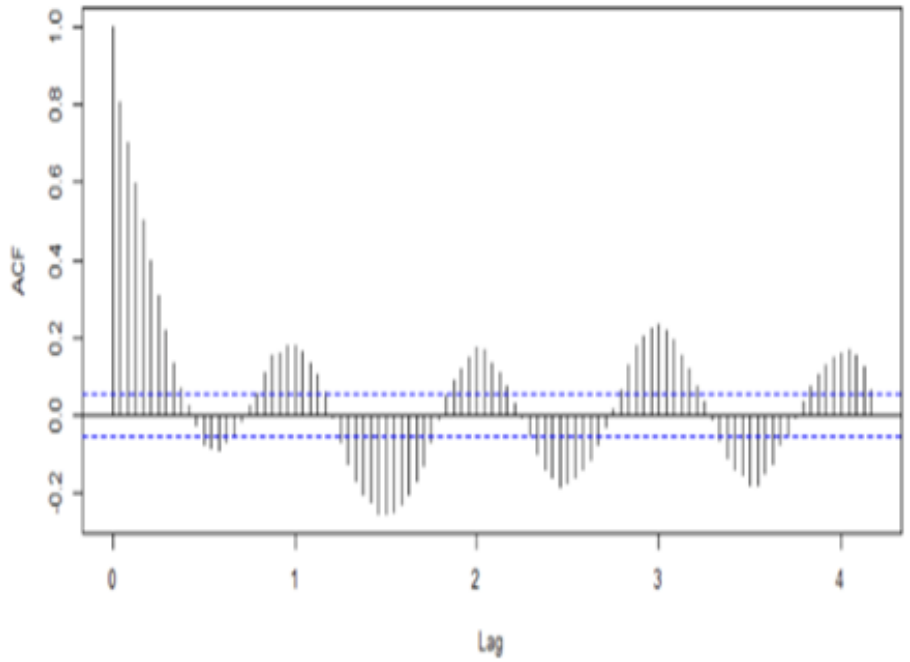


Figure 5. Functions of autocorrelation for observed wind speed data.

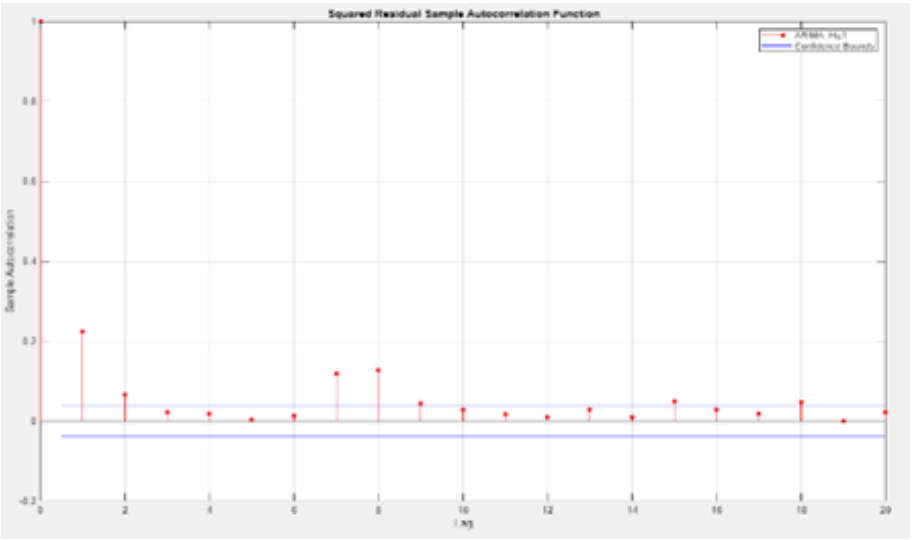


Figure 6. Functions of partial autocorrelation for data on measured wind speed.

Two primary types of variations are evident in the series. Firstly, there are distinct sinusoidal fluctuations characterized by dips and peaks. Secondly, there are periodically repeated fluctuations occurring at a slower frequency.

Non-stationary data are typically unpredictable and challenging to model. However, the periodic behavior observed in this series suggests the possibility of achieving stationarity. Techniques such as differencing can be employed to eliminate the effects of periodicity from the data [48].

Time series exhibiting frequent cycles and repetitive activity are more straightforward to comprehend and model. This implies that the underlying processes may possess regularity, distinguishable by the frequency or speed of oscillations defining the behavior of the parent series.

Non-stationary time series may yield inaccurate results, potentially indicating a lack of correlation between variables. Transforming non-stationary data into stationary data is crucial for obtaining consistent and reliable findings.

A stationary process consistently reverts to its long-run mean and maintains a constant variance over time. Conversely, a non-stationary process lacks these characteristics, exhibiting variable variance, a non-converging mean, or a lack of return to its long-run mean over time.

The gradual decline in autocorrelation function (ACF) values suggests non-stationarity in the data. Thus, transforming the data into a stable series is necessary to arrest the decline of ACF values [46].

4.2. Prediction model LSTM

Constructing an LSTM regression network involved specifying an LSTM-RNN layer with training options [34]. Multiple initial learning rates were tested to identify the optimal training parameter that yielded the lowest RMSE and loss, specifically at a learning rate of 0.01. Figure 7 illustrates how the initial learning rate affects training time, showing an increase in training time as the learning rates decrease. However, achieving the best outcome may pose challenges with a limited number of iterations. A high learning rate can expedite training but may result in divergence or failure to converge if excessively high. Moreover, substantial weight changes may enhance improvement but could also degrade the loss function. Through experimentation with different initial learning rates, a 24-step prediction was conducted during time step testing, resulting in improved training outcomes and a reduction in function loss to a manageable level. These findings underscore the efficacy of LSTM for large time series datasets and highlight the effectiveness of the specified model in minimizing RMSE.

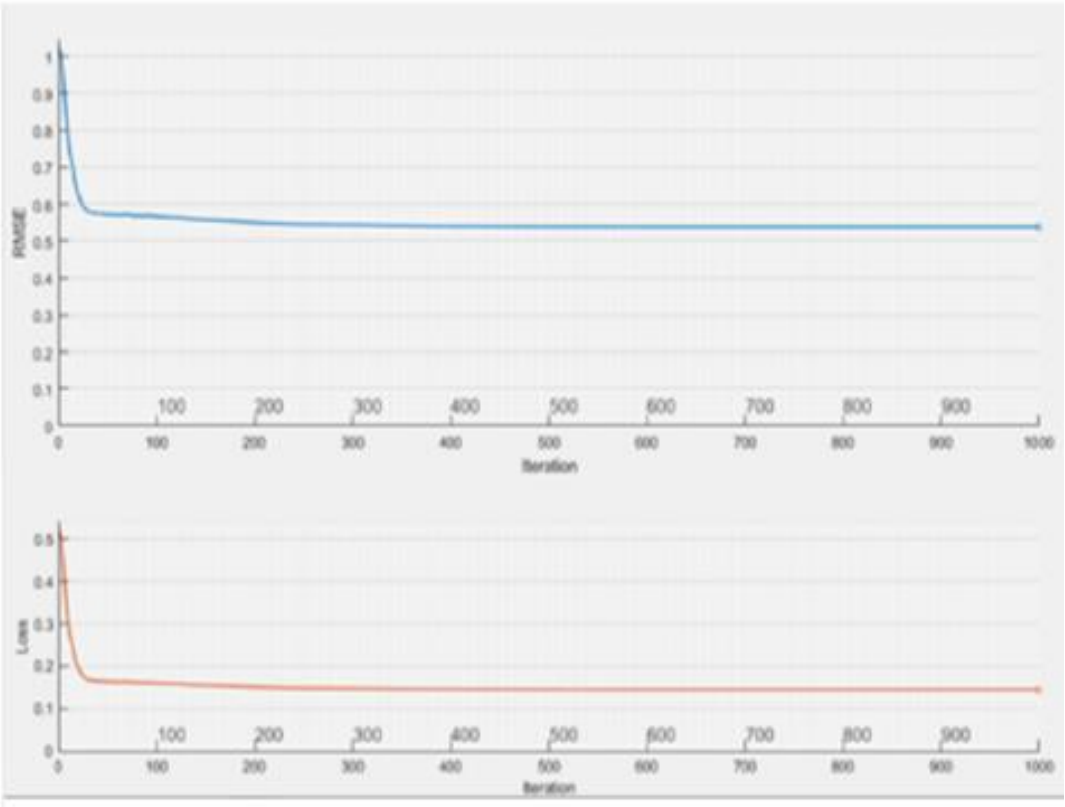


Figure 7. Process of training as learning rate of 0.01 and time step test of 24.

During LSTM-RNN training, the hidden layer receives feedback from the anticipated values of the preceding phase [34]. Throughout the validation process, the model is fine-tuned to fit all of the training data and then updated after each prediction. In this scenario, before generating the subsequent forecast, the model undergoes two additional training cycles. The prediction is then normalized using the previously determined mean and standard deviation, after which the RMSE is computed [35].

The 24-step wind speed forecast results are displayed in Figure 8. It can be observed that there are no noticeable oscillations, and all of the model's training epoch forecast data closely resemble the actual data. Furthermore, the model's total test RMSE score is the lowest, indicating uncommon occurrences of vanishing gradients and gradient bursts for the LSTM algorithm. Moreover, the expected outcomes show no significant deviation, falling within a reasonable range.

Both the ARIMA and LSTM models' outputs are assessed based on two key metrics: MAE and RMSE. These metrics provide insight into the accuracy and precision of the forecasts generated by each model. By comparing the performance of the ARIMA and LSTM models using these metrics, we can determine which model is more effective in predicting wind speed.

Upon analyzing the results presented in Table 2, it becomes evident that the LSTM model outperforms the ARIMA model in terms of efficiency. This suggests that the LSTM model is better able to capture the underlying patterns and dynamics in the wind speed data, resulting in more accurate and reliable forecasts compared to the ARIMA model.

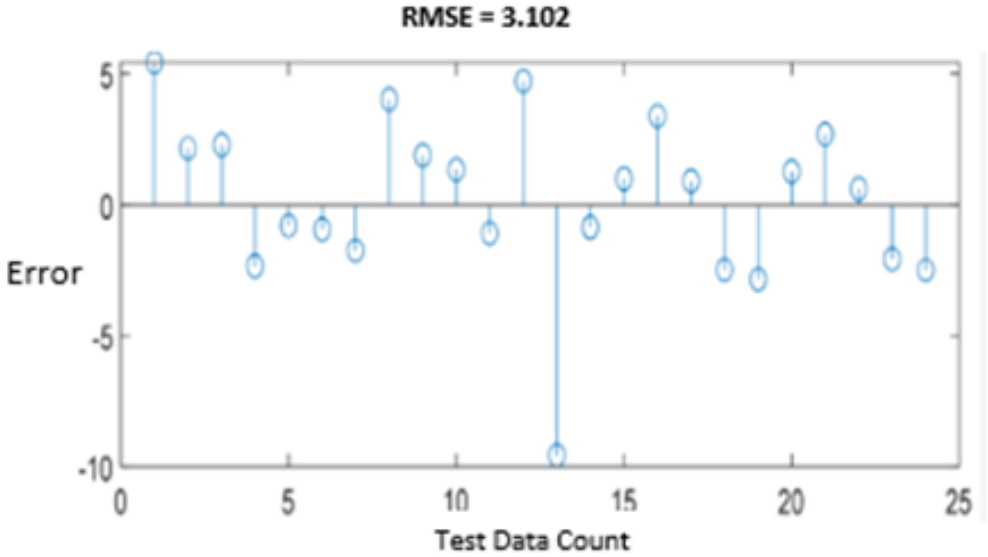


Figure 8. RMSE outcome in 24-step stages.

Table 2. Summary of statistical errors of the test data.

Model	RMSE	MAE
ARIMA	3.312	2.682
LSTM	3.102	2.235

The comparison between the ARIMA and LSTM models reveals important insights into their forecasting performance. Specifically, the MAE and RMSE values offer a quantitative assessment of the accuracy of the forecasts generated by each model.

For the ARIMA model, the MAE value is calculated to be 2.682, while the RMSE value is 3.312. In contrast, the LSTM model achieves a lower MAE value of 2.235 and a lower RMSE value of 3.102. These figures clearly indicate that the LSTM model outperforms the ARIMA model in terms of forecast accuracy.

The RMSE metric measures the degree of deviation between the expected and actual values, while the MAE metric quantifies the average discrepancy between the true and predicted values. By evaluating both metrics, we gain a comprehensive understanding of the forecast accuracy. In this comparison, the LSTM model exhibits a lower RMSE value compared to the ARIMA model, indicating that its predictions are closer to the actual values. Additionally, the LSTM model achieves a lower MAE value than the ARIMA model, further highlighting the improved accuracy of its forecasts.

Figure 9 illustrates the 24-step prediction results of wind speed time series data, presenting the predictions produced by the LSTM deep learning model alongside the observed values.

Upon analysis of the graphs, it becomes apparent that the LSTM model's predictions closely track the observed values. Specifically, when the observed values decrease, the LSTM prediction values also decrease, indicating the model's ability to accurately identify trend changes in time series data. Likewise, when the observed values increase, the LSTM prediction values also increase, demonstrating the model's capability to detect seasonal changes in the data.

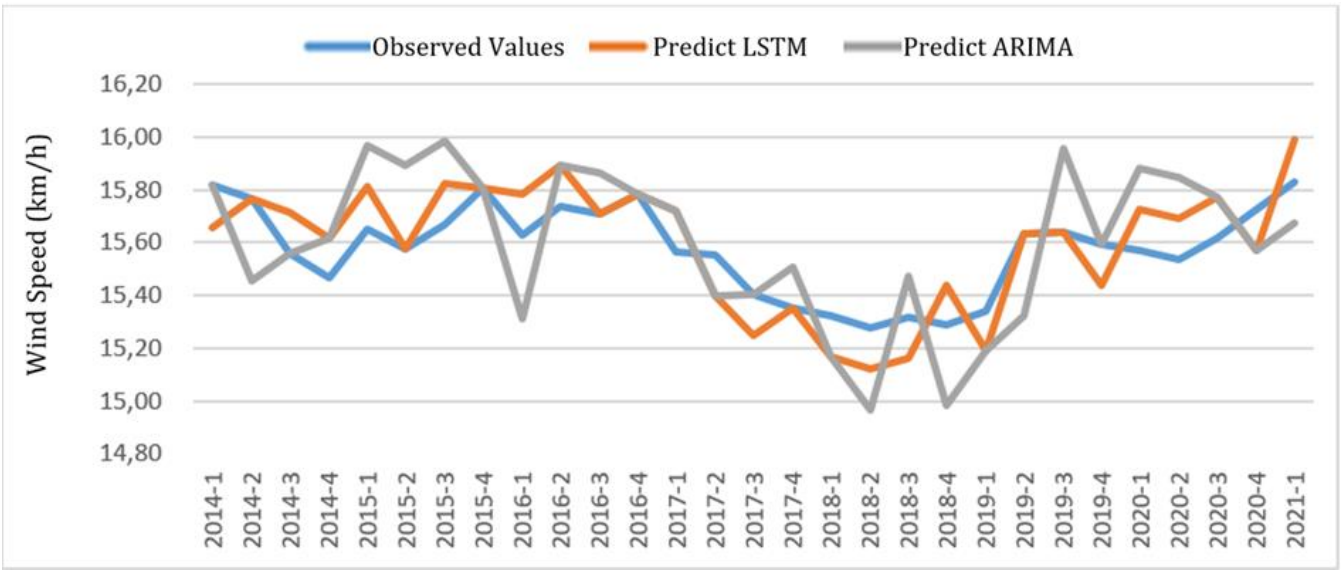


Figure 9. Using ARIMA and LSTM models, compare the three-month average wind speed data.

While the ARIMA model also yields results close to the predicted values, it exhibits discrepancies in trend detection. Notably, instances where the actual values increase are accompanied by decreases in the prediction values from the ARIMA model, despite achieving a lower error rate compared to the LSTM model. This suggests that the ARIMA model may struggle to accurately identify trend changes in time series data.

In conclusion, the graph in Figure 9 underscores the superior effectiveness of the LSTM deep learning model in wind speed prediction. Leveraging its ability to discern trends and seasonal patterns in time series data, the LSTM model produces more precise forecasts compared to the ARIMA model.

5. Discussion

Researchers from diverse fields are increasingly adopting advanced machine learning techniques, particularly deep learning algorithms. It is imperative to comprehensively evaluate the effectiveness and robustness of these modern methods in comparison to traditional approaches [49]. This study focuses on assessing the performance of two models with a specific emphasis on the Gelibolu district of Çanakkale province: a deep learning-based algorithm, the LSTM model, and a classical algorithm, the ARIMA model. A review of the literature suggests that the ARIMA model may yield superior results with limited data, as observed in previous academic research. However, the substantial amount of data utilized in the models developed for this study indicates that LSTM and other deep learning-based algorithms outperform traditional techniques like ARIMA. This underscores the significant potential of deep learning algorithms and approaches, particularly in forecasting complex time series data such as wind speed.

In addition to model performance, it is important to acknowledge the challenges and limitations encountered during the course of this research. Data availability, model selection, parameter tuning, and computational resources posed significant challenges throughout the study. Despite these challenges, the findings of this research provide valuable insights into the effectiveness of deep learning techniques in wind speed prediction. Future research endeavors should explore the application of deep learning techniques to other forecasting problems within the domain of wind speed prediction. Additionally, a more detailed examination of the performance of deep learning techniques across different datasets could provide valuable insights into this research area.

6. Conclusion

The results of this study demonstrate that the LSTM model outperforms the ARIMA model in predicting average wind speed in the Gelibolu district of Çanakkale province. With an average RMSE of 6.3% and MAE of 16.67%, the LSTM-based algorithm exhibits greater accuracy compared to ARIMA, highlighting its effectiveness in forecasting time series data.

The findings of this research emphasize the advantages of utilizing deep learning algorithms, particularly LSTM, in wind speed prediction. Despite previous studies suggesting superior performance of ARIMA with limited data, our results indicate that the abundance of data used in this study favors LSTM and other deep learning-based techniques.

Future research directions may involve exploring the application of deep learning methodologies to additional forecasting challenges within the realm of wind speed prediction. Further investigation is warranted to ascertain the extent of improvement achievable through deep learning approaches across diverse datasets with varying features.

Acknowledgement

The authors express their gratitude to the Türkiye State Meteorological Service for providing the data. Additionally, the authors extend they're thanks to the reviewers for their constructive criticisms, which have significantly enhanced this manuscript.

Author contributions

Adem Demirtop: Data analysis, drafting the manuscript.
Onur Sevli: Defining the methodology, evaluating the results and editing the draft.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Torunoğlu Gedik, Ö. (2015). Türkiye'de yenilenebilir enerji kaynakları ve çevresel etkileri. [Doctoral dissertation, Istanbul Technical University].
2. Makarieva, A. M., Gorshkov, V. G., & Li, B. L. (2008). Energy budget of the biosphere and civilization: Rethinking environmental security of global renewable and non-renewable resources. Ecological Complexity, 5(4), 281-288. https://doi.org/10.1016/j.ecocom.2008.05.005
3. Ssekulima, E. B., Anwar, M. B., Al Hinai, A., & El Moursi, M. S. (2016). Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: a review. IET Renewable Power Generation, 10(7), 885-989. https://doi.org/10.1049/iet-rpg.2015.0477
4. Bokde, N., Feijóo, A., Villanueva, D., & Kulat, K. (2019). A review on hybrid empirical mode decomposition models for wind speed and wind power prediction. Energies, 12(2), 254. https://doi.org/10.3390/en12020254
5. Wohland, J., Omrani, N. E., Keenlyside, N., & Witthaut, D. (2019). Significant multidecadal variability in German wind energy generation. Wind Energy Science, 4(3), 515-526. https://doi.org/10.5194/wes-4-515-2019
6. Sinap, V. (2023). Makine öğrenmesi teknikleri ile counter-strike: Global offensive raunt sonuçlarının tahminlenmesi. Journal of Intelligent Systems: Theory and Applications, 6(2), 119-129.

https://doi.org/10.38016/jista.1235031
7. Çakır, F. (2020). Demiryolu yolcu taşıma talebinin yapay sinir ağları ile tahmini. [Master's thesis, Aksaray University].
8. Akbulut, S., & Adem, K. (2023). Derin öğrenme ve makine öğrenmesi yöntemleri kullanılarak gelişmekte olan ülkelerin finansal enstrümanlarının etkileşimi ile Bist 100 tahmini. Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi, 12(1), 52-63. https://doi.org/10.28948/ngmuh.1131191
9. Eşsiz, E. S. (2022). Short-term wind power prediction with harmony search algorithm: Belen region. Turkish Journal of Engineering, 6(3), 251-255. https://doi.org/10.31127/tuje.970959
10. Balcı, M., Yüzgeç, U., & Dokur, E. (2022, May). Rüzgâr hızı tahmini için ayrıştırmaya dayalı hibrit yöntemlerin karşılaştırmalı bir çalışması. In International Conference on Emerging Sources in Science, 118-135.
11. Balti, H., Abbes, A. B., Mellouli, N., Sang, Y., Farah, I. R., Lamolle, M., & Zhu, Y. (2021). Big data-based architecture for drought forecasting using LSTM, ARIMA, and Prophet: Case study of the Jiangsu Province, China. In 2021 International Congress of Advanced Technology and Engineering (ICOTEN), 1-8. https://doi.org/10.1109/ICOTEN52080.2021.9493513
12. Baykal, T., Taylan, D., & Terzi, Ö. (2023). Isparta İli için gelecekteki olası meteorolojik kuraklık değerlendirmesi. Doğal Afetler ve Çevre Dergisi, 9(1), 90-100. https://doi.org/10.21324/dacd.1165500
13. Canıtez, M. A., & Savaş, S. (2022). Kripto para piyasa değeri tahmini için özellik tabanlı LSTM ve ARIMA karşılaştırması. In 3rd International Conference on Applied Engineering and Natural Sciences, 1311-1317.
14. Dave, E., Leonardo, A., Jeanice, M., & Hanafiah, N. (2021). Forecasting Indonesia exports using a hybrid model ARIMA-LSTM. Procedia Computer Science, 179, 480-487. https://doi.org/10.1016/j.procs.2021.01.031
15. Demirtop, A., & Işık, A. H. (2023). Yapay sinir ağları ile rüzgâr enerji verimliliğine yönelik yeni bir tahmin yaklaşımı: Çanakkale İli Bozcaada Örneği. Uluslararası Mühendislik Tasarım ve Teknoloji Dergisi, 5(1-2), 25-32.
16. Devi, A. S., Maragatham, G., Boopathi, K., & Rangaraj, A. G. (2020). Hourly day-ahead wind power forecasting with the EEMD-CSO-LSTM-EFG deep learning technique. Soft Computing, 24(16), 12391-12411. https://doi.org/10.1007/s00500-020-04680-7
17. Elsaraiti, M., & Merabet, A. (2021). A comparative analysis of the arima and lstm predictive models and their effectiveness for predicting wind speed. Energies, 14(20), 6782. https://doi.org/10.3390/en14206782
18. Erden, C. (2023). Derin Öğrenme ve ARIMA Yöntemlerinin Tahmin Performanslarının Kıyaslanması: Bir Borsa İstanbul Hissesi Örneği. Yönetim ve Ekonomi Dergisi, 30(3), 419-438. https://doi.org/10.18657/yonveek.1208807
19. Ji, L., Zou, Y., He, K., & Zhu, B. (2019). Carbon futures price forecasting based with ARIMA-CNN-LSTM model. Procedia Computer Science, 162, 33-38. https://doi.org/10.1016/j.procs.2019.11.254
20. Kamber, E., Körpüz, S., Can, M., Aydoğmuş, H. Y., & Gümüş, M. (2021). Yapay Sinir Ağlarina Dayali Kisa Dönemli Elektrik Yükü Tahmini. Endüstri Mühendisliği, 32(2), 364-379. https://doi.org/10.46465/endustrimuhendisligi.820509
21. Liu, X., Lin, Z., & Feng, Z. (2021). Short-term offshore wind speed forecast by seasonal ARIMA-A comparison against GRU and LSTM. Energy, 227, 120492. https://doi.org/10.1016/j.energy.2021.120492
22. Othman, M. M. (2023). Modeling of daily groundwater level using deep learning neural networks. Turkish Journal of Engineering, 7(4), 331-337. https://doi.org/10.31127/tuje.1169908
23. Sevinç, A., & Kaya, B. (2021). Derin öğrenme ve istatistiksel modelleme yöntemiyle sıcaklık tahmini ve karşılaştırılması. Avrupa Bilim ve Teknoloji Dergisi, (28), 1222-1228. https://doi.org/10.31590/ejosat.1014106
24. Shao, B., Song, D., Bian, G., & Zhao, Y. (2021). Wind speed forecast based on the LSTM neural network optimized by the firework algorithm. Advances in Materials Science and Engineering, 2021(1), 4874757. https://doi.org/10.1155/2021/4874757
25. Wang, J., & Wang, J. (2023, March). Short-term Wind Speed Forecast Using ARIMA Based on EEMD Decomposition. In Journal of Physics: Conference Series, 2450(1), 012020. https://doi.org/10.1088/1742-6596/2450/1/012020
26. Zhang, R., Guo, Z., Meng, Y., Wang, S., Li, S., Niu, R., ... & Li, Y. (2021). Comparison of ARIMA and LSTM in forecasting the incidence of HFMD combined and uncombined with exogenous meteorological variables in Ningbo, China. International Journal of Environmental Research and Public Health, 18(11), 6174. https://doi.org/10.3390/ijerph18116174
27. Zhang, M., Wang, Y., Zhang, H., Peng, Z., & Tang, J. (2023). A novel and robust wind speed prediction method based on spatial features of wind farm cluster. Mathematics, 11(3), 499. https://doi.org/10.3390/math11030499
28. Zhao, J., Nie, G., & Wen, Y. (2023). Monthly precipitation prediction in Luoyang city based on EEMD-LSTM-ARIMA model. Water Science & Technology, 87(1), 318-335. https://doi.org/10.2166/wst.2022.425
29. Bektaş, B. N. (2019). Understanding the conservation process of gallipoli historical site. [Master's thesis, Middle East Technical University].
30. https://www.mgm.gov.tr/
31. Sahoo, B. B., Jha, R., Singh, A., & Kumar, D. (2019). Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. Acta Geophysica, 67(5), 1471-1481. https://doi.org/10.1007/s11600-019-00330-1

32. Sundermeyer, M., Ney, H., & Schlüter, R. (2015). From feedforward to recurrent LSTM neural networks for language modeling. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23(3), 517-529. https://doi.org/10.1109/TASLP.2015.2400218
33. Kamath, U., Liu, J., & Whitaker, J. (2019). Deep learning for NLP and speech recognition, 84. Cham, Switzerland: Springer.
34. Gers, F. A., Schraudolph, N. N., & Schmidhuber, J. (2002). Learning precise timing with LSTM recurrent networks. Journal of Machine Learning Research, 3, 115-143.
35. Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM--a tutorial into long short-term memory recurrent neural networks. Neural and Evolutionary Computing. https://doi.org/10.48550/arXiv.1909.09586
36. Yang, H., Luo, L., Chueng, L. P., Ling, D., & Chin, F. (2019). Deep learning and its applications to natural language processing. Deep learning: Fundamentals, theory and applications, 89-109. https://doi.org/10.1007/978-3-030-06073-2_4
37. Wang, Y., Jiang, L., Yang, M. H., Li, L. J., Long, M., & Fei-Fei, L. (2018). Eidetic 3D LSTM: A model for video prediction and beyond. In International Conference on Learning Representations.
38. Wang, X., Wang, Y., Yuan, P., Wang, L., & Cheng, D. (2021). An adaptive daily runoff forecast model using VMD-LSTM-PSO hybrid approach. Hydrological Sciences Journal, 66(9), 1488-1502. https://doi.org/10.1080/02626667.2021.1937631
39. Schaffer, A. L., Dobbins, T. A., & Pearson, S. A. (2021). Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions. BMC Medical Research Methodology, 21, 1-12. https://doi.org/10.1186/s12874-021-01235-8
40. Loganathan, N., & Ibrahim, Y. (2010). Forecasting international tourism demand in Malaysia using Box Jenkins Sarima application. South Asian Journal of Tourism and Heritage, 3(2), 50-60.
41. Valipour, M., Banihabib, M. E., & Behbahani, S. M. R. (2013). Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. Journal of Hydrology, 476, 433-441. https://doi.org/10.1016/j.jhydrol.2012.11.017
42. Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. Peerj Computer Science, 7, e623. https://doi.org/10.7717/peerj-cs.623
43. Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A., & Kamaev, V. A. E. (2013). A survey of forecast error measures. World Applied Sciences Journal, 24(24), 171-176. https://doi.org/10.5829/idosi.wasj.2013.24.itmies.80032
44. Barbounis, T. G., Theocharis, J. B., Alexiadis, M. C., & Dokopoulos, P. S. (2006). Long-term wind speed and power forecasting using local recurrent neural network models. IEEE Transactions on Energy Conversion, 21(1), 273-284. https://doi.org/10.1109/TEC.2005.847954
45. Lo, A. W., Siah, K. W., & Wong, C. H. (2019). Machine learning with statistical imputation for predicting drug approvals, 60, 10.1162.
46. Arya, F. K., & Zhang, L. (2015). Time series analysis of water quality parameters at Stillaguamish River using order series method. Stochastic Environmental Research and Risk Assessment, 29, 227-239. https://doi.org/10.1007/s00477-014-0907-2
47. Beard, E., Marsden, J., Brown, J., Tombor, I., Stapleton, J., Michie, S., & West, R. (2019). Understanding and using time series analyses in addiction research. Addiction, 114(10), 1866-1884. https://doi.org/10.1111/add.14643
48. Franses, P. H., & Paap, R. (2004). Periodic time series models. OUP Oxford.
49. Sharifani, K., & Amini, M. (2023). Machine learning and deep learning: A review of methods and applications. World Information Technology and Engineering Journal, 10(07), 3897-3904.

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Predicting the compressive strength of self-compacting concrete using artificial intelligence techniques: A review

Terlumun Sesugh[*1], **Michael Onyia**[2], **Okafor Fidelis** [2]

[1] *Air Force Institute of Technology, Department of Civil and Environmental Engineering, Nigeria, terlumunsesugh8@gmail.com*
[2]*University, University of Nigeria Nsukka, Department of Civil Engineering, Nigeria, michaelonyia2023@gmail.com, okaforfo2023@gmail.com*

Cite this study: Sesugh, T., Onyia, M., & Fidelis, O. (2024). Predicting the compressive strength of self-compacting concrete using artificial intelligence techniques: A review. Turkish Journal of Engineering, 8 (3), 537-550

https://doi.org/10.31127/tuje.1422225

Keywords
Compression strength prediction
Self-compacting concrete
Artificial intelligence

Review Article

Received:18.01.2024
Revised: 13.03.2024
Accepted: 14.03.2024
Published: 11.07.2024

Abstract

Concrete is one of the most common construction materials used all over the word. In estimating the strength properties of concrete, laboratory works need to be carried out. However, researchers have adopted predictive models in order to minimize the rigorous laboratory works in estimating the compressive strength and other properties of concrete. Self-compacting concrete which is an advanced form of construction is adopted mainly in areas where vibrations may not be possible due to complexity of the form work or reinforcement. This work is targeted at predicting the compressive strength of self-compacting concrete using artificial intelligence techniques**.** A comparative performance analysis of all techniques is presented. The outcomes demonstrated that training in a Deep Neural Network model with several hidden layers could enhance the performance of the suggested model. The artificial neural network (ANN) model, possesses a high degree of steadiness when compared to experimental results of concrete compressive strength. ANN was observed to be a strong predictive tool, as such is recommended for formulation of many civil engineering properties that requires predictions. Much time and resources are saved with artificial intelligence models as it eliminates the need for experimental test which sometimes delay construction works.

1. Introduction

The world's most popular building material has historically been concrete [1]. It has become man's most trusted companion due to its widespread use in construction projects across the world. At least since the Roman Empire's era, concrete has been instrumental for the expansion and improvement of many different cultures. In addition to its fluid-to-solid transformation, the "magic" of concrete lies in its inexpensive price, vast availability, malleability, plasticity, adaptability, high compressive strength, stiffness, and durability [1-4]. Many researchers are still enthusiastic about learning more about this niche of concrete engineering. Even for novel substances like reactive powder concrete, the future seems bright. Concrete's compressive strength ranges from 200 to 800 MPa [2-4].

It is standard practice to take a sizable enough sample of a combination to draw conclusions about the entire concrete mix. According to [5-7], testing concrete on day 7 is used to establish the early-age strength. Since concrete achieves its full strength at 28 days, the examination of the samples would also take that long. The result of the 28-day test can be used as a guide to predict the concrete's compressive strength as it ages. The procedures resemble those described in design regulations. However, these design standards fall short of the mark when concrete ingredients move beyond the traditional cement, aggregate, and water. The calculation of a concrete mix's compressive strength grows more complicated as the sorts of components used to manufacture it vary. Additionally, unaccounted-for components like pozzollans and admixtures are utilized when the type of concrete varies (for example, high strength concrete), rendering it impossible to clearly define the inter-relationship of concrete ingredients at the molecular level [8-12]. In cases when the outcome of the compressive strength might be influenced by factors

like "effect of admixtures," "temperature," etc., the empirical relationship suggested by the codes falls short of capturing the real picture, then, a trustworthy model is necessary to ascertain the strength of concrete samples at different ages in order to complete urgent projects that cannot wait for the results of 28-day testing. The time saved waiting for test results can be converted into monetary gains due to the productive nature of the construction process. This requires looking into alternate approaches. There is much room for debate and exploration regarding the usefulness of these non-traditional approaches.

There has been a lot of interest in utilizing artificial neural networks (ANN) to calculate concrete compressive strength over the previous year's [13, 14]. Several researches have used combinatory models, such as fuzzy logic (FL), genetic programming (GP), artificial neural networks (ANN) etc. [15] forecast how ground granulated blast furnace slag will affect concrete's strength over time. Fast learning with only moderately accurate performance was made possible by the cascade correlation type of ANN [16-18] in order to capture the inherently nonlinear character of patterns in the concrete properties. For a certain category of problems, specifically the measurement of compressive strength, this helps save time and money. The outstanding performance outcomes of combinatorial formations of artificial neural networks and meta-heuristic algorithms are making them increasingly popular for dealing with challenging structural engineering difficulties. Concrete compressive strength determination for deep beams attached to sheared walls, for instance, has been shown to benefit from a hybrid multilayer perceptron (HMP) network according to [18-22]. Artificial neural networks (ANNs) and evolutionary search methods like genetic algorithms (GAs) are combined in evolutionary artificial neural networks (EANNs) as described by [20, 21]. Kovačević [22] lays out the framework for evolutionary artificial neural networks (EANNs). EANNs have been tremendously effective for the purpose of detecting defects in concrete's structure and compressive strength [22-24]. To more precisely and affordably estimate the concrete compressive strength, EANN is being studied as a potential replacement for both costly mathematical models and damaging trials.

2. Method

The authors focused on machine learning models and the compressive strength of self-compacting concrete in order to achieve the goals of this study: A range of websites, including Scholarly Journals, Elsevier, Springers, SCOPUS, the Web of Science, Turkish Journal of Engineering, IEEE, and Science Direct, were consulted in order to acquire scholarly sources for this study. Only published books that the authors felt offered the most intelligent and pertinent answers to their research topic were chosen. Following an examination of these sources, 81 sources in all were determined to be very relevant to the study. Figure 1 presents a flowchart that outlines the research approach used in this work.

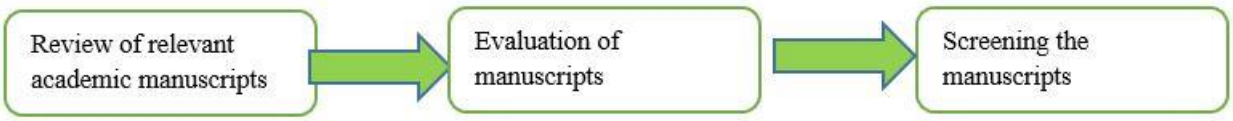


Figure 1. Flow chart of the approach used in obtaining relevant articles for the research.

A total of 80 articles that were researched came from academic journals, which accounted for a greater percentage of the total number of academic sources, also EFNARC guidelines and specifications for self-compacting concrete was also used in this research. Thus, the total number of academic sources was 81. The study applies the idea of modern construction technology to give a methodical assessment of the compressive strength of self-compacting concrete using machine learning models.

2.1. Predicting concrete compressive and flexural strength

There are numerous techniques for predicting compressive strength as adopted by many researchers across the world.

2.1.1. Empirical methods

In an attempt to directly represent the experimental procedure into a given connection, empirical methodologies have been developed. The strength of the connection indicates how important each component was in generating the final result. The created connection must be able to suit the experimental data; therefore, they often rely on extensive experimental analysis. Most empirical relationships for forecasting compressive strength make an effort to connect the strength to the water cement ratio, despite the fact that this ratio typically has a negative effect on the predicted strength. Other times, the models take into account information about compressive strength that is already available and employ experimentally estimated coefficients to establish a connection between the available information and the necessary information [10, 25].

2.1.2. Computational modelling

Complex equations of thermodynamics are the foundation of finite element analysis. Such modelling is often done with the use of computers. They rely on an accurate depiction of the microstructure of the concrete. In order to simulate accurately the hydrations (and other processes) of the different particle sizes in the cement using a computer simulation (pixel based), it is necessary to cast the cement randomly in a unit cell space. It is also possible to encode the processes empirically, after adjusting for experimental data [10, 25, 26].

2.1.3. Mechanical modeling

Parameters in mechanical models are typically represented using a spring-and-dashpot structure. The cement matrix is typically viewed as the spring and the time element (age of testing) as the dashpot when applying this theory to forecast the compression strength of concrete. When the model is compared to a testing dataset, its accuracy declines. Compressive strength is also significantly impacted by the dashpot component, especially at younger ages [10, 25, 26].

2.1.4. Statistical methods

In order to provide the most accurate description of the connections between variables, statistical techniques rely on experimental data and apply mathematical equations. In the realm of statistics, multilinear regression is by far the most used approach. Although statistical approaches are intuitive and simple to implement, their data-heavy nature might be a hindrance in some situations. Also, their effectiveness varies with the mathematical function used to fit the data, making them less reliable than alternative approaches [10, 25-27].

2.1.5. Regression analysis

It is widely agreed that regression analysis is a crucial part of any respectable statistical modelling strategy. The convenience of the statistical method in terms of computing coefficients, which may be interpreted in terms of efficiency gains, [28-30], such an approach, however, is typically convoluted and difficult to follow. This trait is ignored for the time being, in order to create a confidence interval for a set of predictions, statistical models must be mathematically rigorous. Correlation analysis also shows how the main constituents affect the final product. R^2, sometimes termed the correlation coefficient, is a metric used to compare the effectiveness of different regression equations. R^2 measures how well a model can make predictions. It's a metric for assessing the model's ability to account for differences in output that result from differences in input values. A value of 0 indicates that no variance in Y is interpreted by the calculated regression model, whereas a value of 1 indicates that all points are on the regression line. Slump test findings and concrete density were presented by [28-30] to complement mix proportions of component materials in estimating high performance concrete's compressive strength.

This was done so that results from tests at varying ages would more accurately reflect the progression from weakness to strength. The final multi-variable power equation benefited from the addition of these independent variables, with a correlation coefficient of 99.99% being achieved. Good connections between his predicted compressive strength and experimental data were also discovered in the work of [31]. The inputs to his model were standardized according to a fixed matrix formula in order to enhance development of concrete mixtures. Regression methods have long been popular because of their simplicity and the little amount of time they require for modelling. Each regression model has great predictive power for low levels of non-linearity between reactants and products (as is typically the case in concrete). This is evident in the work of [32, 33], in which he attempted to evaluate this strategy in comparison to the artificial neural network approach and concluded that the latter provided much superior forecasts.

2.1.6. Artificial intelligence

According to Webster's New World College Dictionary, AI is "the ability of machines or programs to function in ways that simulate human intelligence in tasks such as reasoning and learning." One can then wonder what kinds of issues call for a computer solution that "thinks like a human brain." Knowledge-based inference with partial or unclear data, different types of perception and learning, and control, prediction, classification, and optimization are only a few examples [34, 35]. They can be used to mimic a phenomenon in civil engineering situations where the underlying process is unclear. Neural networks and genetic algorithms are two types of AI techniques; they have structural and functional similarities to microscopic biological models (neural networks that mimic the brain's) and genes (computer programs that 'mutate' to improve their performance), respectively. Adaptive Network based Fuzzy Inference Systems (ANFIS) Fuzzy systems are two further examples of modern artificial intelligence. Artificial Neural Network Fuzzy Inference System (ANFIS) uses a hybrid of the two. Training using linguistic interpretation of variables with fuzzy logic and neural networks are both strong suits of the network created in this way [34, 35]. Methods based on artificial intelligence are potent due to their ability to learn and provide relevant models, even for a wide variety of data. A 'near to reality' conclusion, however, requires more data than just this.

2.2. Fuzzy logic

It was Lotfi Zadeh who originally proposed the use of fuzzy logic [33-37]. Among the various fuzzy inference system approaches he paved the way for include Mamdani, Takagi, and Sugeno. His work on fuzzy logic is credited with revolutionizing the way people represent the world.

According to [33-37], fuzzy logic is a collection of mathematical principles for knowledge representation based on degrees of membership. This stands in stark contrast to the logical values in traditional Boolean logic of either "completely false" (=0) or "completely true" (=1). The Boolean logic is a representation of information having binary boundaries (black or white, for example). The truth usually differs significantly from expectations. Parameters within a set are not always well delineated, and a "grey area" often exists when there is an element that belongs to both sets to varying degrees. An individual may, for instance, have the heights of 10 persons in a set, with the 'tall' group ranging from 1.75 m to 2.20 m and the 'short' group ranging from 1.50 m to 1.74 m. Consideration of a crisp set may seem insufficient

under these conditions. Some researchers [33-37] summed up this kind of dichotomous thinking by saying that it is "painfully simplified" and "in many circumstances, lacking rapport with reality." Fuzzy sets can be used to simulate the aforementioned crisp set more accurately and robustly. They [33-37] define a fuzzy set as a class of objects having a continuum of degrees of membership. A membership (characteristic) function, in this case, gives each item in the set a score between one and zero based on its membership degree in the set.

Modelling non-linear systems and creating sophisticated controllers are two of the many applications of fuzzy logic control, a strong mathematical tool. In concrete, each component serves a purpose both alone and in concert with the others. This will cause the relationship between the number of ingredients and the resulting compressive strength to be non-linear. FLC can help reduce this nonlinearity since it is implemented when the complexity of a system prevents the use of other modelling approaches [33-37].

2.3. Neural network

The dramatic shift in how people approach problems that neural networks represent is happening now. Since work on artificial intelligence began in the 1950s, neural networks have been at the forefront of efforts to expand robots' limited utility from muscle labor to intellect work. The term "artificial neural network" can refer to a number of different things. In that respect, they are analogous to biological brain networks. In this metaphor, the axon represents the outputs and the synapses represent the weights. In artificial neural networks, the ubiquitous neuron is also referred to as a "processing element." References [36-39] concisely defined artificial neural networks as a class of massively parallel designs that address complex issues by coordinating the efforts of several, relatively uncomplicated processors (or "artificial neurons"). A perceptron is a simple neural network with a one-to-one mapping of inputs to outputs. More complex neural networks may have many layers and use a variety of activation functions. Classification of artificial neural networks helps shed light on their many forms.

Damage detection, identification of structural systems, modelling of material behavior, structural settlement, control and optimization, monitoring ground water, and determining concrete mix proportions are just some of the many areas where ANNs have been put to use [40]. Reference [41] examined the use of artificial neural networks in prediction models to supplement their work in fuzzy logic and attempt to forecast the compressive strength of self-compacting concrete. 500 iterations were performed on the ANN model, which had a hidden layer of 6 neurons. Their successes were comparable, as were the degrees of inaccuracy in their measurements. The resultant model improved upon the fuzzy logic result in that it was able to estimate the strength with an R^2 value of 0.9767. The selected ANN design had a single hidden layer with eleven secret neurons. Adding a second hidden layer improved the accuracy of the model [42]. Using an ANN design with two hidden layers and nine and eight neurons in the first and second hidden layers, respectively, yielded the lowest absolute percentage error (=0.000515). Sarıdemir [43] also considered the advantages of a multi-layer architecture. The material issue was also given proper consideration in his model, as seen by the use of chemical analysis data for fly ash, gradient, and sand chemical compositions. R^2 = 0.9557 was found for compressive strength, while 0.9119 was found for flexural tensile strength, indicating that the prediction model performed well. The influence of water to binder ratio on compressive strength was also considered in the compressive strength forecasting model [43, 44]. Compressive strength was shown to decrease with increasing water-to-binder ratio. Their model has an R^2 of 0.9944 for reproducing experimental outcomes and an R^2 of 0.9767 for predicting testing samples. Concrete strength prediction methodologies were compared with an artificial neural network approach [45-47], the results showed that the ANN approach provided an accurate forecast for low and medium strength concretes. Despite promising findings from a multiple linear regression model, the ANN method performed much better.

2.4. Genetic programming

In order to evaluate a computer's performance on a given job, genetic programmers use a set of "instructions" and a "fitness process" known as "genes expression programming" (GEP). Each node is a piece of code running on a computer, making this a subtype of genetic algorithm (GA). It's a method for improving the performance of a computer programme by moving its components to the most advantageous physical location, as defined by the conditions the it must satisfy. Here are the three genetic manipulations [43-46].

Some of the key adjusting factors are detailed in [43-46] article. The GP building process begins with an initial population size of 49 without cement/FA substitution and 27 with cement/FA substitution of 0.15. The curing time was used to further classify each dataset in the GP model, just as it did in the ANN model. When the desired result is a 28-day compressive strength, four input parameters were selected: water, cement, coarse aggregate, and fine aggregate.

Reference [44-46] presented regression equations for predicting in situ concrete compressive strength based on data from ready-mixed concrete mixture proportions and on-site compressive strength testing. They employed 1442 compressive strength test data from 68 distinct mixture types with specified compressive strengths between 18 and 27 MPa, water to cement ratios between 0.39 and 0.62, and maximum sizes of aggregate was between 25 and 100 mm. They [45-47] tested a proposed model for predicting the compressive strength of concrete using in situ data.

2.5. Parameters in machine learning models

The hyper parameters of a machine learning algorithm are external guidelines that inform decision making and control how the algorithm learns. Machine learning engineers must first set these parameters before

training the algorithm, [48, 49, 50]. Hyper parameters include things like the learning rate, the number of clusters in a clustering algorithm, and the number of branches on a regression tree. In reaction to the training data, parameters start to form as the algorithm is trained and guided by the hyper parameters. The weights and biases that the algorithm develops during training are among these parameters. The model parameters are the final values for a machine learning model that, in an ideal world, fit a data set without exceeding or falling short of [48-50].

2.6. Hyper parameters in machine learning models

A hyper parameter in machine learning is a parameter whose value directs the process of learning. On the other hand, training is used to determine the values of other parameters, which are usually node weights. Hyper parameters can be divided into two categories: model hyper parameters, which pertain to the model selection task and cannot be inferred during machine learning, and algorithm hyper parameters, which impact the speed and caliber of learning but theoretically have no effect on the model's performance. The topology and size of a neural network are two instances of model hyper parameters [48-50].

Hyper parameters of an algorithm include learning rate, batch size, and mini-batch size. A mini-batch size would be a smaller sample set, whereas batch size can refer to the entire data sample. While some straightforward methods (like ordinary least squares regression) require none, other model training procedures require different hyper parameters. The training algorithm extracts the parameters from the data and applies them to these hyper parameters. The selection of a model's hyper parameters can affect how long it takes to train and test. Typically, a hyper parameter is of an integer or continuous form, which results in mixed-type optimization issues. Some hyper parameters can only exist if other hyper parameters are valued, for example, in a neural network, the number of layers may affect the size of each hidden layer [48-50].

2.7. Untrainable parameters

Because hyper parameters aggressively boost a model's capacity and have the potential to push the loss function to an undesirable minimum (overfitting to and picking up noise in the data), rather than accurately mapping the richness of the structure in the data, they can sometimes prevent hyper parameters from being learned from the training set. The degree of a polynomial equation that fits a regression model, for instance, would increase until the model perfectly suited the data if we treated it as a trainable parameter. This would result in reduced training error but poor generalization performance.

2.8. Tuning in ML

Only a few hyper parameters account for the majority of performance variation. How much more performance can be obtained by tweaking an algorithm, hyper parameter, or combination of hyper parameters is called its tuning ability. The two most important hyper parameters for machine learning models are learning rate and network size; batching and momentum have no discernible impact on the model's performance. Research has shown that mini-batch sizes between 2 and 32 yield the best results, despite some studies advocating for mini-batch sizes in the thousands.

2.9. Robustness of ML models

Learning has an inherent stochasticity, which suggests that the performance measured by the empirical hyper parameter may not accurately reflect the true performance. Without considerable simplification and robustness, methods that are not resistant to straightforward modifications in hyper parameters, random seeds, or even various implementations of the same algorithm, cannot be incorporated into mission-critical control systems. In particular, algorithms for reinforcement learning must have their performance evaluated across a large number of random seeds and their sensitivity to hyper parameter selection evaluated. Because of the high variance, their evaluation using a small number of random seeds is unable to fully capture performance. Certain reinforcement learning techniques—like Deep Deterministic Policy Gradient (DDPG) are more responsive to selections of hyper parameters than others.

2.10. Optimization

Through the use of hyper parameter optimization, a tuple of hyper parameters is found to produce an optimal model on test data that minimizes a predetermined loss function. The corresponding loss is returned by the objective function after receiving a tuple of hyper parameters.

2.11. Reproduction in ML models

In addition to fine-tuning hyper parameters, machine learning entails reproducibility checks, parameter and result organization, and storage. When there's no strong infrastructure in place for this kind of thing, research code tends to change quickly and jeopardizes important features like reproducibility and bookkeeping. Machine learning online collaboration platforms go one step further by enabling scientists to automatically exchange, arrange, and communicate about experiments, data, and algorithms. Deep learning models can be especially challenging to reproduce [51].

2.12. Creating machine learning models

Using either labeled, unlabeled, or a combination of the two, algorithms are trained to create machine learning models [51,52]. There are four main machine learning algorithms available:

2.12.1. Supervised learning

When an algorithm is trained on "labeled data," or data that has been labeled so that an algorithm can

successfully learn from it, supervised learning takes place. Training labels aid in ensuring that the final machine learning model is capable of classifying data in the way that the researcher has intended.

2.12.2. Unsupervised learning

To train an algorithm, unsupervised algorithms use unlabeled data. The algorithm generates its own data clusters during this process by looking for patterns in the data itself. Researchers who are trying to discover patterns in data that are currently unknown to them can benefit from unsupervised learning and pattern recognition.

2.12.3. Semi-supervised learning

This technique trains an algorithm by combining labeled and unlabeled data. This procedure involves training the algorithm with a small amount of labeled data initially, and then training it with a much larger amount of unlabeled data.

2.12.4. Reinforcement learning

This machine learning approach assigns positive and negative values to actions that are desired and undesirable. In order to maximize rewards through trial and error, the intention is to incentivize programs to steer clear of the unfavorable training examples and toward the positive ones. Unsupervised machine learning can be guided by reinforcement learning.

2.13. Types of machine learning models

Classification and prediction problems are the two main categories of machine learning problems. Models derived from algorithms created for either classification or regression (a technique used for predictive modeling) are used to approach these problems [53]. Depending on how it is trained, an algorithm may occasionally be used to produce regression or classification models. A list of well-known algorithms for building regression and classification models can be seen below. Models of classification Support vector machines, random forests, decision trees, logistic regression, naive Bayes, and K-nearest neighbor (KNN) Models of regression K-nearest neighbor (KNN) regression, decision trees, random forests, neural network regression, and linear regression

2.14. Overfitting in machine learning (ML)

In machine learning (ML), overfitting occurs when a model performs poorly on new, unseen data because it is too closely aligned with the training data it was trained on. Its primary goals are to detect, enable the model to generalize to new data, and guarantee that machine learning becomes less specialized to the training set of data [54,55]. Overfitting is used in machine learning tasks such as image recognition and natural language processing. After overfitting is eliminated, these models may also be able to predict new data more precisely and produce reliable forecasts for practical uses. It happens when there is insufficient data, when there is a lot of unnecessary information in the data set, when training on a specific data set is prolonged, and when the model's complexity needs to be reduced. In machine learning (ML), the term "overfitting" refers to the situation in which a model becomes excessively complex and fits too closely to the training set of data. Helping the model achieve better generalization performance on new data is the main goal of addressing overfitting. It can be found by keeping an eye on all losses, analyzing the learning curve, incorporating regularization terms, running cross-validation on the model, and visually examining the prediction to make sure it closely matches the training set.

The following methods can be used to avoid overfitting in machine learning: dropout, feature selection, early stopping, cross-validation, regularization using L1 and L2, and data augmentation. When a machine learning model becomes overly specialized in the training data and is unable to generalize well to new, unseen data, this is known as overfitting. This occurs in neural networks when the machine learning model gives less significant information in the training data more weight. Because it is unable to distinguish between relevant, crucial data that form the pattern and noisy data, this model requires assistance in order to produce accurate forecasts about new data.

2.15. Causes of overfitting

The model's training dataset is dirty and contains a lot of noise; additionally, the training dataset should be larger, when only a portion of the available data is used to build the model, resulting in an inaccurate representation of the entire dataset. Therefore, by looking at validation metrics such as accuracy and loss in overfitting decision trees, one can identify an overfitting model. Due to the effect of overfitting, these metrics typically have a tendency to rise to a certain point before beginning to decline or plateau. In addition, this model works to achieve an ideal fit, and once it does, the trend of these metrics starts to flatten or decline. To solve this, it is crucial to strike the correct balance between the complexity of the model and the amount of training data that is available.

2.16. Identifying overfitting in ML models

The overfitting can be detected in a few different ways, including the following:

1. All losses need to be kept an eye on, even if the validation loss rises while the training loss falls.
2. It is important to keep a close eye on the machine learning curve because any divergence between the training and validation curves indicates overfitting. Since the regularization term avoids overfitting, it must be included in the loss function.
3. Lastly, visual inspection of the model's prediction is necessary to determine whether the model fits the training data too closely. If so, this indicates overfitting.

2.17. Preventing ML from overfitting

The following methods are among the many that aid in preventing overfitting in machine learning:

1. The model increases the loss function by adding a penalty term through the joint Regularization L1 & L2 techniques. It thus deters the model from fitting the training set too closely, which avoids overfitting.
2. Cross-validation aids in its avoidance. The training automatically ends just before the point of overfitting when the early stopping technique in the model is applied.
3. Using the data augmentation technique improves the creation of new data from previously used training data. This keeps the model from overfitting by increasing its exposure to a wider range of data samples.
4. The model can be trained to prevent overfitting to noisy data by reducing the number of features through feature selection.
5. The dropout technique forces the model to learn more robust and generalizable representations, preventing data overfitting. Consequently, by using these techniques, the overfitting issue may be lessened, leading to the creation of a machine-learning model that is more precise and useful.

2.18. Fuzzy logic approach

The four main components of each fuzzy system are, Fuzzy rule base, Fuzzification, and Defuzzification and Inference Engine [56]. Figure 2 present a general fuzzy logic model architecture.

The Fuzzification is the first step that estimates the fuzzy input and converts it into one or more fuzzy sets. Gaussian and trapezoidal fuzzifications are the two types. In that kind of fuzzy logic, any item can belong to any number of subsets of the universal set. The foundation of fuzzy rules is linguistic IF-THEN-constructions with the overall procedure "IF A THEN B," where A and B are schemes incorporating linguistic variables. The premise is designated as A, while the rule's importance is designated as B. In order to account for ambiguity and inaccuracy, linguistic variables and fuzzy IF-THEN rules are used. Depending on how unique the challenges are, there are two different types of rule techniques. In a fuzzy inference engine, fuzzy rule base is integrated and used in the analysis of fuzzy outputs [56,57]. The Defuzzification process transforms outputs from a fuzzy inference engine into precise numerical values and crisp output. The flow chart for the GP is shown Figure 3.

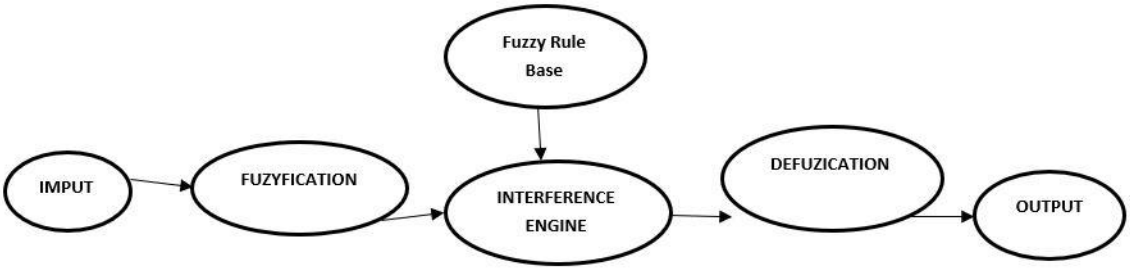


Figure 2. Fuzzy Logic architecture.

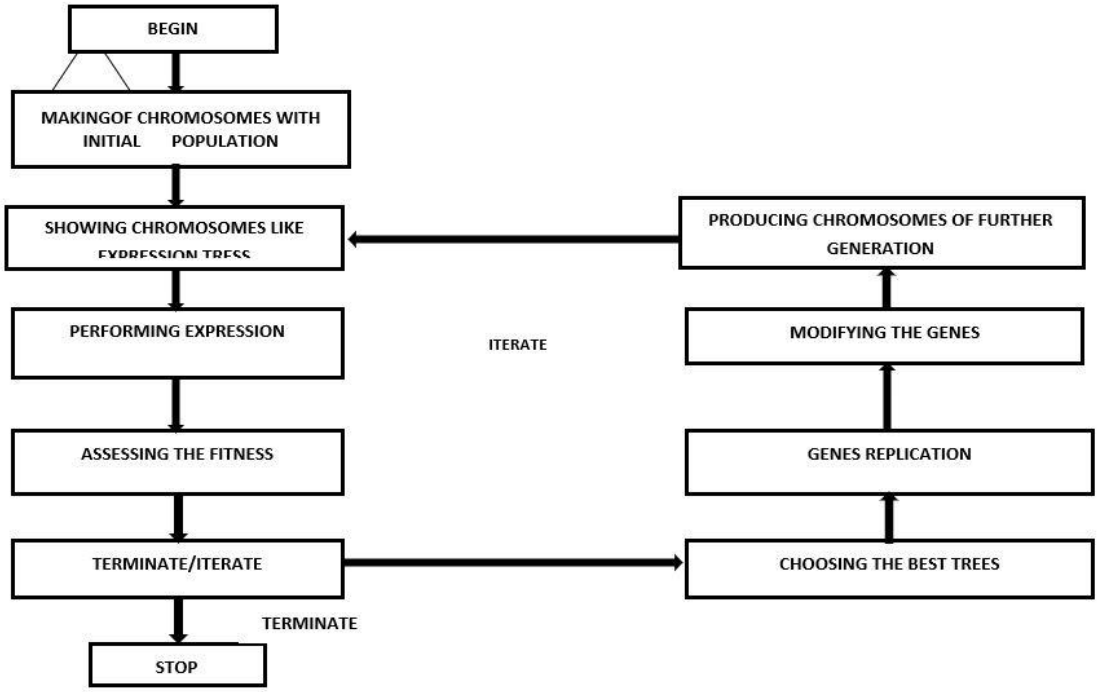


Figure 3. Genetic algorithm flow chart diagram.

2.19. Evaluating the performance of the GP model

For a model to be deemed adequate, it is recommended that the dataset ratio to the total number of input features equal 3, and a ratio of 5 is desired. The GEP model can be validated using various statistical computations from the training, testing, and validation sets. In order to assess if the model was effectively trained and whether there is a significant association between the model and experimental data with little error, the values of the RMSE, MAE, and RSE parameters will be computed for the testing phase [56-58].

With the aid of several statistical techniques, the GEP model will also be examined for external validation. It is suggested that one of the regression lines (k or k′) passing through the origin's slope should be close to 1. The same dataset will also be used in this study's linear regression model, which will be used to calculate the SCC's fc. It is important to remember that the robustness and generalizability of the resulting model depend on the fitting parameters. The fitting parameters for the GEP algorithm will be determined using test runs or experimental results. How long the program will run depends on the population size (the number of chromosomes). Depending on the quantity and complexity of the prediction model, the levels will be taken into account as the population size. The head size and the number of genes were two variables that the algorithm that created the models in this study used to define the architecture of the models [59-61].

The head size, or "head size," of the model, which refers to the size of the model's "head," contributes to determining how complex each term is inside the model. On the other hand, the number of sub-ETs (basic data structures) that make up the model depends on the number of genes. Five alternative head sizes will be considered in this study: 8, 9, 12, 10, and 14. There will be either three or four genes [61, 62]. The GEP algorithm will be used to develop the precise parameters for each model.

2.20. Model evaluation criteria

One common performance indicator is the correlation coefficient (R). R is insensitive to the division and multiplication of output values by constants, hence it cannot be used as a primary indicator of how well the model predicts. As a result, the study will also evaluate the mean absolute error (MAE), relative squared error (RSE), and relative root mean square error (RRMSE) [63-67]. To evaluate the model's performance in regard to both the RRMSE and R, Gandomi and Roke propose a performance index [56]. Equations 1 to 7 provide the mathematical formulas for these error functions.

$$Fc = f(\alpha + \beta_1 Y_1 + \beta_2 Y_2 + \beta_3 Y_3 + \beta_4 Y_4 \ldots\ldots + \beta_n Y_n) + £_1 \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(ei - mi)^2}{n}} \quad (2)$$

$$MAE = \frac{\sum_{i=1}^{n}|ei - mi|}{n} \quad (3)$$

$$RSE = \frac{\sum_{i=1}^{n}(mi - ei)^2}{\sum_{i=1}^{n}(\bar{e} - ei)^2} \quad (4)$$

$$RRMSE = \frac{1}{|\bar{e}|}\sqrt{\frac{\sum_{i=1}^{n}(ei - mi)^2}{n}} \quad (5)$$

$$R = \frac{\sum_{i=1}^{n}(ei - \bar{e}i)(mi - \overline{mi})}{\sqrt{\sum_{i=1}^{n}(ei - \bar{e}i)^2 \sum_{i=1}^{n}(mi - \overline{mi})^2}} \quad (6)$$

$$\rho = \frac{RRMSE}{1 + R} \quad (7)$$

where Fc is the compressive strength, β1 to βn is the regression coefficients, α is the regression constant, $£_{1\ is}$ the error term

2.21. Artificial neural network

The compressive and flexural strength of the mixed design specimens will be measured. Following casting for 7, 14, 21, and 28 days, all mix design samples will be examined. Compressive strength will be determined using cube samples, and flexural strength will be determined using beam specimens. After 28 days of casting, cube samples will be weighed to determine their density [68-71]. The flow chart for the ANN is shown in Figure 4.

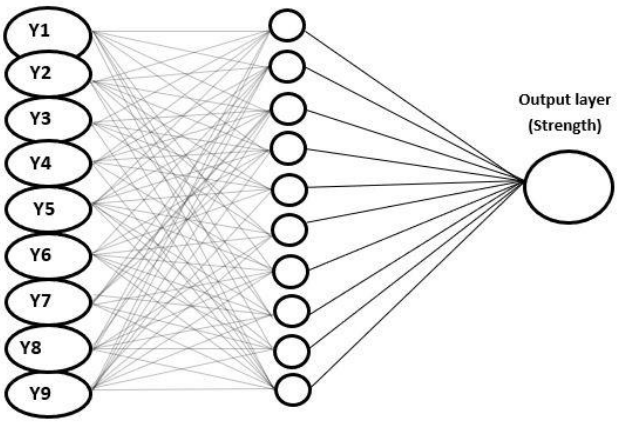


Figure 4. ANN architecture.

Reference [72-75], A evaluated the feasibility of MARS and GEP models in predicting the compressive strength of SCC at 28 days. In their research, values of compressive strength were estimated using multivariate adaptive regression spline approach [76-81]. The research indicated high ability of both the GEP and MARS models in predicting SCC strength. The parameters selected for

the GEP models were defined as shown in Table 1. From the work, predictive models were developed and the model tree shown in Figure 5 while Figure 6 represent the scatter plot for training of the GEP models and Figure 7 shows the scatter plot for testing of the GEP models.

Table 1.: Parameters of optimized GEP models [80].

Parameters	Description of parameters	Setting of parameters
P1	Function set	+,-, x,l,exp, power
P2	Mutation rate	0.138
P3	Inversion rate	0.546
P4	One point and two point recombination rate	0.277
P5	Gene recombination rate	0.277
P6	Gene tramsportatiom rate	0.277
P7	Maximum tree depth	6
P8	Number of genes	3
P9	Number of chromosones	30

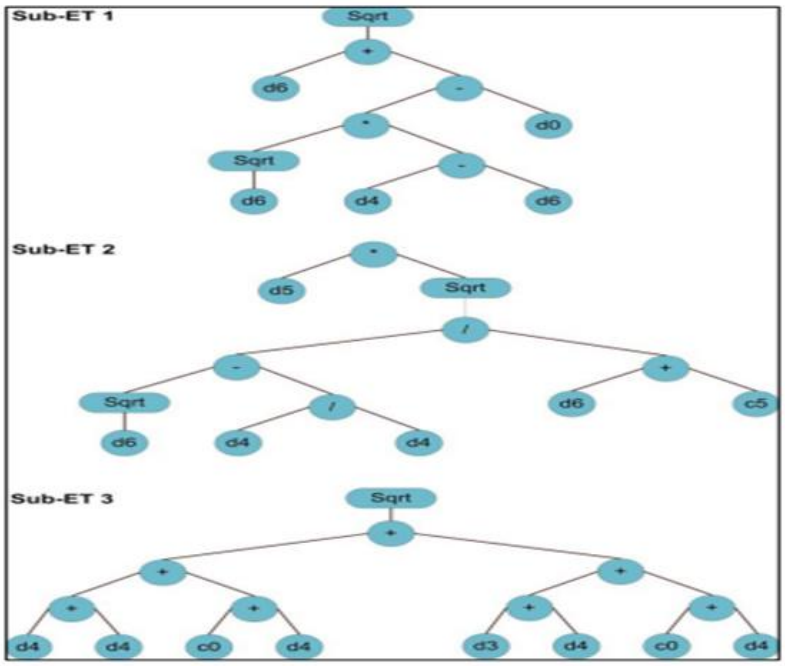


Figure 5. GEP models in concrete strength prediction [80].

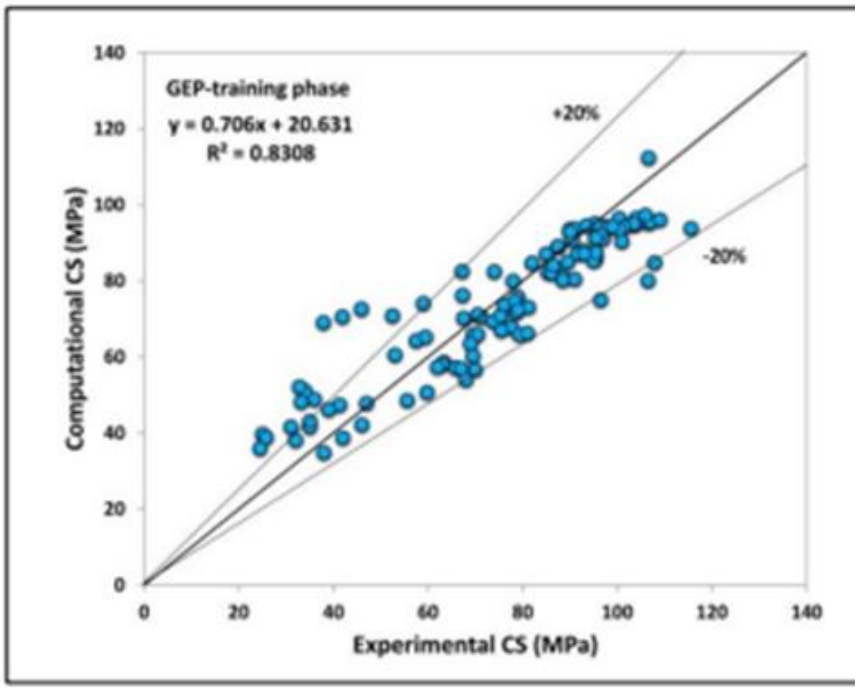


Figure 6. Scatter plot of observed and predicted compressive strength for training of GEP model [80].

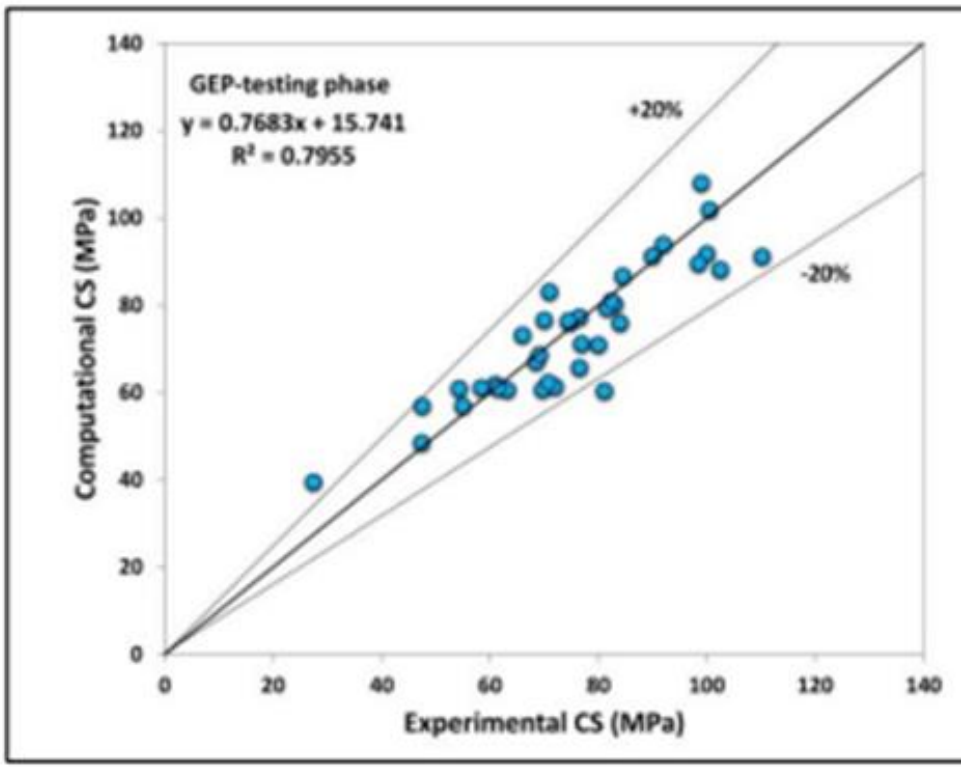


Figure 7. Scatter plot of observed and predicted compressive strength for testing of GEP model [80].

3. Conclusion

From the findings of this work, it was observed that use of artificial intelligence techniques in predicting the compressive strength of self-compacting concrete give approximate strength values close to experimental investigations. The research has shown that if a model gives R^2 > 0.8, then there is a correlation between predicted and experimental values for the available data in the dataset. The review observed that all the techniques studied, R^2 was always greater than 0.8 for 28 days' compressive strength which proves both of the models can be used for accurate prediction. The ANN model was identified to possesses a high degree of steadiness when compared to experimental results of concrete compressive strength, hence, ANN is a strong predictive tool, for both in situ and experimental prediction and as such is recommended for formulation of many civil engineering properties that requires predictions. Much time and resources are saved with artificial intelligence models as it eliminates the need for experimental test which sometimes delay construction works. Reinforcement learning techniques like Deep Deterministic Policy Gradient (DDPG) are more responsive to selections of hyper parameters than others. Through the use of hyper parameter optimization, a tuple of hyper parameters can produce an optimal model on test data that minimizes a predetermined loss function. Overfitting occurs in machine learning when a model performs poorly on new or unseen data because it is too closely aligned with the training data that was used for the model but over fitting can be avoided through, cross validation, augmentations, drop out and feature selection techniques.

Acknowledgement

The authors appreciate the staff of the Department of Civil Engineering, University of Nigeria Nsukka for considering this work for a post graduate program, we remain grateful for all your recommendations to improve this work.

Author contributions

Sesugh Terlumun: Conceptualization, Methodology, Writing-original draft preparation of the manuscript
Onyia Michael: Visualization, Writing-Reviewing and Editing
Okafor Fidelis: Validation, Writing-Reviewing and Editing

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Gaimster, R., & Dixon, N. (2003). Self-compacting concrete. Advanced Concrete Technology, 3, 1-23. https://doi.org/10.1016/B978-075065686-3/50295-0
2. Falliano, D., De Domenico, D., Ricciardi, G., & Gugliandolo, E. (2018). Experimental investigation on the compressive strength of foamed concrete: Effect of curing conditions, cement type, foaming agent and dry density. Construction and Building Materials, 165, 735-749. https://doi.org/10.1016/j.conbuildmat.2017.12.241

3. Lee, S. C. (2003). Prediction of concrete strength using artificial neural networks. Engineering Structures, 25(7), 849-857. https://doi.org/10.1016/S0141-0296(03)00004-X
4. Madandoust, R., & Mousavi, S. Y. (2012). Fresh and hardened properties of self-compacting concrete containing metakaolin. Construction and Building Materials, 35, 752-760. https://doi.org/10.1016/j.conbuildmat.2012.04.109
5. Tufail, R. F., Naeem, M. H., Ahmad, J., Waheed, H., Majdi, A., Farooq, D., ... & Butt, F. (2022). Evaluation of the fresh and mechanical properties of nano-engineered self compacting concrete containing graphite nano/micro platelets. Case Studies in Construction Materials, 17, e01165. https://doi.org/10.1016/j.cscm.2022.e01165
6. Shi, C., Wu, Z., Lv, K., & Wu, L. (2015). A review on mixture design methods for self-compacting concrete. Construction and Building Materials, 84, 387-398. https://doi.org/10.1016/j.conbuildmat.2015.03.079
7. Hamada, H., Alattar, A., Tayeh, B., Yahaya, F., & Thomas, B. (2022). Effect of recycled waste glass on the properties of high-performance concrete: A critical review. Case Studies in Construction Materials, 17, e01149. https://doi.org/10.1016/j.cscm.2022.e01149
8. Efnarc, S. (2002). Guidelines for Self-Compacting Concrete, Rep. from EFNARC. 44, 32.
9. Tejaswini, G. L. S., & Rao, A. V. (2020). A detailed report on various behavioral aspects of self-compacting concrete. Materials Today: Proceedings, 33, 839-844. https://doi.org/10.1016/j.matpr.2020.06.273
10. Danish, P., & Ganesh, G. M. (2021). Self-compacting concrete—optimization of mix design procedure by the modifications of rational method. In 3rd International Conference on Innovative Technologies for Clean and Sustainable Development: ITCSD 2020 3, 369-396. https://doi.org/10.1007/978-3-030-51485-3_25
11. Esmaeilkhanian, B., Khayat, K. H., Yahia, A., & Feys, D. (2014). Effects of mix design parameters and rheological properties on dynamic stability of self-consolidating concrete. Cement and Concrete Composites, 54, 21-28. https://doi.org/10.1016/j.cemconcomp.2014.03.001
12. Ashish, D. K., & Verma, S. K. (2019). An overview on mixture design of self-compacting concrete. Structural Concrete, 20(1), 371-395. https://doi.org/10.1002/suco.201700279
13. Bayer, İ. R., Turanlı, L., & Mehta, P. K. (2019). Mass concrete construction using self-compacting mortar. Turkish Journal of Engineering, 3(3), 110-119. https://doi.org/10.31127/tuje.462548
14. Kandiri, A., Golafshani, E. M., & Behnood, A. (2020). Estimation of the compressive strength of concretes containing ground granulated blast furnace slag using hybridized multi-objective ANN and salp swarm algorithm. Construction and Building Materials, 248, 118676. https://doi.org/10.1016/j.conbuildmat.2020.118676
15. Golafshani, E. M., Rahai, A., & Sebt, M. H. (2015). Artificial neural network and genetic programming for predicting the bond strength of GFRP bars in concrete. Materials and Structures, 48, 1581-1602. https://doi.org/10.1617/s11527-014-0256-0
16. Gesoğlu, M., Güneyisi, E., Özturan, T., & Özbay, E. (2010). Modeling the mechanical properties of rubberized concretes by neural network and genetic programming. Materials and Structures, 43, 31-45. https://doi.org/10.1617/s11527-009-9468-0
17. Belalia Douma, O., Boukhatem, B., Ghrici, M., & Tagnit-Hamou, A. (2017). Prediction of properties of self-compacting concrete containing fly ash using artificial neural network. Neural Computing and Applications, 28, 707-718. https://doi.org/10.1007/s00521-016-2368-7
18. Siddique, R., Aggarwal, P., & Aggarwal, Y. (2011). Prediction of compressive strength of self-compacting concrete containing bottom ash using artificial neural networks. Advances in Engineering software, 42(10), 780-786. https://doi.org/10.1016/j.advengsoft.2011.05.016
19. Behnood, A., & Golafshani, E. M. (2018). Predicting the compressive strength of silica fume concrete using hybrid artificial neural network with multi-objective grey wolves. Journal of Cleaner Production, 202, 54-64. https://doi.org/10.1016/j.jclepro.2018.08.065
20. Chakravarthy HG, N., Seenappa, K. M., Naganna, S. R., & Pruthviraja, D. (2023). Machine Learning Models for the Prediction of the Compressive Strength of Self-Compacting Concrete Incorporating Incinerated Bio-Medical Waste Ash. Sustainability, 15(18), 13621. https://doi.org/10.3390/su151813621
21. Behnood, A., Behnood, V., Gharehveran, M. M., & Alyamac, K. E. (2017). Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm. Construction and Building Materials, 142, 199-207. https://doi.org/10.1016/j.conbuildmat.2017.03.061
22. Kovačević, M., Lozančić, S., Nyarko, E. K., & Hadzima-Nyarko, M. (2021). Modeling of compressive strength of self-compacting rubberized concrete using machine learning. Materials, 14(15), 4346. https://doi.org/10.3390/ma14154346
23. Aiyer, B. G., Kim, D., Karingattikkal, N., Samui, P., & Rao, P. R. (2014). Prediction of compressive strength of self-compacting concrete using least square support vector machine and relevance vector machine. KSCE Journal of Civil Engineering, 18, 1753-1758. https://doi.org/10.1007/s12205-014-0524-0
24. Golafshani, E. M., Behnood, A., & Arashpour, M. (2020). Predicting the compressive strength of normal and High-Performance Concretes using ANN and ANFIS hybridized with Grey Wolf Optimizer. Construction and Building Materials, 232, 117266. https://doi.org/10.1016/j.conbuildmat.2019.117266
25. Ikeagwuani, C. C. (2021). Estimation of modified expansive soil CBR with multivariate adaptive regression splines, random forest and gradient

boosting machine. Innovative Infrastructure Solutions, 6(4), 199. https://doi.org/10.1007/s41062-021-00568-z
26. Onyia, M. E., Ambrose, E. E., Okafor, F. O., & Udo, J. J. (2023). Mathematical modelling of compressive strength of recycled ceramic tile aggregate concrete using modified regression theory. Journal of Applied Sciences and Environmental Management, 27(1), 33-42. https://doi.org/10.4314/jasem.v27i1.6
27. Aicha, M. B., Al Asri, Y., Zaher, M., Alaoui, A. H., & Burtschell, Y. (2022). Prediction of rheological behavior of self-compacting concrete by multi-variable regression and artificial neural networks. Powder Technology, 401, 117345. https://doi.org/10.1016/j.powtec.2022.117345
28. Patel, R., Hossain, K. M. A., Shehata, M., Bouzoubaa, N., & Lachemi, M. (2004). Development of statistical models for mixture design of high-volume fly ash self-consolidating concrete. Materials Journal, 101(4), 294-302. https://doi.org/10.14359/13363
29. Razavi Tosee, S. V., & Nikoo, M. (2019). Neuro-fuzzy systems in determining light weight concrete strength. Journal of Central South University, 26(10), 2906-2914. https://doi.org/10.1007/s11771-019-4223-3
30. Günal, A. Y., & Mehdi, R. (2024). Application of a new fuzzy logic model known as" SMRGT" for estimating flow coefficient rate. Turkish Journal of Engineering, 8(1), 46-55. https://doi.org/10.31127/tuje.1225795
31. Zhou, Q., Wang, F., & Zhu, F. (2016). Estimation of compressive strength of hollow concrete masonry prisms using artificial neural networks and adaptive neuro-fuzzy inference systems. Construction and Building Materials, 125, 417-426. https://doi.org/10.1016/j.conbuildmat.2016.08.064
32. Behnood, A., & Golafshani, E. M. (2020). Machine learning study of the mechanical properties of concretes containing waste foundry sand. Construction and Building Materials, 243, 118152. https://doi.org/10.1016/j.conbuildmat.2020.118152
33. Madani, H., Kooshafar, M., & Emadi, M. (2020). Compressive strength prediction of nanosilica-incorporated cement mixtures using adaptive neuro-fuzzy inference system and artificial neural network models. Practice Periodical on Structural Design and Construction, 25(3), 04020021. https://doi.org/10.1061/(ASCE)SC.1943-5576.0000499
34. Chiew, F. H., Ng, C. K., Chai, K. C., & Tay, K. M. (2017). A fuzzy adaptive resonance theory-based model for mix proportion estimation of high-performance concrete. Computer-Aided Civil and Infrastructure Engineering, 32(9), 772-786. https://doi.org/10.1111/mice.12288
35. Lee, S. C. (2003). Prediction of concrete strength using artificial neural networks. Engineering structures, 25(7), 849-857. https://doi.org/10.1016/S0141-0296(03)00004-X
36. Lai, S., & Serra, M. (1997). Concrete strength prediction by means of neural network. Construction and Building Materials, 11(2), 93-98. https://doi.org/10.1016/S0950-0618(97)00007-X
37. Ibrahim, S. M., Ansari, S. S., & Hasan, S. D. (2023). Towards white box modeling of compressive strength of sustainable ternary cement concrete using explainable artificial intelligence (XAI). Applied Soft Computing, 149, 110997. https://doi.org/10.1016/j.asoc.2023.110997
38. Erzin, Y. (2007). Artificial neural networks approach for swell pressure versus soil suction behaviour. Canadian Geotechnical Journal, 44(10), 1215-1223. https://doi.org/10.1139/T07-052
39. Silva, F. A., Delgado, J. M., Cavalcanti, R. S., Azevedo, A. C., Guimarães, A. S., & Lima, A. G. (2021). Use of nondestructive testing of ultrasound and artificial neural networks to estimate compressive strength of concrete. Buildings, 11(2), 44. https://doi.org/10.3390/buildings11020044
40. Duan, Z. H., Kou, S. C., & Poon, C. S. (2013). Prediction of compressive strength of recycled aggregate concrete using artificial neural networks. Construction and Building Materials, 40, 1200-1206. https://doi.org/10.1016/j.conbuildmat.2012.04.063
41. Alade, I. O., Bagudu, A., Oyehan, T. A., Abd Rahman, M. A., Saleh, T. A., & Olatunji, S. O. (2018). Estimating the refractive index of oxygenated and deoxygenated hemoglobin using genetic algorithm–support vector regression model. Computer Methods and Programs in Biomedicine, 163, 135-142. https://doi.org/10.1016/j.cmpb.2018.05.029
42. Iqbal, M. F., Liu, Q. F., Azim, I., Zhu, X., Yang, J., Javed, M. F., & Rauf, M. (2020). Prediction of mechanical properties of green concrete incorporating waste foundry sand based on gene expression programming. Journal of Hazardous Materials, 384, 121322. https://doi.org/10.1016/j.jhazmat.2019.121322
43. Sarıdemir, M. (2010). Genetic programming approach for prediction of compressive strength of concretes containing rice husk ash. Construction and Building Materials, 24(10), 1911-1919. https://doi.org/10.1016/j.conbuildmat.2010.04.011
44. Gandomi, A. H., & Roke, D. A. (2015). Assessment of artificial neural network and genetic programming as predictive tools. Advances in Engineering Software, 88, 63-72. https://doi.org/10.1016/j.advengsoft.2015.05.007
45. Khan, M. A., Zafar, A., Akbar, A., Javed, M. F., & Mosavi, A. (2021). Application of Gene Expression Programming (GEP) for the prediction of compressive strength of geopolymer concrete. Materials, 14(5), 1106. https://doi.org/10.3390/ma14051106
46. Faradonbeh, R. S., Hasanipanah, M., Amnieh, H. B., Armaghani, D. J., & Monjezi, M. (2018). Development of GP and GEP models to estimate an environmental issue induced by blasting operation. Environmental Monitoring and Assessment, 190, 1-15. https://doi.org/10.1007/s10661-018-6719-y
47. Khandelwal, M., Shirani Faradonbeh, R., Monjezi, M., Armaghani, D. J., Majid, M. Z. B. A., & Yagiz, S. (2017). Function development for appraising brittleness of intact rocks using genetic programming and non-

linear multiple regression models. Engineering with Computers, 33, 13-21. https://doi.org/10.1007/s00366-016-0452-3
48. Ferreira, C. (2002). Gene expression programming in problem solving. Soft Computing and Industry: Recent Applications, 635-653. https://doi.org/10.1007/978-1-4471-0123-9_54
49. Ferreira, C. (2006). Gene expression programming: mathematical modeling by an artificial intelligence, 21. Springer.
50. Gholampour, A., Gandomi, A. H., & Ozbakkaloglu, T. (2017). New formulations for mechanical properties of recycled aggregate concrete using gene expression programming. Construction and Building Materials, 130, 122-145. https://doi.org/10.1016/j.conbuildmat.2016.10.114
51. Asteris, P. G., Skentou, A. D., Bardhan, A., Samui, P., & Pilakoutas, K. (2021). Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models. Cement and Concrete Research, 145, 106449. https://doi.org/10.1016/j.cemconres.2021.106449
52. Neira, P., Bennun, L., Pradena, M., & Gomez, J. (2020). Predviđanje tlačne čvrstoće betona pomoću umjetnih neuronskih mreža. Građevinar, 72(07), 585-592. https://doi.org/10.14256/JCE.2438.2018
53. Mehmannavaz, T., Khalilikhorram, V., Sajjadi, S. M., & Samadi, M. (2014). Presenting an Appropriate Neural Network for Optimal Mix Design of Roller Compacted Concrete Dams. Research Journal of Applied Sciences, Engineering and Technology, 7(9), 1872-1877. https://doi.org/10.19026/rjaset.7.475
54. Khan, M. I. (2012). Predicting properties of high performance concrete containing composite cementitious materials using artificial neural networks. Automation in Construction, 22, 516-524. https://doi.org/10.1016/j.autcon.2011.11.011
55. Onyelowe, K. C., Iqbal, M., Jalal, F. E., Onyia, M. E., & Onuoha, I. C. (2021). Application of 3-algorithm ANN programming to predict the strength performance of hydrated-lime activated rice husk ash treated soil. Multiscale and Multidisciplinary Modeling, Experiments and Design, 4, 259-274. https://doi.org/10.1007/s41939-021-00093-7
56. Prasad, B. R., Eskandari, H., & Reddy, B. V. (2009). Prediction of compressive strength of SCC and HPC with high volume fly ash using ANN. Construction and Building Materials, 23(1), 117-128. https://doi.org/10.1016/j.conbuildmat.2008.01.014
57. Nguyen, T., Kashani, A., Ngo, T., & Bordas, S. (2019). Deep neural network with high-order neuron for the prediction of foamed concrete strength. Computer-Aided Civil and Infrastructure Engineering, 34(4), 316-332. https://doi.org/10.1111/mice.12422
58. Dias, W. P. S., & Pooliyadda, S. P. (2001). Neural networks for predicting properties of concretes with admixtures. Construction and Building Materials, 15(7), 371-379. https://doi.org/10.1016/S0950-0618(01)00006-X
59. Abunassar, N., Alas, M., & Ali, S. I. A. (2023). Prediction of compressive strength in self-compacting concrete containing fly ash and silica fume using ANN and SVM. Arabian Journal for Science and Engineering, 48(4), 5171-5184. https://doi.org/10.1007/s13369-022-07359-3
60. Yaman, M. A., Abd Elaty, M., & Taman, M. (2017). Predicting the ingredients of self compacting concrete using artificial neural network. Alexandria Engineering Journal, 56(4), 523-532. https://doi.org/10.1016/j.aej.2017.04.007
61. Belalia Douma, O., Boukhatem, B., Ghrici, M., & Tagnit-Hamou, A. (2017). Prediction of properties of self-compacting concrete containing fly ash using artificial neural network. Neural Computing and Applications, 28, 707-718. https://doi.org/10.1007/s00521-016-2368-7
62. Golafshani, E. M., Behnood, A., & Arashpour, M. (2020). Predicting the compressive strength of normal and High-Performance Concretes using ANN and ANFIS hybridized with Grey Wolf Optimizer. Construction and Building Materials, 232, 117266. https://doi.org/10.1016/j.conbuildmat.2019.117266
63. Demir, V., & Doğu, R. (2024). Prediction of elevation points using three different heuristic regression techniques. Turkish Journal of Engineering, 8(1), 56-64. https://doi.org/10.31127/tuje.1257847
64. Salami, B. A., Iqbal, M., Abdulraheem, A., Jalal, F. E., Alimi, W., Jamal, A., ... & Bardhan, A. (2022). Estimating compressive strength of lightweight foamed concrete using neural, genetic and ensemble machine learning approaches. Cement and Concrete Composites, 133, 104721. https://doi.org/10.1016/j.cemconcomp.2022.104721
65. Shariati, M., Armaghani, D. J., Khandelwal, M., Zhou, J., & Khorami, M. (2021). Assessment of longstanding effects of fly ash and silica fume on the compressive strength of concrete using extreme learning machine and artificial neural network. Journal of Advanced Engineering and Computation, 5(1), 50-74. http://dx.doi.org/10.25073/jaec.202151.308
66. Mogaraju, J. K. (2024). Machine learning empowered prediction of geolocation using groundwater quality variables over YSR district of India. Turkish Journal of Engineering, 8(1), 31-45. https://doi.org/10.31127/tuje.1223779
67. Acı, M., Acı, Ç. İ., & Avcı, M. (2018). Performance comparison of ANFIS, ANN, SVR, CART AND MLR techniques for geometry optimization of carbon nanotubes using CASTEP. Turkish Journal of Engineering, 2(3), 119-124. https://doi.org/10.31127/tuje.408976
68. Kiani, B., Gandomi, A. H., Sajedi, S., & Liang, R. Y. (2016). New formulation of compressive strength of preformed-foam cellular concrete: an evolutionary approach. Journal of Materials in Civil Engineering, 28(10), 04016092. https://doi.org/10.1061/(ASCE)MT.1943-5533.0001602

69. Nehdi, M., Djebbar, Y., & Khan, A. J. M. J. (2001). Neural network model for preformed-foam cellular concrete. Materials Journal, 98(5), 402-409.
70. Othman, M. M. (2023). Modeling of daily groundwater level using deep learning neural networks. Turkish Journal of Engineering, 7(4), 331-337. https://doi.org/10.31127/tuje.1169908
71. Tiza, M. T., Ogunleye, E., Jiya, V. H., Onuzulike, C., Akande, E. O., & Terlumun, S. (2023). Integrating Sustainability into Civil Engineering and the Construction Industry. Journal of Cement Based Composites, 4(1), 1-11. https://doi.org/10.36937/cebacom.2023.5756
72. Algaifi, H. A., Bakar, S. A., Alyousef, R., Sam, A. R. M., Alqarni, A. S., Ibrahim, M., ... & Salami, B. A. (2021). Machine learning and RSM models for prediction of compressive strength of smart bio-concrete. Smart Structural System, 28, 535-551. https://doi.org/10.12989/sss.2021.28.4.535
73. Chou, J. S., & Pham, A. D. (2013). Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength. Construction and Building Materials, 49, 554-563. https://doi.org/10.1016/j.conbuildmat.2013.08.078
74. Imam, A., Salami, B. A., & Oyehan, T. A. (2021). Predicting the compressive strength of a quaternary blend concrete using Bayesian regularized neural network. Journal of Structural Integrity and Maintenance, 6(4), 237-246. https://doi.org/10.1080/24705314.2021.1892572
75. Ly, H. B., Nguyen, M. H., & Pham, B. T. (2021). Metaheuristic optimization of Levenberg–Marquardt-based artificial neural network using particle swarm optimization for prediction of foamed concrete compressive strength. Neural Computing and Applications, 33(24), 17331-17351. https://doi.org/10.1007/s00521-021-06321-y
76. Shariati, M., Mafipour, M. S., Mehrabi, P., Bahadori, A., Zandi, Y., Salih, M. N., ... & Poi-Ngian, S. (2019). Application of a hybrid artificial neural network-particle swarm optimization (ANN-PSO) model in behavior prediction of channel shear connectors embedded in normal and high-strength concrete. Applied Sciences, 9(24), 5534. https://doi.org/10.3390/app9245534
77. Shariati, M., Mafipour, M. S., Mehrabi, P., Shariati, A., Toghroli, A., Trung, N. T., & Salih, M. N. (2021). A novel approach to predict shear strength of tilted angle connectors using artificial intelligence techniques. Engineering with Computers, 37, 2089-2109. https://doi.org/10.1007/s00366-019-00930-x
78. Shariati, M., Mafipour, M. S., Mehrabi, P., Ahmadi, M., Wakil, K., Trung, N. T., & Toghroli, A. (2020). Prediction of concrete strength in presence of furnace slag and fly ash using Hybrid ANN-GA (Artificial Neural Network-Genetic Algorithm). Smart Structures and Systems, An International Journal, 25(2), 183-195. https://doi.org/10.12989/sss.2020.25.2.183
79. Pham, A. D., Ngo, N. T., Nguyen, Q. T., & Truong, N. S. (2020). Hybrid machine learning for predicting strength of sustainable concrete. Soft Computing, 24(19), 14965-14980. https://doi.org/10.1007/s00500-020-04848-1
80. Babajanzadeh, M., & Azizifar, V. (2018). Compressive strength prediction of self-compacting concrete incorporating silica fume using artificial intelligence methods. Civil Engineering Journal, 4(7), 1542-1552. http://dx.doi.org/10.28991/cej-0309193
81. Özcan, F., Atiş, C. D., Karahan, O., Uncuoğlu, E., & Tanyildizi, H. (2009). Comparison of artificial neural network and fuzzy logic models for prediction of long-term compressive strength of silica fume concrete. Advances in Engineering Software, 40(9), 856-863. https://doi.org/10.1016/j.advengsoft.2009.01.005

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Comparative study on obtaining paper and thread-based microfluidics via simple fabrication techniques

Nagihan Okutan Arslan [*1], **Ragheid Mohammed Helmy Atta** [2], **Levent Trabzon** [3]

[1] *Istanbul Technical University, Department of Nanoscience and Nanoengineering, Türkiye, nokutan@itu.edu.tr*
[2] *Taibah University, Department of Electrical Engineering, Kingdom of Saudi Arabia, ratta@taibahu.edu.sa*
[3] *Istanbul Technical University, Department of Manufacturing Engineering, Türkiye, levent.trabzon@gmail.com*

Cite this study: Okutan Arslan, N., Atta, R. M. H., & Trabzon, L. (2024). Comparative study on obtaining paper and thread-based microfluidics via simple fabrication techniques. Turkish Journal of Engineering, 8 (3), 551-562

https://doi.org/10.31127/tuje.1432125

Keywords
Microfluidic fabrication techniques
µPAD
µTAD
Organic solvent compatibility

Research Article

Received: 05.02.2024
Revised: 22.02.2024
Accepted: 01.03.2024
Published: 13.07.2024

Abstract

Microfluidic paper-based analytical devices (µPADs) and microfluidic thread-based analytical devices (µTADs) have recently been introduced as a new class of on-site monitoring devices. Creating hydrophilic channels with hydrophobic barriers on papers/threads produces µPADs/µTADs. Fabrication is a crucial step in creating durable µPADs/µTADs that can withstand various liquids and impact the device's performance. Fabrication materials with distinct physicochemical properties allow microfluidic systems with sophisticated functions to be customized for specific applications. We present flexible and low-cost fabrication methods for µPAD and µTAD platforms. Platform designs and fabrications were implemented using a trial-and-error method for various designs with varying parameters. All production methods presented in the method section were used in µPAD production. For comparison studies, only the dipping method was used in µTAD production due to its ease of application. In this study, we tried to reveal the strengths and weaknesses of the production techniques and the resulting microfluidic platforms. A leaching test was performed with water solutions containing red ink. The compatibility of the hydrophobic walls of the platforms was tested with several solvents (isopropanol, methanol, and acetone), deionized (DI) water, and phosphate buffer solution PBS and compared. Patterning paper with polydimethylsiloxane (PDMS), white glue, alkyl ketene dimer (AKD), beeswax, and paraffin are much more flexible and simpler than traditional photoresist-based fabrications. The advantages and disadvantages of fabrication techniques; solvent resistance and wicking behaviors of platforms were discussed in the last part. The fabricated microfluidic platforms can be functionalized and used in many areas where analytical tests are applied. Studies on diversifying channel geometries and increasing resolution need to be continued. It should be investigated which devices can be used to obtain qualitative and quantitative results. To make simple and cheap production techniques suitable for mass production, studies should be carried out from different branches.

1. Introduction

Recent advances in material science have resulted in the convergence of numerous sciences and technological fields, including engineering, chemistry, physics, biology, and medicine. Microanalysis systems have further transformed these fields [1]. The reduction to the microchannel scale not only miniaturizes and integrates the analytical device but also exerts many micro and nano effects on the fluid, endowing microfluidic chips with greatly improved performance compared to that of a traditional analysis system [2]. At the microscale, various forces exert a greater influence compared to those encountered in everyday life. The effects that gain prominence in microfluidics encompass laminar flow, diffusion, fluidic resistance, surface area to volume ratio, and surface tension [3]. Microfluidics was defined as the integration of basic operating units such as sample preparation, biological and chemical reactions, separation, and detection to complete various biological or chemical reaction processes and analyze their products [4] allowing for precise control of micro-nano

liquids [5]. Scientists have proposed microfluidic systems as a method for developing new tools [6] for various fields [7]. Microfluidics offers many advantages over traditional systems [8] due to the benefits of high sensitivity, rapid detection, compact size, low reagent consumption, and wastage, low cost, [7] assay automation, enriched detection performance, user-friendliness [9] faster processing times, and improved analytical performance [8].

The evolution of microscale devices from closed to open device embodiments, including open microfluidic capillary systems [10]. Among microfluidics, self-driven microfluidics typically employs surface hydrophilic properties or capillary forces for fluid transport and handling. It is distinguished by its self-driven nature, and the absence of an additional pumping and energy source [11]. Paper-based microfluidics are devices made of paper or other porous membranes that use capillary action to manipulate small (10^{-6} to 10^{-9} L) volumes of fluids [12]. Low Reynolds numbers (Re) in microfluidic devices ($10^{-3} > Re > 10^{-5}$) result in laminar flow, which allows effective control of critical reaction parameter such as temperature, mixing velocity, and reagent concentration [13].

µPADs first described by Whitesides' group [14], are sustainable alternatives to traditional polymer or glass-based microfluidic devices [15]. In comparison to regular microfluidic devices, µPADs have flexibility, hydrophilicity, lightweight, and biocompatibility, and can be operated without pump power to transport fluids of small size [16]. µPADs have recently been introduced as a new class of on-site monitoring devices specifically for use in developing countries [17]. To carry out the desired physical, chemical, and biological processing, multiple components, such as reaction vessels, mixing structures, or separation chambers, can be patterned in an integrated manner onto a piece of paper [18]. Since their introduction, they've been widely used in a variety of fields, including food safety, cell culture, and point-of-care (POC) applications [19].

Since its invention, paper has played an important role in everything from writing scriptures to modern-day chemistry [20]. Paper was first introduced as a potential material for microfluidics in 1949 [21] when a heated dye was used to impregnate a filter paper with a wax barrier, the goal was to create a restricted flowing channel to reduce the amount of sample used [22]. The paper was widely used due to its unique fiber structure, porosity, capillary force drive, absorbency, air permeability [23,24], and sustainability [25]. The biological function of cellulose, as well as its numerous applications, are based on its distinct fiber morphology. Elementary fibrils, microfibrils, and microfibrillar bands define the morphological hierarchy [26]. Paper's porosity is caused by spaces between the fibers, uncollapsed fiber lumens, and the inherent porosity of the fiber walls [27]. Paper surface modification is typically accomplished by fabricating µPADs through chemical modification, physical deposition, or physical blocking of its pores to tune the paper material to have the desired properties for direct use in critical experiments/applications [28]. These agents alter the paper's liquid-wetting properties, allowing the formation of hydrophilic-hydrophobic patterns [29]. Creating hydrophilic channels with hydrophobic barriers on papers results in µPADs [30]. The two parallel hydrophobic lines function as channels because the hydrophilic sample solution cannot penetrate the hydrophobic line or barrier, causing the liquid to flow in the channels due to capillary action [31] also known as wicking in contrast to traditional silicone-based microfluidic chips [32]. Capillary action is caused by intermolecular forces between the solution and the surrounding porous material surfaces and is unaffected by gravity, inertia, or any other external forces [33].

There are two fundamental ideas for creating open channels: the first is to confine liquid by designing a solid structure, such as grooves, suspended microfluidics, and hanging droplets. The second step is to modify the substrate to be hydrophilic/hydrophobic, and then confine the aqueous samples at the hydrophilic pattern [34]. A µPAD device consists of a detection zone, a sample zone, and a flowing channel that transports the sample solution from the sample zone to the detection zone [35]. The most common techniques for defining barriers and zones are cutting, printing, drawing, dip-coating, plotting, and photolithography (Figure 1) [36]. While most reports used photolithography to create barriers, newer, simpler, faster, and less expensive technologies have emerged, enabling implementation in research groups with limited resources [37]. Recent advancements in fabrication methods have concentrated on reducing fabrication time, cost, and complexity, increasing spatial resolution, and shrinking device sizes [38]. Fabrication is a critical step in the development of effective µPADs that are resistant to a variety of liquids [39] and strongly influence the final device's performance [40]. Fabrication materials with distinct physicochemical properties enable microfluidic systems with sophisticated functions to be tailored to specific applications [41].

Aside from paper, threads of cotton, nylon, silk, polyester, or other fibers have been proposed as substrates for microfluidic diagnostic systems, also known as µTADs [42]. They are more flexible and wearable alternatives [43] with higher mechanical strengths under wet conditions [19]. In addition to similar advantages provided by paper threads have varying wicking and adsorptive properties to meet the needs of various research needs [44].

Threads have stretchability, high mechanical strength, and durability, making them an ideal substrate for the development of wearable analytical devices [46].

Detailed studies must be done about substrate properties, wicking rate, material durability, and fabrication methods. Here we present flexible and low-cost fabrication methods to obtain µPAD and µTAD platforms. All production methods presented in the method section were used in µPAD production. For comparison studies, only the dipping method was used in µTAD production due to its ease of application. The advantages and disadvantages of fabrication techniques; solvent resistance and wicking behaviors of platforms were discussed and summarized in the last part.

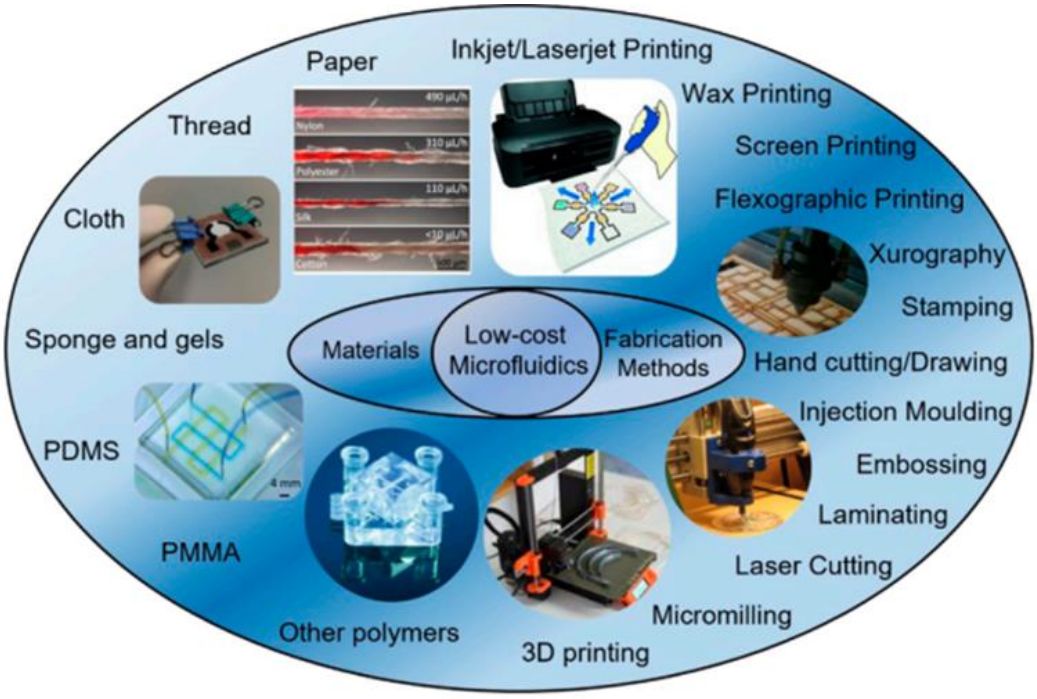


Figure 1. Microfluidic fabrication materials and techniques [45].

2. Materials and method

2.1. Materials

Coarse filter paper (ISOLAB); hexane, n-heptane, methanol, isopropanol, and acetone (MERCK); PBS (Sigma-Aldrich), were purchased from AS Kimya Ltd. PDMS and, PDMS curing agent (SYLGARD™ 184 Silicone Elastomer, DOW Corning) were purchased from C3 Teknoloji Ltd. White glue, paraffin, beeswax, pens, and magnets were purchased from local markets. AKD (Diper Chemical Company) was taken from the forest engineering department of Istanbul University. Evolon® (Freudenberg) was taken from the textile engineering department of Istanbul Technical University.

2.2. Fabrication methods

The filter paper was used for all µPAD platform fabrications. Platform designs and fabrications were implemented using a trial-and-error method for various designs with varying parameters. Evolon® fabric was also used for the dipping method to obtain the µTAD platform.

2.2.1. Creating channels and test zones with PDMS and white glue

In the first approach, hydrophilic channels or zones of interest are defined by drawing with hydrophobic inks. Commercial ballpoint pens were used to draw hydrophobic barriers on the filter papers. The tubes of pens were emptied and replaced with the prepared PDMS and white glue solutions. Various solutions were prepared with hexane during PDMS studies. PDMS was first used by dissolving it alone in hexane at ratios of 2:1 and 3:1. Solutions at the same ratio were also prepared after mixing PDMS and curing agent at a ratio of 10:1. After patterning, paper samples were kept at 70 °C for 1 h in an oven [47].

Another material used in the drawing technique is white glue. This substance is an emulsion of poly (vinyl acetate) (PVAc) in an aqueous solution. It was used undiluted and in 50% (v/v) dilution form. Drawings were made on single and double sides of the filter paper and the results were compared. Glue-patterned samples were kept at 130 °C for 1 h in an oven to obtain an insoluble structure. The glue-coated papers were then exposed to UV/Vis light (SUSS MicroTec Mask Aligner) at 800 W lamp power and 55-56 mW/cm^2 light intensity (30 min) for cross-linking to increase chemical resistance [37].

2.2.2. Creating channels and test zones with AKD

Patterns were created on a computer using the CATIA V5R21 software, then transferred to a printer and printed on filter papers. The printer was modified by replacing the ink cartridge with the AKD-heptane solution at ratios of 0.6, 2, and 4% (w/v). Printer cartridges were filled with AKD solutions and the designed channels were printed on A4-size filter papers. After evaporation of heptane paper samples were heated in an oven at 100 °C for 45 min to cure the AKD [48].

Another technique suitable for AKD is plasma treatment. The filter papers were hydrophobized using AKD-heptane solutions at the ratio of 0.6, 2, 4, 10 % (w/v). After evaporation of heptane, paper samples were heated in an oven at 100 °C for 45 min to cure the AKD and then placed into a vacuum O_2 plasma reactor (HARRICK PLASMA-Plasma Cleaner- USA). Heavy metallic masks were used to create test zone geometries. O_2 was set at 5 bar and taken into the plasma cabin for 3 minutes. Plasma was applied for 10 min at a high plasma level of the device (18W) [49].

2.2.3. Creating test zones with paraffin and beeswax

For the dipping method, paraffin and beeswax pellets were placed in a beaker and heated (75 °C) on a hotplate (IKA C-MAG HS7) until melted. Filter papers were cut into pieces, and round magnets were placed on both sides of the papers to hold them together temporarily. The assembly was then immersed for 1 second in a chamber of melted wax. Magnets were peeled off the paper after it had cooled to room temperature [50]. Since the dipping method can be easily applied to the threads, hydrophilic areas were created on the Evolon® for comparison purposes.

2.3. Evaluating the solvent compatibility of the hydrophobic walls

Various channel and test zone models have been designed and fabricated using various techniques to determine the suitable geometries. A leaching test was performed with water solutions containing red ink. The compatibility of the hydrophobic walls was tested with several solvents (isopropanol, methanol, and acetone), DI water, and PBS and compared. To observe leakage and understand the minimum volume required to fill the channels and test zones 5, 10, 20, 40, 60, 80, and 100 µL solvents were loaded on test zones and one end of the channels. In comparison studies with thread-based platforms, 200, 300, and 400 µL were also tested.

3. Results

Patterning is the first step to fabricate any paper or thread-based device [51]. Non-covalent cellulose treatment is based on filling the void spaces with a hydrophobic material (solid melt or polymer solution), creating a barrier to aqueous liquids [52].

The methods can be further divided into additive and subtractive methods [53], additive methods selectively hydrophobized only the barriers, while subtractive methods hydrophobized the entire paper before selectively restoring hydrophilicity to the channels [54]. The drawing, printing, and dipping methods used in this study can be shown as examples of additive methods. Plasma treatment is a subtractive method.

Photolithography, the first reported fabrication method, is known for its high resolution [55] but is laborious, time-consuming, has a high cost per device; and necessitates sophisticated instrumentation housed in expensive cleanrooms [56]. Alternative and more efficient methods and substrates are required for paper and thread-based microfluidics to achieve its true purpose. In this study, we proceeded with simple non-covalent patterning techniques. The first stage of the study is the production of various channels and the second stage is the test of the solvent resistance of these channels.

3.1. Drawing with PDMS and white glue

In addition to traditional fabrication methods, pen-on-paper (PoP) strategies, which are based on directly writing functional materials on paper with various pens, have emerged as simple and alternative approaches to creating hydrophobic barriers on paper substrates [57]. The first part of the study focuses on the fabrication of hand-drawn test zones (Figure 2). For comparison, two different materials, PDMS and scholar glue, were used in varying ratios. Enough pressure and weight, along with the pen nib, allow polymers to fully penetrate the filter paper. The solvent evaporates almost immediately, leaving the polymers on the paper to form the channels' hydrophobic walls. When we use PDMS alone, it does not plasticize and remains as a gel. Mixing with hexane improves the paper's absorption. As the hexane ratio increases, so does the ability to absorb; however, spread occurs. The addition of a curing agent improves plasticization, but because it dries faster, it may interfere with the paper's absorption. The mixture with the curing agent becomes unusable the following day; only PDMS can be used for a few days.

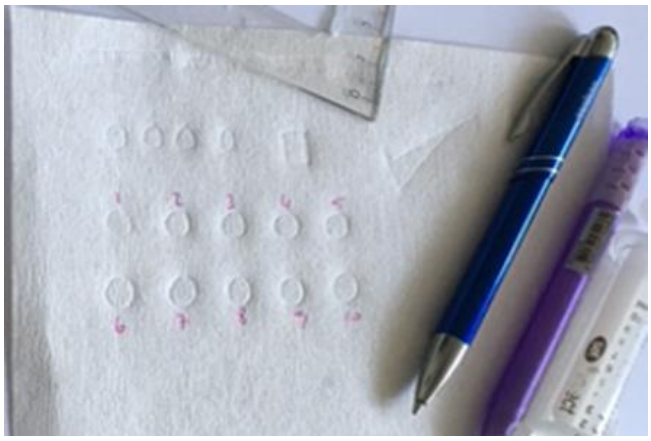


Figure 2. PDMS patterned filter paper.

Advantages of PDMS include its transparency, ease of availability, nontoxicity, odorlessness, and ease of dilution in hexanes or other organic solvents [58]. However, it required special preparation of PDMS diluted in hexanes and an extra heating step for polymerization after deposition. Because of its low solid-vapor interfacial free energy (21 erg/cm^2), PDMS can wet the entire depth of the paper. Solutions with higher viscosity (i.e., with greater than 80% PDMS (w/w) did not penetrate the paper; solutions with less than 67% PDMS (w/w) led to irregular lines [47]. The ink viscosity and tip pressure must be optimized to ensure that the hydrophobic ink penetrates the thickness of a piece of paper [59].

Another material used in the drawing method is white glue, which is cheaper and has a more viscous structure than PDMS. This PVAc-based glue was used both in nominative and in aqueous solution (50% (v/v) form. Although it is soluble in water, it becomes insoluble after baking at 130 °C. Exposing the UV light cross-linked the polymer and increased chemical resistance. Glue is attractive because it is inexpensive, globally available, non-toxic, and most importantly, does not require organic solvents [37]. Glue is percolated on both sides of the paper, which is required for the creation of a leak-proof device. To fabricate solvent and surfactant-

resistant structures, an additional heating step for polymerization and UV treatment is required.

PDMS drawings require the use of an organic solvent (hexane), which can damage the original structure of the filter paper [60]. While for white glue there is no such a problem. To take this technique to further levels, drawings were made with devices such as plotters and 3D pens.

Walia et al. [61] demonstrated a pen-plotter as a fast-writing tool filled with an inexpensive, aqueous phase ink composed of bovine serum albumin (BSA). Different concentrations such as 15-20-30% were tried and it was concluded that 20% concentration viscosity was suitable for printing BSA with a plotter. The proper viscosity of the used ink material helps block the paper pores effectively and form a good hydrophobic barrier. Nuchtavorn and Macka [62] present a novel, highly flexible, and low-cost fabrication method based on a desktop digital craft plotter. The method can print two or more water-resistant inks quickly, in a variety of designs, and without alignment issues. No additional heating stages are required when using this type of ink. Sousa et al. [63] described for the first time the use of a 3D pen to pattern papers with acrylonitrile butadiene styrene (ABS).

Using a ballpoint pen to define the test zones reduces the number of reagents used and waste, making it green engineering. This technique is inexpensive, but the pattern resolution is limited [59]. Also, it is almost impossible to produce devices on an upgraded scale and at a high speed [48].

3.2. Patterning with AKD

Inkjet technology has evolved into a versatile tool for a variety of industrial fabrication processes [64]. In its most basic form, an inkjet printer dispenses picoliter-sized droplets of liquid (ink) onto a substrate at a user-defined position [52]. In this study, PDMS and white glue were also used in the printing process to produce high-resolution patterned papers, but they clogged the inkjet printer cartridges. Therefore, patterning was continued with AKD. AKD was used as a solution in heptane at the ratios of 0.6, 2, and 4% (w/v). Since AKD is colorless, coloring studies were carried out with pigment-based paints to reveal the patterns on the paper. Unfortunately, these particle-based colorings resulted in nozzle clogging.

Examples of microfluidic devices with one, two, or four detection zones have been successfully developed (Figure 3) with 2 and 4% AKD solutions. However, due to the AKD spread of 0.6 %, it is difficult to produce the exactly designed patterns with high resolution. Thickness uniformity results can be influenced by the polymer ink's physical properties and interaction with the substrate [65]. Fluid properties must be within an acceptable range to dispense fluids other than consumer-grade commercial inks. Typically, the permitted range of physical properties is as follows: The viscosity η (1-25 cP), surface tension σ (20-50 dyne/cm), ink density ρ (0.9-1.1 g mL^{-1}), and particle size should be less than 1 μm [66].

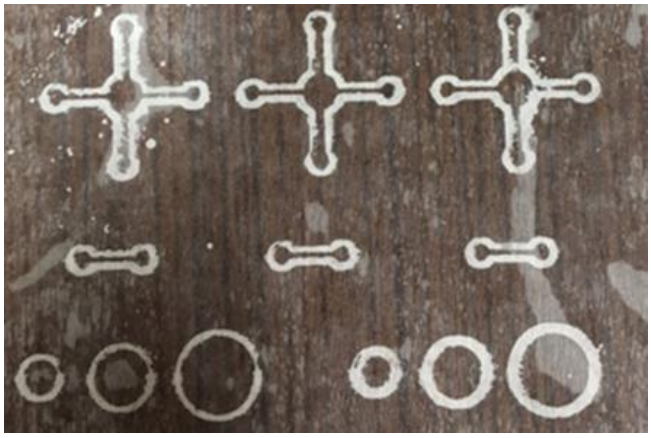


Figure 3. AKD printed filter paper.

AKD needs to be bonded to the paper with hydrogen bonds by keeping it in the oven at 100 °C for 5 min. So, it can finish penetrating throughout the paper and form a hydrophobic barrier while not moving sideways enough to block the channel [67].

Instead of AKD, chemical sensing ink formulations can be printed on different types of [63]. The laser printer's toner can be directly printed over a paper matrix and re-flow with heat to create a hydrophobic barrier within the paper thickness [68]. Wax-based materials can be deposited on chromatography papers via specific wax printers [69].

Inkjet printing's non-contact nature enables the deposition of inks on a variety of substrates, including papers, fabrics, polymers, metals, and biomaterials [70]. It is a non-contact and maskless approach with reduced material waste, low cost, and scalability for large-area manufacturing. Due to these characteristics, it is an appropriate technique for combinatorial studies [71]. Polymers such as PDMS and AKD are non-biodegradable, necessitate organic solvents, and pose risks to human health during their life cycle. Approaches like screen-printing for μPAD fabrication from simple biodegradable polymer Polycaprolactone (PCL) can solve this kind of problem [72].

Plasma treatment is among the chemical methods that change the surface properties of paper. After the paper is coated with a suitable hydrophobic material, the whole surface is made hydrophilic again by plasma treatment [73]. In our experiment, following plasma treatment, the AKD-covered regions at all tested ratios could be rendered hydrophilic. The desired test zones could not be created with the applied metallic masks. Due to excessive plasma etching, the paper samples returned to their fully hydrophilic form. According to the studies conducted, hydrophilic areas could be created in designed shapes using materials that adhere strongly to each other from both sides of the paper, such as magnets.

3.3. Dipping method in paraffin and beeswax melts

The paper is clipped in the center of the round magnets and dipped in melted paraffin or beeswax. After cooling and solidification, hydrophobic barriers are formed due to the penetration around the magnets (Figure 4). The final size of the hydrophilic channel is

determined by the size of the round magnet used. We designed 8-9 mm circles via this technique which serve as barriers. Many biochemical detection methods, such as ELISA, require arrays of hydrophilic dots or microzone plates [74]. The test zones produced can be used in this kind of detection method.

Figure 4. Beeswax patterned filter paper.

The advantages of this method include low reagent costs and a simple procedure. However, due to poor reproducibility and the need for metal molds of a specific shape that are not readily available, mass production is not feasible [75]. It is a more efficient and cost-effective patterning method compared to other tried techniques. It only requires wax dipping, and the test zone was created in less than a minute with subsequent soaking and standard heating techniques. One disadvantage of materials like AKD, paraffin, and beeswax is that they require high temperatures to work with. Another limitation of magnet-based methods is that the resulting test zone shapes are determined by the shape of the magnet. This limits the ability to create the desired designs.

3.4. Evaluating the compatibility of the hydrophobic walls

The type of paper used is entirely determined by the user's application, but it can and will have a significant and predictable impact on performance and fluidic transport [76]. Filter papers were used in all mPAD platform productions. This kind of cellulose chromatography paper had an accessible specific surface area of 9.76 m^2/g [77].

In paper device fabrication, the resolution is typically reported in terms of the minimum functional channel and barrier widths, in this case, meaning that channels wick fluid and barriers do not leak [54]. The next step in our study is to observe the resistance of hydrophobic channels to water, PBS, and organic solvents. All channels and test zones were able to keep the red dye solution within their barriers. Cellulose is insoluble in water and most organic solvents due to its supramolecular structure [26]. Therefore, the determining factor of the study is the behavior of alcohols and acetone within the channels and test areas.

When PDMS is used in its pure form during the drawing process, it cannot create a complete barrier, so leaks occur. When used with a curing agent, we could create hydrophobic channels and test zones that are suitable for water and PBS. It has been observed that water can move through the capillary effect in all channels obtained with PDMS+curing agent, white glue, and AKD. Capillary flow is the natural wicking of liquids in small spaces without the aid of external forces [78]. When the energy reduction associated with the larger liquid-solid surface area outweighs the energy increase associated with the larger liquid-air interface area, the fluid front advances in the channel [79]. However, some real samples may contain solvents or reagents that can break the integrity of the hydrophobic barrier [52]. Areas created with PDMS could not be used with solvents such as alcohol and acetone. These solvents crossed the channel walls and found dispersed to other parts of the chromatography papers. In areas obtained through AKD, alcohol, and acetone behaved similarly. When the fluid fills the entire microchannel, it forms a meniscus at the end. This meniscus is the same size as the liquid head meniscus, but it faces the opposite direction. These opposing capillary forces keep the solution within the microchannel [80]. In a study, leakage has been observed in printed channels when using solvents that have a surface tension lower than 30m Nm^{-1} [81]. The motion of the liquid-solid contact line is dominated by viscous and surface tension forces rather than inertial and gravitational forces [82]. When the surface tension of a liquid is lower than a certain value (Table 1), the liquid tries to move not only in the hydrophilic channels but also in some regions of the barriers that remain hydrophilic. While materials such as AKD and wax hydrophobize the paper, they also reduce the surface energy [29].

Table 1. Surface tensions of the liquids used in the study [83].

Solvent	Surface tension (mN/m)
Isopropanol	23.00
Methanol	22.70
Acetone	25.20
Water	72.80
PBS	69.50*

When white glue is drawn on one side of the paper, the resulting channels leak; however, when drawn on both sides of the paper, no leakage occurs. While the white glue used in the drawing technique is normally a water-soluble material, it becomes insoluble after baking and UV cross-linking processes and acts as a hydrophobic barrier. Because the glue has higher walls than PDMS, alcohols and acetone can flow through the channel. In a study, it was observed that the channels produced from resin using a 3D pen were resistant to organic solvents other than ethanol [63]. It was found that once formed, the glue barriers demonstrated high resistance to aggressive solutions such as surfactants, organic solvents, and strong acids and bases [37]. Studies have shown that the reason why solvents with low surface tension such as alcohol and acetone cannot penetrate paraffin and wax walls is that the channel walls are high due to the production technique. Paraffin and beeswax strengthened the resistance of microchannel barriers to organic solvents, preventing leakage. In a study laser ablation method allowed for the fabrication of parafilm barriers resistant to various organic solvents [39]. Microscale detection is a distinguishing feature of the microfluidic technique. As a result, optimizing the volume used is a necessary step [35]. Since the resolution was low in the hand-drawn channels/test zones, the maximum liquid volume that could be used in the

channel could not be determined. However, in studies conducted with AKD printing, it was observed that a minimum of 5 and a maximum of 20 µL of liquid loading was required to ensure capillary flow in the channels. Capillary-driven microflows differ from pressure- or flow-rate-driven ones in that the engine of motion is located at the liquid's surface. As a result, the engine's strength (capillary force), is determined by the geometry discovered by the advancing interface [84]. The maximum amount increased up to 100 µL due to the increased size of the test zones obtained by the paraffin/beeswax dipping method. It was observed that the EVOLON® worked with the dipping method could hold up to 4 times more liquid than chromatography paper. Fluid confinement is improved in threads, and less fluid is lost unused in the fibers [85]. System modeling is required based on wettability design to understand fluid mechanics [86].

4. Conclusion

In this study, we tried to reveal the strengths and weaknesses of the production techniques and the resulting microfluidic platforms as summarized in Table 2. Patterning paper with PDMS, white glue, AKD, beeswax, and paraffin is much more flexible and simpler than photoresist. Resolution and cost are two of the most important criteria for evaluating the µPAD fabrication process [73]. When compared to the drawing method itself, it was seen that more durable barriers were created with glue, which is a cheaper material compared to other materials we use. Although the method could be applied with a cheap and easily accessible material, the channel resolutions were very low. Inkjet printing provides more control over the pattern because of the deposition technique. However, AKD required the use of heptane, an organic solvent. The printing technique, which enables the production of many different geometries at once, has a high resolution, but its resistance to solvents and reagents needs to be increased. It may be necessary to work with different hydrophobic materials in this regard. In this study, water-soluble white glue was tested using the printing method, but production was halted due to nozzle clogging. Methods that can produce higher resolution with glue can be investigated.

Simplicity allows scientists with limited experience in device fabrication to begin independent experimentation and adaptation of µPADs and µTADs. Detailed research into capillary wicking could help improve device design and suggest new capabilities. The development of new methods of manufacturing will expand their capabilities [87]. The cost of a microfluidic device is also important, and it should be as low as possible, especially if it is intended to be used in resource-limited settings. It is possible to achieve such cost savings by using inexpensive materials in conjunction with a cost-effective manufacturing process for mass production of microfluidic devices [56]. Minimizing the number of chemical solvents lowers manufacturing costs, particularly with the beeswax and paraffin patterned method we described here. It takes about 20-25 minutes to complete the process. Beeswax or paraffin pellets, a beaker, and heating equipment are all inexpensive and widely available. This kind of fabrication method is quite appealing to implement in laboratories with limited resources.

Surface properties of materials are among the most important features that should be evaluated during the production stages [88]. As an alternative to the filter paper used in this study, chromatography papers with different fiber thicknesses and pore structures should also be evaluated for the purpose. Thread types with designable fiber properties can provide different flow patterns and adjustable wettability. Studies can be continued with different papers and threads for the purpose.

By determining the working zones/channels on the paper/thread specific steps of a clinical analysis can be performed, such as sampling, pretreatment, and chemical reactions [89]. Paper and thread-based microfluidic platforms developed in this study can be used in many fields where analytical tests are used. Polymers and polymer-based composites are increasingly used structurally in many industries [90,91]. Various additives and fillers can be used depending on the desired physical data in the field of use as the final product [92]. Creating controllable, high-mechanical composite surfaces with nanomaterials [93] will help researchers in this field develop new perspectives. Natural threads, such as cotton, have superior tangible and synthetic properties in terms of water absorption and stability [94] may be a sustainable alternative. Further research is required to integrate paper and thread-based microfluidics with portable and low-cost platforms. Different channel/test zone geometries can be produced for specific purposes.

Direct methods are simple and ideal for rapid prototyping, whereas the use of intermediaries by indirect methods allows their adaptation in mass production [95]. Other microchannel fabrication methods exist that reduce fabrication cost and complexity, allowing the research community greater access to microfluidics [96]. Studies need to be continued to improve the disadvantages caused by the materials used and to discover a simple, cheap, and high-resolution method that can be accepted universally.

The fabricated microfluidic platforms can be functionalized and used in many areas where analytical tests are applied. Studies on diversifying channel geometries and increasing resolution need to be continued. It should be investigated which devices can be used to obtain qualitative and quantitative results. To make simple and cheap production techniques suitable for mass production, studies should be carried out from different branches.

Acknowledgement

This work was supported by The Scientific and Technological Research Council of Turkey (TUBITAK) Project No: 218M528 and partially by Istanbul Technical University – Scientific Research Projects Unit under the BAP project number 40707.

Table 2. Advantages and disadvantages experienced during work.

Fabrication method	Equipment	Hydrophobic Material	Solvents	Experienced advantages	Experienced disadvantages
Drawing	Ballpoint pen	PDMS White glue	Hexane DI water	Easy and flexible fabrication Easy access to pen and white glue Low cost for pen and white glue The water solubility of white glue Solvent resistance of white glue walls	Poor reproducibility Poor resolution Difficult access to PDMS and its curing agent Organic solvent needs for PDMS Multiple-step procedure Need for heat and UV treatment Cost of heat and UV treatment equipment Requirement of drawing on the double side to prevent leakage The low solvent resistance of PDMS walls
Inkjet Printing	Inkjet printer	AKD	n-Heptane	Easy and rapid fabrication High reproducibility High resolution Possibility of production of different channel designs Obtaining thinner channels compared to other tried methods in the study Low cost of AKD Possibility to work with lower sample/reagent/solvent volumes due to channel structures	Cost of printer Need for organic solvent and heat Multiple-step procedure The necessity of adjusting viscosity and concentration in a printable hydrophobic material
Plasma treatment	Plasma cleaner	AKD	n-Heptane	Easy fabrication Low cost of AKD	Need for plasma equipment High cost and accessibility of plasma equipment Customized masks are required for high-resolution plasma patterning Need for organic solvent and heat Multiple-step procedure
Dipping	Heater	Paraffin Beeswax	-	Easy fabrication One step procedure Low cost of paraffin, beeswax, and heater No solvent need High organic solvent resistance of patterned walls Applicable to both paper and thread	Poor reproducibility Poor resolution Customized magnets are required for different channel geometries

Author contributions

Nagihan Okutan Arslan: Design, fabrication, analysis, interpretation, writing, and editing**. Raghied Mohammed Helmy Atta:** Design and fabrication via drawing, plasma etching, printing techniques, interpretation**. Levent Trabzon:** Conception, design, reviewing, editing

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Aralekallu, S., Boddula, R., & Singh, V. (2023). Development of glass-based microfluidic devices: A review on its fabrication and biologic applications. Materials & Design, 225, 111517. https://doi.org/10.1016/j.matdes.2022.111517
2. Gao, B., Li, X., Yang, Y., Chu, J., & He, B. (2019). Emerging paper microfluidic devices. Analyst, 144, 6497–6511. https://0-doi-org/10.1039/C9AN01275C
3. Mousaabadi, K.Z., Vandishi, Z.T., Kermani, M., Arab, N., & Ensafi, A.A. (2023). Recent developments toward microfluidic point-of-care diagnostic sensors for viral infections. Trends in Analytical Chemistry, 169, 117361. https://doi.org/10.1016/j.trac.2023.117361
4. Shi, L., Li, Y., Jia, C., Shan, J., Wang, S., Liu, S., Sun, J., Zhang, D., Ji, Y., & Wang, J. (2023). An overview of fluorescent microfluidics into revealing the mystery of food safety analysis: Mechanisms and recent applications. Trends in Food Science & Technology, 138, 100–115. https://doi.org/10.1016/j.tifs.2023.05.016
5. Li, T., Yang, N., Pan, X., Zhang, X., & Xu, L. (2024). A portable microfluidic photometric detection method

based on enzyme linked immunoscatter enhancement. Biosensors and Bioelectronics, 244, 115794. https://doi.org/10.1016/j.bios.2023.115794
6. Thakur, R., & Fridman, G. Y. (2022). Low cost, ease-of-access fabrication of microfluidic devices using wet paper molds. Micromachines (Basel), 13(9), 1408. https://doi.org/10.3390/mi13091408
7. Hassan, M. M., Yi, X., Zareef, M., Li, H., & Chen, Q. (2023). Recent advancements in optical, electrochemical, and photoelectrochemical transducer-based microfluidic devices for pesticide and mycotoxins in food and water. Trends in Food Science & Technology, 142, 104230. https://doi.org/10.1016/j.tifs.2023.104230
8. Lu, S.-Y., Tseng, C.-C., Yu, C.-X., Chen, T.-L., Huang, K.-H., Fu, L.-M., & Wu, P.-H. (2024). Rapid microfluidic fluorescence detection platform for determination of whole blood sodium. Sensors & Actuators: B. Chemical, 400, 134839. https://doi.org/10.1016/j.snb.2023.134839
9. Pou, K.R.J., Raghavan, V., & Packirisamy, M. (2022). Microfluidics in smart packaging of foods. Food Research International, 161, 111873. https://doi.org/10.1016/j.foodres.2022.111873
10. Berthier, E., Dostie, A.M., Lee, U.N., Berthier, J., & Theberge, A.B. (2019). Open Microfluidic Capillary Systems. Anal Chemistry, 91(14), 8739–8750. https://doi.org/10.1021/acs.analchem.9b01429
11. Zhang, Y., Yu, Y., Yang, X., Yuan, X., & Zhang, J. (2024). Pb(II) inhibits CRISPR/Cas12a activation and application for paper-based microfluidic biosensor assisted by smartphone. Sensors & Actuators: B. Chemical, 398, 134732. https://doi.org/10.1016/j.snb.2023.134732
12. Nishat, S., Jafry, A. T., Martinez, A. W., & Awan, F. R. (2021). Paper-based microfluidics: Simplified fabrication and assay methods. Sensors and Actuators: B. Chemical, 336, 129681. https://doi.org/10.1016/j.snb.2021.129681
13. Agustini, D., Caetano, F. R., Quero, R. F., da Silva, J. A. F., Bergamini, M. F., Marcolino-Junior, L. H., & de Jesus, D. P. (2021). Microfluidic devices based on textile threads for analytical applications: state of the art and prospects. Anal. Methods, 13, 4830. https://0-doi-org/10.1039/D1AY01337H
14. Martinez, A.W., Phillips, S.T., Butte, M.J., & Whitesides, G.M. (2007). Patterned paper as a platform for inexpensive, low-volume, portable bioassays. Angewandte Chemie International Edition, 46(8), 1318-20. https://doi.org/10.1002/anie.200603817
15. Teymoori, M. & Yalçınkaya, A.D. (2023). A low-cost microwave metamaterial-inspired sensing platform for quantitative paper microfluidic analytical devices. Sensors & Actuators: A. Physical, 363, 114684. https://doi.org/10.1016/j.sna.2023.114684
16. Sanjayan, C. G., Ravikumar, C. H., & Balakrishna, R. G. (2023). Perovskite QD-based paper microfluidic device for simultaneous detection of lung cancer biomarkers – Carcinoembryonic antigen and neuron-specific enolase. Chemical Engineering Journal, 464, 142581. https://doi.org/10.1016/j.cej.2023.142581
17. Saiboh, T., Malahom, N., Prakobjij, A., Seebunrueng, K., Amatatongchai, M., Chairam, S., Sameenoi, Y., & Jarujamrus, P. (2023). Visual detection of formalin in food samples by using a microfluidic thread-based analytical device. Microchemical Journal, 190, 108685. https://doi.org/10.1016/j.microc.2023.108685
18. Hamidon, N. N., Hong, Y., Salentijn, G. IJ., & Verpoorte, E. (2018). Water-based alkyl ketene dimer ink for user-friendly patterning in paper microfluidics. Analytica Chimica Acta, 1000, 180-190. https://doi.org/10.1016/j.aca.2017.10.040
19. Derakhshani, M., Jahanshahi, A., & Ghourchian, H. (2023). Addressing the sample volume dependency of the colorimetric glucose measurement on microfluidic paper-based and thread/paper-based analytical devices using a novel low-cost analytical viewpoint. Microchemical Journal, 195, 109545. https://doi.org/10.1016/j.microc.2023.109545
20. Hussain, G. H., Jafry, A. T., Malik, S., Shah, S. F., Nishat, S., & Awan, F. R. (2023). Multifunctional rotational active valve for flow control in paper-based microfluidic devices. Sensors & Actuators: B. Chemical, 378, 133142. https://doi.org/10.1016/j.snb.2022.133142
21. Müller, R. H. & Clegg, D. L. (1949). Automatic paper chromatography. Analytcial Chemistry, 21, (9), 1123-1125. https://pubs.acs.org/doi/abs/10.1021/ac60033a032
22. Sarabi, M., Yigci, D., Alseed, M. M., Mathyk, B. A., Ata, B., Halicigil, C., & Tasoğlu, S. (2022). Disposable paper-based microfluidics for fertility testing. iScience, 25, 104986. https://doi.org/10.1016/j.isci.2022.104986
23. Kulkarni, M. B., Ayachit, N. H., Aminabhavi, T. M., & Pogue, B. W. (2023). Recent advances in microfluidics-based paper analytical devices (μPADs) for biochemical sensors: From fabrication to detection techniques. Biochemical Engineering Journal, 198, 109027. https://doi.org/10.1016/j.bej.2023.109027
24. Yuan, M., Li, C., Zheng, Y., Cao, H., Ye, T., Wu, X., Hao, L., Yin, F., Yu, J., & Xu, F. (2024). A portable multi-channel fluorescent paper-based microfluidic chip based on smartphone imaging for simultaneous detection of four heavy metals. Talanta, 266, 125112. https://doi.org/10.1016/j.talanta.2023.125112
25. Çelik, M. (2022). An experimental study of the performance of a low-cost paper-based membraneless direct hydrogen peroxide fuel cell. Turkish Journal of Engineering, 6(2), 161-165. https://doi.org/10.31127/tuje.891626
26. Klemm, D., Heublein, B., Fink, H. -P., & Bohn, A. (2005). Cellulose: Fascinating Biopolymer and Sustainable Raw Material. Angewandte Chemie International Edition, 44, 3358 – 3393. https://0-doi-org/10.1002/anie.200460587
27. Pelton, R. (2009). Bioactive paper provides a low-cost platform for diagnostics. Trends Analyt Chemistry, 28(8), 925–942. https://doi.org/10.1016/j.trac.2009.05.005
28. Ray, R., Prabhu, A., Prasad D., Garlapati, V.K., Amnabhavi, M., Mani, N.K., & Simal-Gandara, J. (2022).

Paper-based microfluidic devices for food adulterants: Cost-effective technological monitoring systems. Food Chemistry, 390, 133173. https://doi.org/10.1016/j.foodchem.2022.133173
29. Li, X., Ballerini, D.R., & Shen, W. (2012). A perspective on paper-based microfluidics: Current status and future trends. Biomicrofluidics, 6(1), 11301-1130113. https://doi.org/10.1063/1.3687398
30. Guan, H., Du, S., Han, B., Zhang, Q., & Wang, D. (2023). A rapid and sensitive smartphone colorimetric sensor for detection of ascorbic acid in food using the nanozyme paper-based microfluidic chip. LWT - Food Science and Technology, 184, 115043. https://doi.org/10.1016/j.lwt.2023.115043
31. Xia, Y., Si, J., & Li, Z. (2016). Fabrication techniques for microfluidic paper-based analytical devices and their applications for biological testing: A review. Biosensors and Bioelectronics, 77, 774–789. http://dx.doi.org/10.1016/j.bios.2015.10.032
32. Pantoja, R., Nagarah, J. M., Starace, D. M., Melosh, N. A., Blunck, R., Bezanilla, F., & Heath, J. R. (2004). Silicon chip-based patch-clamp electrodes integrated with PDMS microfluidics. Biosens Bioelectronics, 20, 509–517. https://doi.org/10.1016/j.bios.2004.02.020
33. Chung, S., Loh, A., Jennings, C. M., Sosnowski, K., Ha, S. Y., Yim, H., & Jeong-Yeol, Y. (2023). Capillary flow velocity profile analysis on paper-based microfluidic chips for screening oil types using machine learning. Journal of Hazardous Materials, 447, 130806. https://doi.org/10.1016/j.jhazmat.2023.130806
34. Zhang, Q., Feng, S., Lin, L., Mao, S., & Lin, J.-M. (2021). Emerging open microfluidics for cell manipulation. Chemical Society Reviews, 50, 5333–5348. https://doi.org/10.1039/D0CS01516D
35. Chaikan, P., Udnan, Y., Sananmuang, R., Ampiah-Bonney, R.J., & Chaiyasith, W.C. (2020). A low-cost microfluidic paper-based analytical device (μPAD) with column chromatography preconcentration for the determination of paraquat in vegetable samples. Microchemical Journal, 159, 105355. https://doi.org/10.1016/j.microc.2020.105355
36. Nery, E. & Kubota, L. T. (2013). Sensing approaches on paper-based devices: a review. Anal Bioanal Chemistry, 405, 7573–7595. https://0-doi-org/10.1007/s00216-013-6911-4
37. Cardoso, T. M. G., de Souza, F. R., Garcia, P. T., Rabelo, D., Henry, C. S., & Coltro, W. K. T. (2017). Versatile fabrication of paper-based microfluidic devices with high chemical resistance using scholar glue and magnetic masks. Analytica Chimica Acta, 974, 63-68. http://dx.doi.org/10.1016/j.aca.2017.03.043
38. Silva-Neto, H. A., Arantes, I. V. S., Ferreira, A. L., do Nascimento, G. H. M., Meloni, G. N., de Araujo, W. R., Paixão, T. R. L. C., & Coltro, W. K. T. (2023). Recent advances on paper-based microfluidic devices for bioanalysis. TrAC Trends in Analytical Chemistry, 158, 116893. https://doi.org/10.1016/j.trac.2022.116893
39. Safiabadi Tali, S. H., Hajimiri, H., Sadiq, Z., & Jahanshahi-Anbuhi, S. (2023). Engineered detection zone to enhance color uniformity on paper microfluidics fabricated via Parafilm®-heating-laser-cutting. Sensors and Actuators B: Chemical, 380, 133324. https://doi.org/10.1016/j.snb.2023.133324
40. Stefano, J. S., Orzari, L. O., Silva-Neto, H. A., de Ataíde, V. N., Medes, L. F., & Coltro, W. K. T. (2022). Different approaches for fabrication of low-cost electrochemical sensors. Current Opinion in Electrochemistry, 32, 100893. https://doi.org/10.1016/j.coelec.2021.100893
41. Liao, X., Zhang, Y., Zhang, Q., Zhou, J., Ding, T., & Feng, J. (2023). Advancing point-of-care microbial pathogens detection by material-functionalized microfluidic systems. Trends in Food Science & Technology, 135, 115–130. https://doi.org/10.1016/j.tifs.2023.03.022
42. Tan, W., Powles, E., Zhang, L., & Shen, W. (2021). Go with the capillary flow. Simple thread-based microfluidics. Sensors and Actuators: B. Chemical, 334, 129670. https://doi.org/10.1016/j.snb.2021.129670
43. Baysal, G., Önder, S., Göcek, İ., Trabzon, L., Kızıl, H., Kök, F. N., & Kayaoğlu, B. K. (2015). Design and fabrication of a new nonwoven-textile-based platform for biosensor construction. Sensors and Actuators B, 208, 475–484. https://dx.doi.org/10.1016/j.snb.2014.11.042
44. Chen, L., Ghiasvand, A., & Paull, B. (2023). Applications of thread-based microfluidics: Approaches and options for detection. Trends in Analytical Chemistry, 161, 117001. https://doi.org/10.1016/j.trac.2023.117001
45. Mesquita, P., Gong, L., & Lin, Y. (2022). Low-cost microfluidics: Towards affordable environmental monitoring and assessment. Frontiers in Lab on a Chip Technologies, 1, 1074009. https://doi.org/10.3389/frlct.2022.1074009
46. Tzianni, E. I., Sakkas, V. A., & Prodromidis, M. I. (2024). Wax screen-printable ink for massive fabrication of negligible-to-nil cost fabric-based microfluidic (bio)sensing devices for colorimetric analysis of sweat. Talanta, 269, 125475. https://doi.org/10.1016/j.talanta.2023.125475
47. Bruzewicz, D. A., Reches, M., & Whitesides, G. M. (2008). Low-cost printing of poly (dimethylsiloxane) barriers to define microchannels in paper. Anal. Chemistry, 80, 3387-3392. https://0-doi-org/10.1021/ac702605a
48. Li, X., Tian, J., Garnier, G., & Shen, W. (2010). Fabrication of paper-based microfluidic sensors by printing. Colloids and Surfaces B: Biointerfaces, 76, 564–570. https://doi.org/10.1016/j.colsurfb.2009.12.023
49. Li., X., Tian, J., Nguyen, T., & Shen, W. (2008). Paper-Based Microfluidic Devices by Plasma Treatment. Anal Chemistry, 2008, 80, 9131–9134. https://doi.org/10.1021/ac801729t
50. Songjaroen, T., Dungchai, W., Chailapakul, O., & Laiwattanapaisal, W. (2011). Novel, simple and low-cost alternative method for fabrication of paper-based microfluidics by wax dipping. Talanta, 85, 2587–2593. https://dx.doi.org/10.1016/j.talanta.2011.08.024
51. Bandopadhyay, A. A., & Das, P. K. (2023). Paper based microfluidic devices: a review of fabrication

techniques and applications. The European Physical Journal Special Topics, 232(6), 781-815. https://doi.org/10.1140/epjs/s11734-022-00727-y
52. Yamada, K., Henares, T.G., Suzuki, K., & Citterio, D. (2015). Paper-Based Inkjet-Printed Microfluidic Analytical Devices. Angewandte Chemie International Edition., 54, 5294–5310. https://0-doi-org/10.1002/anie.201411508
53. Levine, L. M. & Campbell, T. (2007). Combining Additive and Subtractive Techniques in the Design and Fabrication of Microfluidic Devices. NSTI Nanotechnology Conference and Trade Show., 3, 385.
54. Roller, R. M. & Lieberman, M. (2023). Beyond wax printing: The future of paper analytical device fabrication. Sensors and Actuators: B. Chemical, 392, 134059. https://doi.org/10.1016/j.snb.2023.134059
55. Trinh, K. T. L., Chae, W. R., & Lee, N. Y. (2022). Recent advances in the fabrication strategies of paper-based microfluidic devices for rapid detection of bacteria and viruses. Microchemical Journal, 180, 107548. https://doi.org/10.1016/j.microc.2022.107548
56. Coltro, W. K. T., Chao-Min, C., Carrilho, E., & de Jesus, D. P. (2014). Recent advances in low-cost microfluidicplatforms for diagnostic applications. Electrophoresis, 35, 2309–2324. https://0-doi-org/10.1002/elps.201400006
57. Li, Z., Liu, H., He, X., Xu, F., & Li, F. (2018). Pen-on-paper strategies for point-of-care testing of human health. TrAC Trends in Analytical Chemistry, 108, 50-64. https://doi.org/10.1016/j.trac.2018.08.010
58. Lee, J. N., Park, C., & Whitesides, G. M. (2003). Solvent Compatibility of Poly(dimethylsiloxane)-Based Microfluidic Devices. Anal Chemistry, 75, 6544-6554. https://0-doi-org /10.1021/ac0346712
59. Oyola-Reynoso, S., Heim, A. P., Halbertsma-Black, J., Zhao, C., Tevis, I. D., Çınar, S., Cademartiri, R., Liu, X., Bloch, J.-F., & Thuo, M. M. (2015). Draw your assay: Fabrication of low-cost paper-based diagnostic and multi-well test zones by drawing on a paper. Talanta, 144, 289–293. http://dx.doi.org/10.1016/j.talanta.2015.06.018
60. Lu, Y., Shi, W., Jiang, L., Qin, J, & Lin, B. (2009). Rapid prototyping of paper-based microfluidics with wax for low-cost, portable bioassay. Electrophoresis, 30, 1497–1500. https://0-doi-org/10.1002/elps.200800563
61. Walia, S., Bhatnagar, I., Liu, J., Mitra, S.K., & Asthana, A. (2021). A novel method for fabrication of paper-based microfluidic devices using BSA-ink. International Journal of Biological Macromolecules, 193, 1617–1622. https://doi.org/10.1016/j.ijbiomac.2021.10.224
62. Nuchtavorn, N. & Macka, M. (2016). A novel highly flexible, simple, rapid and low-cost fabrication tool for paper-based microfluidic devices (mPADs) using technical drawing pens and in-house formulated aqueous inks. Analytica Chimica Acta, 919, 70-77. http://dx.doi.org/10.1016/j.aca.2016.03.018
63. Sousa, L. R., Duarte, L. C., & Coltro, W. K. T. (2020). Instrument-free fabrication of microfluidic paper-based analytical devices through 3D pen drawing. Sensors & Actuators: B. Chemical, 312, 128018. https://doi.org/10.1016/j.snb.2020.128018
64. Abe, K., Suzuki, K., & Ctterio, D. (2008). Inkjet-Printed Microfluidic Multianalyte Chemical Sensing Paper. Anal Chemistry, 80,6928–6934. https://0-doi-org/10.1021/ac800604v
65. Gassend, V., Hauf, C. R., & Chen, J. (2022). Research and Applications of Inkjet Printing for OLED Mass Production. SID Symposium Digest of Technical Papers, 53, 1, 398-401. https://0-doi-org/10.1002/sdtp.15505
66. Waasdorp, R., van den Heuvel, O., Versluis, F., Hajee, B., & Ghatkesar, M. K. (2018). Accessing individual 75-micron diameter nozzles of a desktop inkjet printer to dispense picoliter droplets on demand. RSC Advances, 18, 8(27), 14765–14774. https://doi.org/10.1039%2Fc8ra00756j
67. Espinosa, A., Diaz, J., Vazquez, E., Acosta, L., Santago, A., & Cunci, L. (2022). Fabrication of paper-based microfluidic devices using a 3D printer and a commercially-available wax filament. Talanta Open, 6, 100142. https://doi.org/10.1016/j.talo.2022.100142
68. Ghosh, R., Gopalakrishnan, S., Savitha, R., Renganathan, T., & Pushpavanam, S. (2019). Fabrication of laser printed microfluidic paper-based analytical devices (LP-μPADs) for point-of-care applications. Scientific Reports, 9, 7896. https://doi.org/10.1038%2Fs41598-019-44455-1
69. Carrilho, E., Martinez, A. W., & Whitesides, G. M. (2009). Understanding wax printing: A simple micropatterning process for paper-based microfluidics. Anal Chemistry, 81, 16, 7091–7095. https://doi.org/10.1021/ac901071p
70. Zhan, Z., An, J., Wei, Y., Tran, V. T., & Du, H. (2017). Inkjet-printed optoelectronics. Nanoscale, 9, 965–993. https://0-doi-org/10.1039/C6NR08220C
71. Singh, B. M., Haverinen, H. M., Dhagat, P., & Jabbour, G. E. (2010). Inkjet Printing—Process and Its Applications. Advanced Materials, 22, 673–685. https://0-doi-org/10.1002/adma.200901141
72. Mettakoonpitak, J., Khongsoun, K., Wongwan, N., kaewbutdee, S., Siripinyanond, A., Kuharuk, A., & Henry, C. S. (2021). Simple biodegradable plastic screen-printing for microfluidic paper-based analytical devices. Sensors and Actuators B: Chemical, 331, 15, 129463. https://doi.org/10.1016/j.snb.2021.129463
73. He, Y., Wu, Y., Fu, J. Z., & Wu, W. B. (2015). Fabrication of paper-based microfluidic analysis devices: a review. RSC Advances, 5, 78109–78127. https://0-doi-org/10.1039/C5RA09188H
74. Chitnis, G., Ding, Z., Chang, C. L., Savran, C. A., & Ziaie, B. (2011). Laser-treated hydrophobic paper: an inexpensive microfluidic platform. Lab Chip, 11, 1161–1165. https://0-doi-org/10.1039/C0LC00512F
75. Tong, X., Ga, L., Zhao, R., & Ai, J. (2021). Research progress on the applications of paper chips. RSC Advances, 11(15), 8793–8820. https://doi.org/10.1039%2Fd0ra10470a
76. Cate, D. M., Adkins, J. A., Mettakoonpitak, J., & Henry, C. S. (2015). Recent Developments in Paper-Based Microfluidic Devices. Anal Chemistry, 87, 19–41. https://doi.org/10.1021/ac503968p

77. Hong, J., Ye, X., & Zhang, Y.-H. (2007). Quantitative determination of cellulose accessibility to cellulase based on adsorption of a nonhydrolytic fusion protein containing CBM and GFP with its applications. Langmuir, 23, 12535-12540. https://0-doi-org/10.1021/la7025686
78. Kolliopoulos, P. & Kumar, S. (2021). Capillary flow of liquids in open microchannels: overview and recent advances. npj Microgravity 7, 51, 9242. https://doi.org/10.1038/s41526-021-00180-6
79. Casavant, B. P., Berthier, E., Theberge, A. B., Berthier, J., Montanez-Sauri, S. I., Bischel, L. L., Brakke, K., Hedman, C. J., Bushman, W., Keller, N. P., & Beebe, D. J. (2013). Suspended microfluidics. Proceedings of the National Academy of Sciences, 18, 110(25), 10111-10116. https://doi.org/10.1073%2Fpnas.1302566110
80. Xian, Z., Dai, P., Su, W., Sun, C., Liu, L., You, H., & Liu, Y. (2023). A novel microfuidics PMMA/paper hybrid bioimmunosensor for laser-induced fuorescence detection in the determination of alpha-fetoprotein from serum. Microchemical Journal, 195, 109476. https://doi.org/10.1016/j.microc.2023.109476
81. Maejima, K., Tomikawa, S., Suzuki, K., & Citterio, D. (2013). Inkjet printing: an integrated and green chemical approach to microfluidic paper-based analytical devices. RSC Advances, 3, 9258-9263. https://doi.org/10.1039/C3RA40828K
82. Boylu, M. A., & Ceyhan, U. Controlling the Motion of Interfaces in Capillary Channels with Non-uniform Surface Wettability. Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 25(75), 675-691. https://doi.org/10.21205/deufmd.2023257513
83. Janiszewska, N., Rackowska, J., Budkowski, A., Gajos, K., Stetsyshyn, Y., Michalik, M., & Awsiuk, K. (2020). Dewetting of Polymer Films Controlled by Protein Adsorption. Langmuir, 36, 11817−11828. https://doi.org/10.1021/acs.langmuir.0c01718
84. Berthier, J., Gosselin, D., Pham, A., Delapierre, G., Belgacem, N., & Chaussy, D. (2016). Capillary Flow Resistors: Local and Global Resistors. Langmuir, 32, 3, 651-928. https://doi.org/10.1021/acs.langmuir.5b02090
85. Berthier, J., Brakke, K.A., Gosselin, D., Berthier, E., & Navarro, F. (2017). Thread-based microfluidics: Flow patterns in homogeneous and heterogeneous microfiber bundles. Medical Engineering and Physics, 48, 55–61. https://doi.org/10.1016/j.medengphy.2017.08.004
86. Zheng, Q., Wang, B., & Guo, Z. (2024). Recent advances in microfluidics by tuning wetting behaviors. Materials Today Physics, 40, 101324. https://doi.org/10.1016/j.mtphys.2023.101324
87. Martinez, A. W., Plillips, S. T., Whitesides, G. M., & Carrilho, E. (2010). Diagnostics for the Developing World: Microfluidic Paper-Based Analytical Devices. Analytical Chemistry, 82, 1, 3-10. https://doi.org/10.1021/ac9013989
88. Buldum, B. B., & Cagan, S. C. (2017). The optimization of surface roughness of AZ91D magnesium alloy using ANOVA in ball burnishing process. Turkish Journal of Engineering, 1(1), 25-31. https://doi.org/10.31127/tuje.316860
89. Brazaca, L. C., Imamura, A. H., Blasques, R. V., Camargo, J. R., Janegitz, B. C., & Carrilho, E. (2024). The use of biological fluids in microfluidic paper-based analytical devices (μPADs): Recent advances, challenges and future perspectives. Biosensors and Bioelectronics, 246, 115846. https://doi.org/10.1016/j.bios.2023.115846
90. Kepır, Y., Gunoz, A., & Kara, M. (2022). Repairing of damaged composite materials and self-healing composites. Turkish Journal of Engineering, 6(2), 149-155. https://doi.org/10.31127/tuje.866955
91. Özbek, Ö., Bozkurt, Ö. Y., & Erkliğ, A. (2020). Low velocity impact behaviors of basalt/epoxy reinforced composite laminates with different fiber orientations. Turkish Journal of Engineering, 4(4), 197-202. https://doi.org/10.31127/tuje.644025
92. Güngör, A. (2023). The effect of Cumin Black (Nigella Sativa L.) as bio-based filler on chemical, rheological and mechanical properties of epdm composites. Turkish Journal of Engineering, 7(4), 279-285. https://doi.org/10.31127/tuje.1180753
93. Güler, Ö., gökhan Albayrak, M., Takgün, M., & Güler, S. H. (2017). The investigation on electrical and optical properties of CdO/CNT nanocomposite. Turkish Journal of Engineering, 1(2), 61-65. https://doi.org/10.31127/tuje.317778
94. Koruyucu, A. (2019). Removal of colour pollutions in dye baths with mordants. Turkish Journal of Engineering, 3(4), 201-205. https://doi.org/10.31127/tuje.556349
95. Holman, J. B., Shi, Z., Fadahunsi, A. A., Li, C., & Ding, W. (2023). Advances on microfluidic paper-based electroanalytical devices. Biotechnology Advances, 63, 108093. https://doi.org/10.1016/j.biotechadv.2022.108093
96. Guler, M. T., & Bilican, İ. (2020). A new method for the measurement of soft material thickness. Turkish Journal of Engineering, 4(2), 97-103. https://doi.org/10.31127/tuje.636350

Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Investigation of the effect of additives on the microstructure of clay

Atila Demiröz [1], Onur Saran [*2]

[1] *Konya Technical University, Department of Civil Engineering, Türkiye, ademiroz@ktun.edu.tr*
[2] *Van Yüzüncü Yıl University, Department of Civil Engineering, Türkiye, onursaran@yyu.edu.tr*

Cite this study: Demiröz, A., & Saran, O. (2024). Investigation of the effect of additives on the microstructure of clay. Turkish Journal of Engineering, 8 (3), 563-571

https://doi.org/10.31127/tuje.1439113

Keywords
Soil stabilization
Basalt fiber
XRD
SEM

Research Article

Received: 18.02.2024
Revised: 19.03.2024
Accepted: 25.03.2024
Published: 15.07.2024

Abstract

In civil engineering, some soils cause many problems in terms of geotechnical engineering. Especially high plasticity clayey soils cause serious problems in road, airport, pavement and highway construction. Such soils can be stabilized using the chemical stabilization method. Additives such as lime, cement, fly ash and blast furnace slag are generally used in chemical stabilization. In addition, in recent years, there have been many studies on the use of natural and artificial fibers in ground stabilization. In this study, the effects of basalt fiber and mineral additives on the microstructure of clay soil were examined. Microstructures of pure and additively mixed clayey soil specimens were investigated for this purpose. X-ray diffraction (XRD) and scanning electron microscopy (SEM) analyses were performed on compacted soil specimens. In sum, it has been shown that the addition of lime, fly ash, and silica fume is effective in improving the compactivity properties. Pore sizes in SEM images vary depending on additive addition. SEM images showed that the soil particles adhered to the basalt fiber surface, which contributed to the force and friction between the soil particles and the basalt fiber.

1. Introduction

Expansive soils are frequently encountered in geotechnical engineering applications. In general, soils with low strength, high compressibility, and high volumetric changes cause a slew of problems. Before using clayey soils in roads, airports, sidewalks, and highways, they must be improved. Soil stabilization leads to a decrease in the soil's plasticity and swelling potential while increasing soil strength and workability [1-2]. Chemical or mechanical techniques can be used to achieve stabilization. In situations where mechanical stabilization is inadequate, chemical stabilization is utilized. Chemical stabilization involves adding various chemical additives to the soil to initiate reactions that enhance its strength, durability, and compression properties. Chemical stabilization is usually accomplished by incorporating additives into the soil, such as cement, asphalt, lime, and fly ash [3-5].

Stabilizations with cement and lime are widely used globally. Stabilized soils go through cementation and ion exchange reactions, which enhance the soil's chemical and mechanical characteristics. Previous studies have focused on the improved soils' geotechnical behavior [6-9].

Lime and cement are useful additives for fine-grained soils. When lime or cement additives are mixed with soil, cation exchange, flocculation and agglomeration, cementitious hydration, and pozzolanic reactions take place, which improve the clay's properties [10]. While cation exchange occurs in a short time, pozzolanic reactions take place over a longer period. When lime is added to the soil, pozzolanic reactions result in the formation of products such as hydrated calcium silicate (CS), hydrated calcium aluminate/silicate (CAS/CAH) [11-13]. Calcium silicate/alumina hydrate gels improve soil's mechanical properties by filling voids. A number of researchers have investigated the relationship between improved soil microstructure and geotechnical engineering behavior [14-16].

In the last few years, in addition to such additives, the number of studies involving fibers in ground stabilization has increased. One of these fibers is basalt fiber. Basalt fiber is obtained by melting basalt rock at high temperatures and then pulling it as fibers. Studies have shown that fibers have good properties and can interact with soil particles to increase overall strength when mixed with soil. In geotechnical engineering, the use of fiber reinforcement in soil stabilization enhances soil strength, prevents the formation of tensile cracks,

prevents swelling tendencies in expansive soils, improves hydraulic conductivity and liquefaction resistance, and reduces soil brittleness. Furthermore, basalt fiber, among other fibers, is an environmentally friendly and pollution-free high-tech fiber. The three-dimensional structure created among the fibers reduces soil deformations, forms a friction surface between the soil particles and the fibers, and improves the mechanical properties of the soil [17-21].

In this study, the effect of basalt fiber and mineral additives (lime, fly ash, and silica fume) on the microstructure of high plasticity clayey soil has been investigated. To investigate the effects of additives on soil composition, both X-ray diffraction (XRD) and scanning electron microscopy (SEM) were conducted on samples with and without additives.

2. Method

2.1. Soil properties

In the study, a high plasticity clay (CH) soil exhibiting swelling properties was used. The clayey soil used in the study was obtained from an area in Konya province, Selçuklu district, 2nd Organized Industrial Zone. About 5 tons of clayey soil were excavated from a depth of 6–7 m below the surface, put into a truck, and driven to the geotechnical laboratory at Konya Technical University. Sieve analysis, hydrometer test, Atterberg limits, and pycnometer test were conducted according to the relevant ASTM standard to determine the basic physical properties of the soil (Table 1).

Table 1. Physical properties of soil.

Liquid limit, ω_L (%)	105
Plastic limit, ω_p (%)	31
Plasticity index (I_P) (%)	74
Specific gravity	2.65
Maximum dry density, ρ_{kmax} (g/cm^3)	1.45
Optimum moisture content, w_{opt} (%)	28.5
Soil classification (USCS)	CH

2.2. Properties of additive materials

Lime (L): In the study, Erciyes TS EN 459-1 CL 80-S hydrated lime has been used.

Fly ash (FA): The experimental study used fly ash obtained from the Kütahya Seyitömer Thermal Power Plant. The obtained fly ash meets the ASTM C 618 standard, with a combined SiO_2+Al_2O_3+Fe_2O_3 content of 84.34% (S+A+F > 70%) and a CaO content of less than 10%, categorizing it as Class F (low-calcium) fly ash.

Silica fume (SF): The silica fume additive used in the study was obtained from Antalya Elektrometalurji A.Ş.

The compositions of the chemicals of the additives are shown in Table 2. The chemical analysis results of the additives have been obtained from the companies from which they have been supplied.

2.3. Basalt fiber (BF)

Basalt fibers have been supplied by Spinteks Textile Construction Industry and Trade Inc. (Denizli/Türkiye) (Figure 1). The fibers are 6, 12, 18, and 24 mm in length.

Table 2. Chemical composition of additives used in the study.

Chemical Composition	Silica Fume (SF) (%)	Fly ash (FA) (%)	Lime (L)
SiO_2	79.94	54.49	-
Al_2O_3	0.83	20.58	-
Fe_2O_3	0.41	9.27	-
CaO	2.53	4.26	≥80
MgO	7.68	4.48	≤5
K_2O	-	2.01	-
SO_3	-	0.52	≤2
C	1.22	-	-
S	0.923	-	-
CO_2	-	-	≤7
Free lime			≥65
Loss of ignition	2.96	3.01	

Figure 1. Basalt fibers.

Two different experimental design tables have been created using the Taguchi method for experimental studies. The Taguchi method is an experimental design method that tries to minimize the variability in the product and process by choosing the most appropriate combination of levels of controllable factors against the uncontrollable factors that create variability in the product and process. The Taguchi experimental design method involves a series of steps: Firstly, identifying the problem requiring improvement, then selecting and standardizing the variables affecting performance characteristics. Next, selecting an appropriate orthogonal array, assigning factors and interactions to columns, and choosing quality loss functions along with performance statistics. Following this, conducting experiments, recording results, analyzing data, and determining the optimal values for controllable variables. Within this method, measurement outcomes are transformed into Signal/Noise (S/N) ratios. These ratios represent the ratio of the desired actual value from the system to a factor not initially considered but influencing the test outcome. In this context, 'signal' is denoted by S, and 'noise' by N. Enhanced product quality is achieved through higher S/N ratios, achieved by maximizing signal and minimizing noise. Calculation methods for S/N ratios vary, with the most common being 'Larger is Better', 'Smaller is Better', and 'Nominal is Best' approaches [22].

3% lime was added to all designs containing basalt fiber-fly ash and basalt fiber-silica fume additives to increase pozzolanic interaction. The percentage of lime added to the soil may vary depending on its use to stabilize the soil or increase its workability. Small

amounts of lime are required to increase workability, while large percentages of lime are required to increase stability. Previous studies have shown that a minimum of 3% lime addition is required for the stabilization of high plasticity clay soil [23-24]. For this reason, the lime percentage was chosen as 3% in this experimental study. Additionally, additive contents were determined by taking into account the optimum additive percentages obtained in studies in the literature.

Unconfined compressive tests (UCS) were performed on samples prepared at optimum water content and maximum dry density values. Average unconfined compressive strengths (q_u) were obtained by performing experiments on three samples for each design. The test design table, compaction parameters and unconfined compressive strengths are given in Table 3 and Table 4.

Table 3. Taguchi L_{16} orthogonal experiment table (BF+SF) and unconfined compressive strength values

Design No	Parameters				$\rho_{k,max}$ (g/cm³)	w_{opt} (%)	q_u (kPa)
	A	B	C	D			
	BF Length (mm)	BF content (%)	Curing Time (days)	SF content (%)			
1	6	0	1	0	1.394	32.0	296.7
2	6	0.5	7	5	1.380	32.6	874.9
3	6	1	28	10	1.360	32.7	1299.9
4	6	1.5	56	15	1.336	33.7	1326.5
5	12	0	7	10	1.358	33.0	1081.4
6	12	0.5	1	15	1.333	33.5	593.6
7	12	1	56	0	1.386	31.9	961.3
8	12	1.5	28	5	1.367	32.7	1157
9	18	0	28	15	1.340	33.7	1181.6
10	18	0.5	56	10	1.351	33.0	1521.3
11	18	1	1	5	1.375	32.7	558.5
12	18	1.5	7	0	1.379	31.8	761.6
13	24	0	56	5	1.374	32.7	881.5
14	24	0.5	28	0	1.384	31.9	847.3
15	24	1	7	15	1.339	33.8	1282.4
16	24	1.5	1	10	1.352	32.9	565.4

Table 4. Taguchi L_{16} orthogonal experiment table (BF+FA) and unconfined compressive strength values

Design No	Parameters				$\rho_{k,max}$ (g/cm³)	w_{opt} (%)	q_u (kPa)
	A	B	C	D			
	BF Length (mm)	BF content (%)	Curing Time (days)	FA content (%)			
1	6	0	1	0	1.394	32.0	296.7
2	6	0.5	7	5	1.349	33.0	813.7
3	6	1	28	10	1.340	33.5	1205.1
4	6	1.5	56	15	1.314	34.4	1194.0
5	12	0	7	10	1.338	33.7	1005.7
6	12	0.5	1	15	1.311	34.3	529.7
7	12	1	56	0	1.386	31.9	961.3
8	12	1.5	28	5	1.356	33.1	1064.4
9	18	0	28	15	1.320	34.3	1119.7
10	18	0.5	56	10	1.332	33.4	1354.0
11	18	1	1	5	1.350	33.0	513.8
12	18	1.5	7	0	1.379	31.8	761.6
13	24	0	56	5	1.356	33.2	819.8
14	24	0.5	28	0	1.384	31.9	847.3
15	24	1	7	15	1.311	34.4	1179.8
16	24	1.5	1	10	1.327	33.6	531.1

XRD and SEM analyses were carried out on the pure soil sample, Design 10 samples with the highest strengths, and Design 7 and Design 9 samples with lower strengths to determine the effect of changes in fly ash and silica fume content on microscopic structure. XRD and SEM analyzes were carried out on selected designs in order to observe and interpret the microstructures of pure and added soil samples prepared by compaction at optimum water content and maximum dry density (Figure 2). Additionally, the microstructures of soil-lime-basalt fiber and fly ash/silica fume mixtures were investigated in correlation with UCS results.

3. Results and discussion

The crystal phases determined by X-ray diffraction (XRD) analysis for different mixtures and the pure soil sample are given in Figure 3-8. The XRD analysis results revealed the formation of Tobermorite (calcium silicate/aluminate hydrate (C-S-H)) mineral due to the pozzolanic reaction of lime, fly ash, and silica fume additives. The calcium silicate hydrate peak formed at 29.7°, and according to the analysis results, the presence of quartz (SiO_2) has been observed in the samples. The density of quartz has decreased due to the pozzolanic

reaction of the additives. Tiwari et al. [25] reported that this situation is due to the consumption of quartz minerals during the C-S-H formation.

The density of the calcium silica/alumina hydrate mineral was found to be higher in the designs with the highest strength (design no 10). This indicates that the formation of hydrated gel during curing occurs at higher concentrations with longer curing times and the use of optimal strength parameter levels. XRD analysis results of Design-7 (3% lime and basalt fiber-added sample) and Design-10 (3% lime, 10% fly ash/silica fume and basalt fiber-added samples) cured for 56 days showed that the intensity of C-S-H mineral increased by adding fly ash or silica fume to the lime.

Figure 2. Performing XRD and SEM analyses.

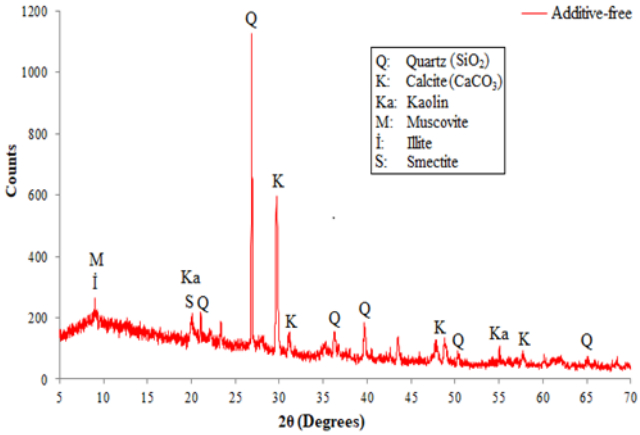


Figure 3. XRD analysis results of pure soil.

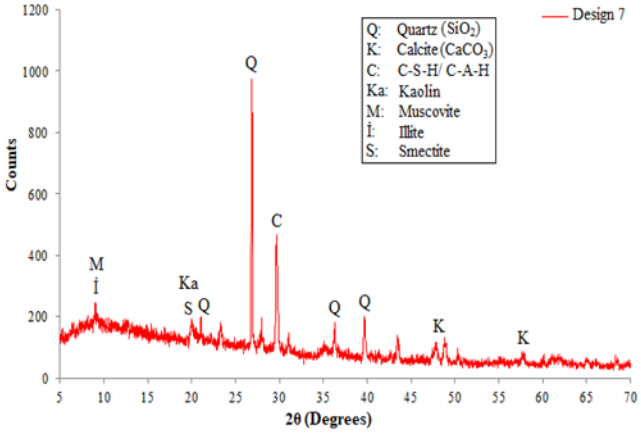


Figure 4. XRD analysis results of Design 7.

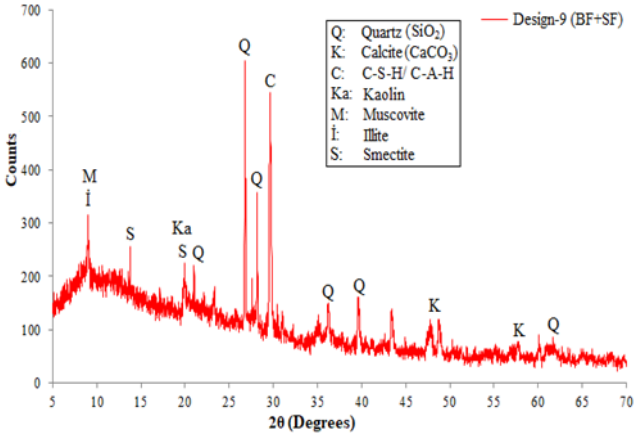


Figure 5. XRD analysis results of Design 9 (BF+SF).

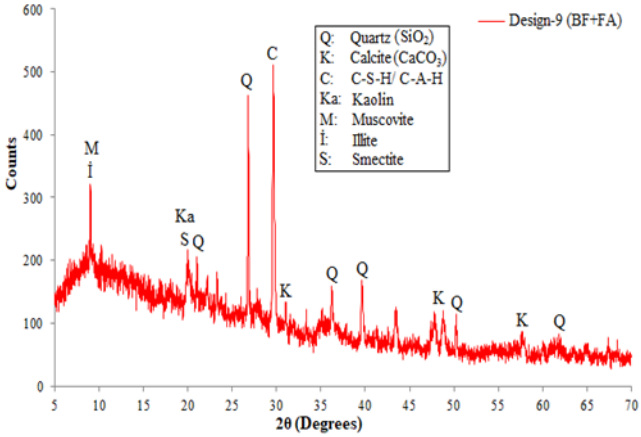


Figure 6. XRD analysis results of Design 9 (BF+FA).

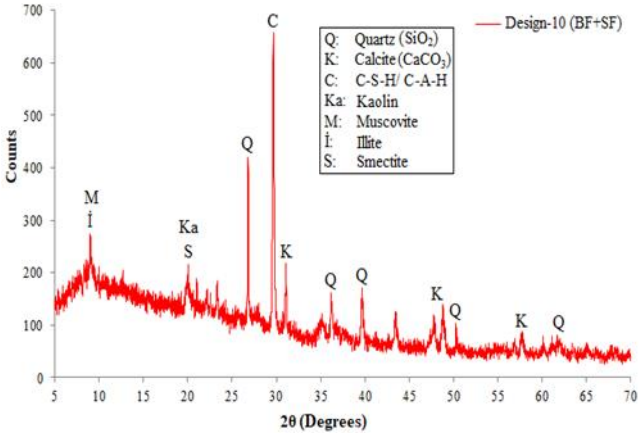


Figure 7. XRD analysis results of Design 10 (BF+SF).

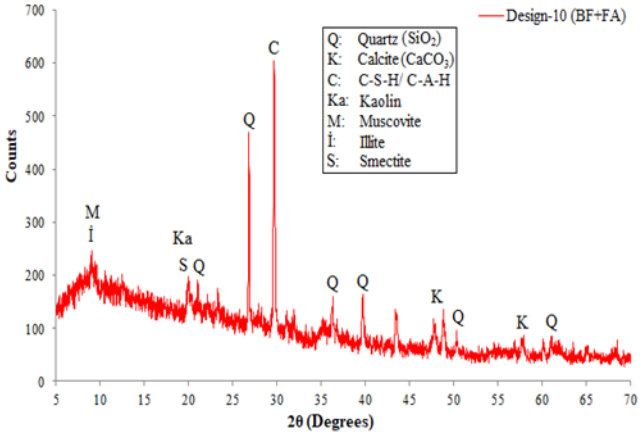


Figure 8. XRD analysis results of Design 10 (BF+FA).

SEM images obtained from selected design samples are given in Figure 9-14.

When the SEM images of the pure soil sample were examined, they showed a more dispersed and void

structure (Figure 9). When adding lime to the soil, the lime comes together to form a more cohesive mass due to pozzolanic and cation exchange reactions. The clayey soil exhibited a more granular structure with 3% lime contribution (Figure 10).

In addition to the lime additive, the addition of fly ash and silica fume led the particles to reorganize and increase structural integrity (Figure 11-13). This finding can be explained by the increased silica concentration due to the addition of both fly ash and silica fume and the rapid agglomeration of particles, leading to an increase in particle size. The addition of fly ash/silica fume developed the arrangement of clay particles to change from dispersed to lumped, resulting in the formation of some new cement-like compounds. The surfaces of the particles are generally coated with hydration gels and the soil pores are largely filled. This situation is caused by the beneficial effects of lime + fly ash or lime + silica fume additives on the formation of additional cementitious phase.

This improvement in soil properties is attributable to the quick development of the silicate gel and the increased synthesis of a new cementitious compound composed of calcium ions (Ca^{+2}) from lime and silicon dioxide (SiO_2) from silica fume [24, 26]. This can quickly fill soil voids and firmly interlock clay particles. As a result of the increased additive content, the soil's geotechnical properties improved. Increased lime and silica fume or lime and fly ash mixtures produced higher strength values than mixtures with 3% lime additives, owing to larger particle size and aggregation dimension.

SEM images show that soil particles adhere to the surface of basalt fiber, contributing to the force and friction between soil particles and basalt fiber (Figure 14). Tang et al. [27] reported that the contact area of fibers and surface roughness have significant effects on the micro-mechanical properties of the fiber-clay interface and that soil particles adhering to the fiber surface improve the interface bond.

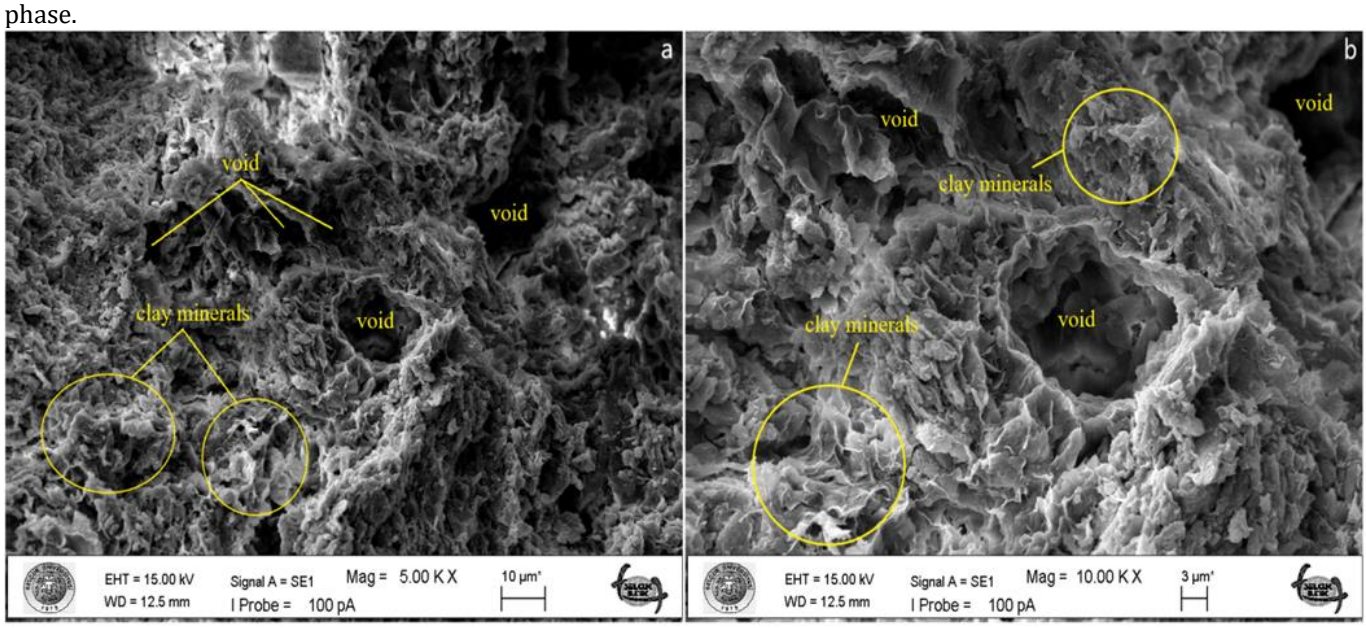


Figure 9. SEM image of clay soil with no additives; a. 5000X magnification b. 10000X magnification.

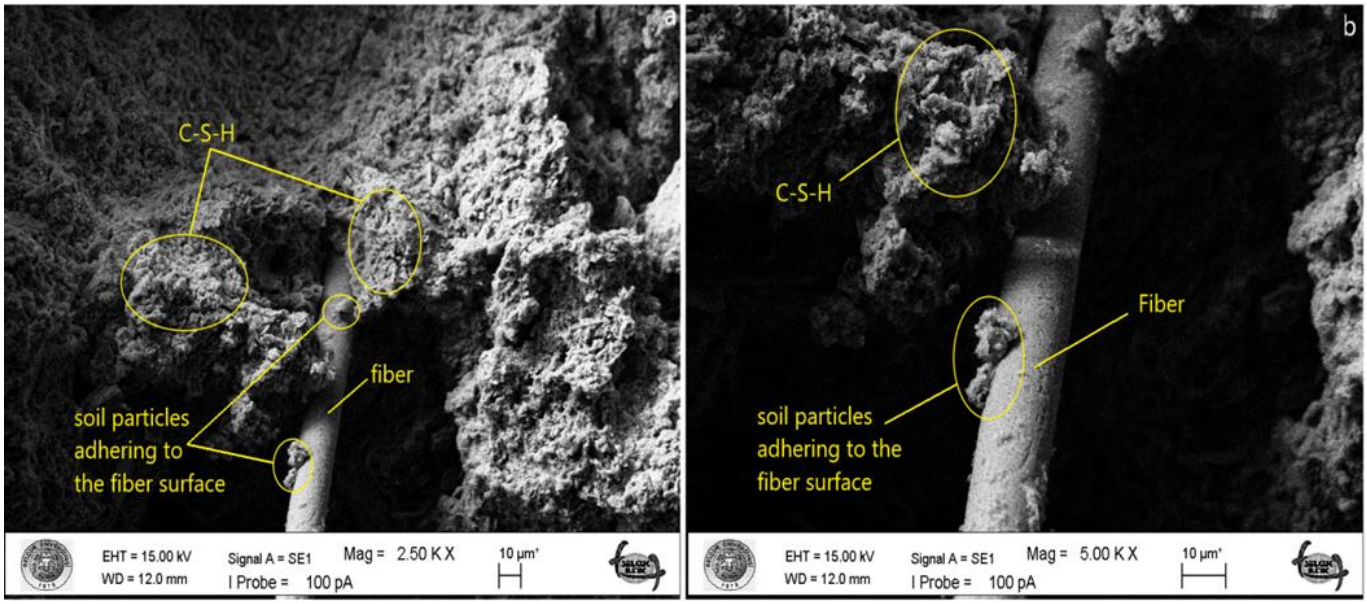


Figure 10. SEM image of design 7 (3%L+1%BF additives) a. 2500X magnification; b. 5000X magnification.

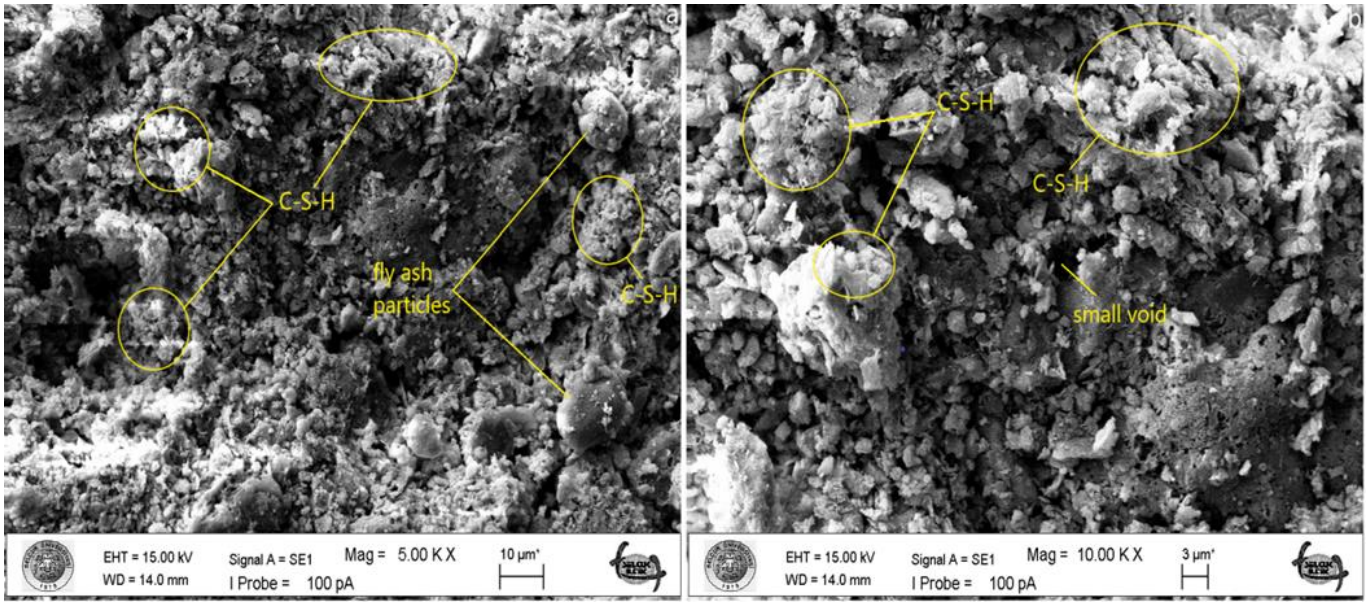


Figure 11. SEM image of design 9 (%3L+%15FA additives) a. 2500X magnification; b. 5000X magnification.

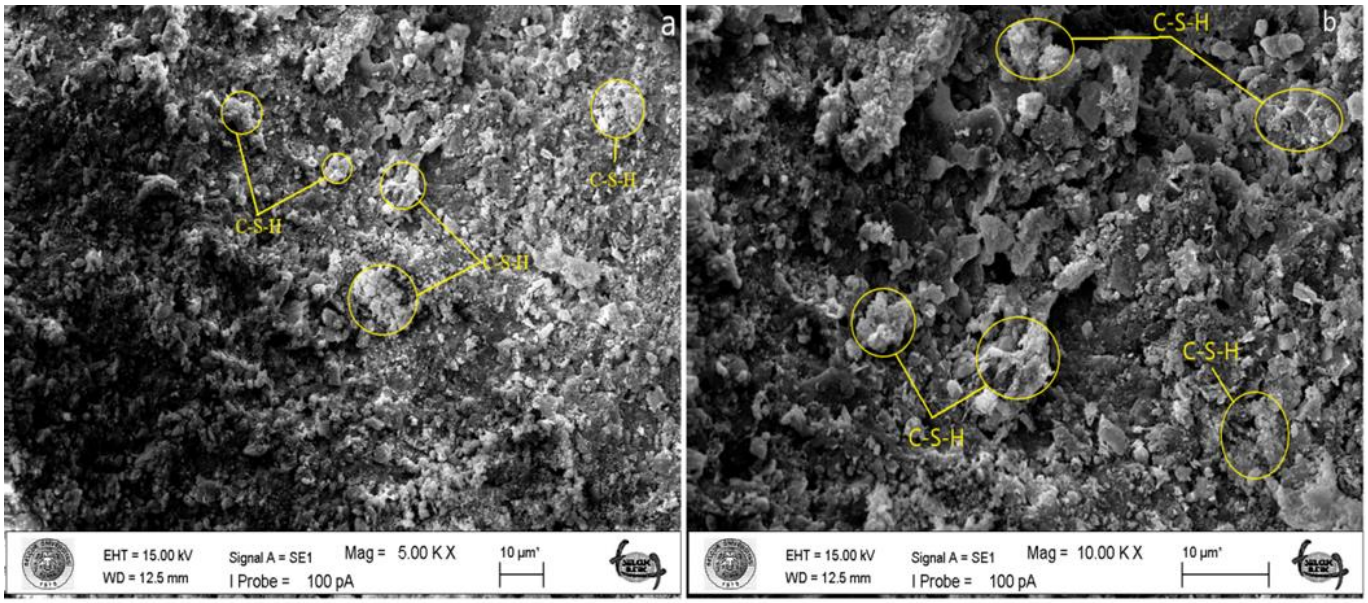


Figure 12. SEM image of design 9 (%3L+%15FA additives) a. 2500X magnification; b. 5000X magnification.

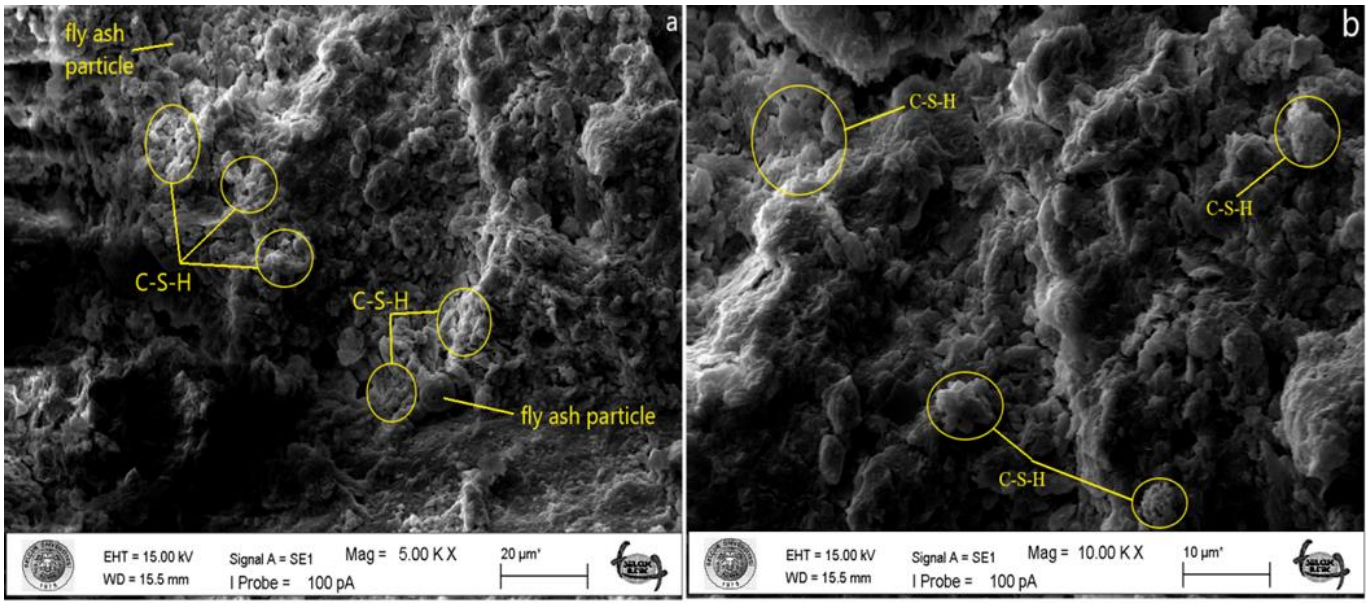


Figure 13. SEM image of design 9 (%3L+%15FA additives) a. 2500X magnification; b. 5000X magnification.

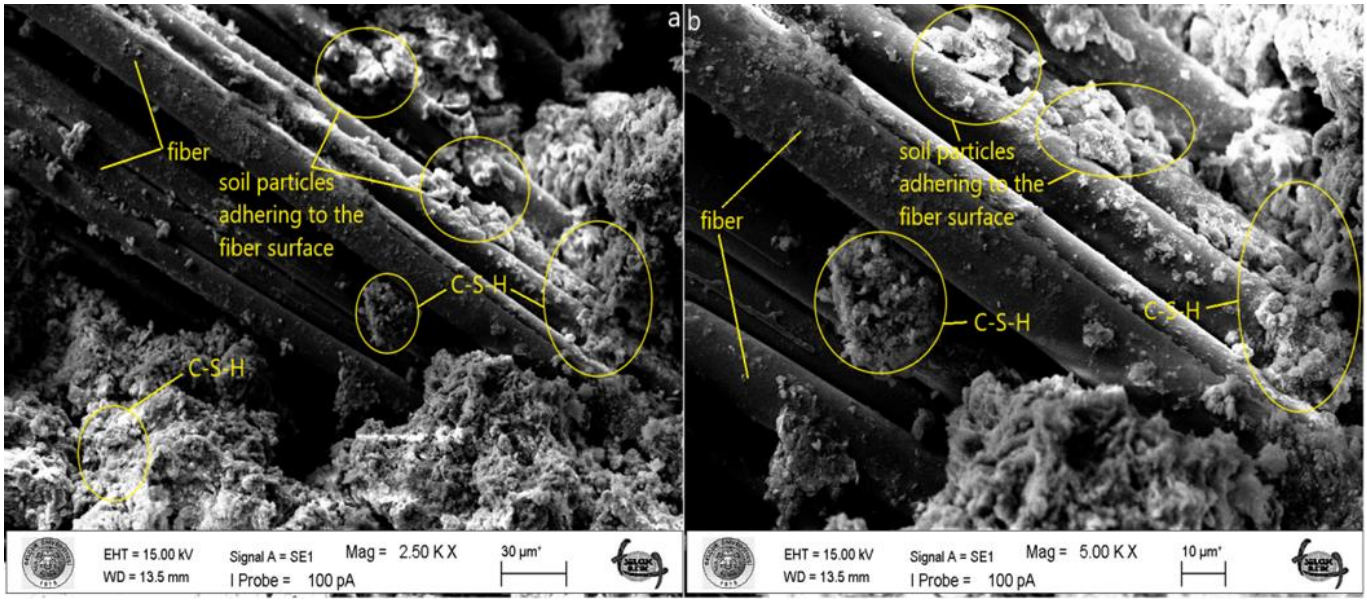


Figure 14. SEM image of design -10 (%3L+%0.5BF+%10SF additives) a. 2500X magnification; b. 5000X magnification.

In the area shown in Figure 14, it is observed that the basalt fiber surface, which is a good indicator of a strong bond between clay particles and basalt fibers, is held by many clay minerals. Due to the tightly reinforced interaction between basalt fibers and clay particles, they can collectively carry external loads, and the interfacial force between clay particles and basalt fibers gradually increases to resist the external load applied to the treated test specimens [28-29]. In this case, the surface roughness of the fiber plays an important role in improving the interfacial friction between the fiber and clay particles and enhances its resistance to external loading. Additionally, the main hydration products resulting from pozzolanic reactions (C-S-H gels) formed enhance the physical bonding strength among clay particles by cementing fine grains and filling pore voids. Hydration products also facilitate better adhesion of basalt fiber with the cementitious soil matrix.

When lime is added to clay, calcium ions react with clay minerals to form silicate gel. The filling effect of C-S-H also strengthens the bond between fiber and composite material, resulting in a more stable interfacial structure. Figure 14 clearly shows cementitious products on the fiber.

4. Conclusion

In this study, experimental studies were conducted to determine the effect of combinations of basalt fiber with fly ash and silica fume on the microstructure of a highly plastic clay. In order to increase the pozzolanic interaction, 3% lime was added to all mixtures in the experimental studies.

Based on SEM images, the untreated soil sample had a dispersed and porous structure as opposed to the treated combinations. Samples treated with just lime, lime combined with fly ash, and lime combined with silica fume showed a denser and more compact matrix structure, which was attributed to pozzolanic and cation exchange reactions. Particle surfaces were predominantly coated with hydration gels, which effectively filled soil pores. The images obtained supported the results of the strength tests.

SEM images revealed that soil particles adhered to the surface of the basalt fiber, enhancing the friction between the soil particles and the basalt fiber.

XRD analysis demonstrated the formation of calcium silica/alumina hydrate (C-S-H) minerals attributable to the pozzolanic reaction involving lime, fly ash, and silica fume additives. Moreover, the analysis results indicated the presence of quartz (SiO_2) in the samples. The density of quartz decreased due to the pozzolanic reaction induced by the additives.

The density of the calcium silicate/alumina hydrate mineral has been found to be higher in designs with high strengths. This situation indicates that during the curing period, the formation of hydrated gel occurs at higher concentrations with the use of optimum parameter levels for strength and an increase in curing time.

Author contributions

Atila Demiröz: Conceptualization, Methodology, Writing-Reviewing and Editing. **Onur Saran:** Laboratory Experiments, Writing-Original draft preparation, Software, Investigation.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Attom, M. F., Taqieddin, S. A., & Mubeideen, T. (2000). Shear strength and swelling stabilization of unsaturated clayey soil using pozzolanic material. Advances in Unsaturated Geotechnics, 275-288. https://doi.org/10.1061/40510(287)19

2. Ural, N. (2016). Effects of additives on the microstructure of clay. Road Materials and Pavement Design, 17(1), 104-119. https://doi.org/10.1080/14680629.2015.1064011
3. Indraratna, B., Vinod, J. S., & Athukorala, R. (2023). Chemically Treated Soils. DSC/HISS Modeling Applications for Problems in Mechanics, Geomechanics, and Structural Mechanics, 149-164.
4. Demiröz, A., Saran, O., & Hamed, E. A. A. (2023). The influence of various additives on the plasticity properties of clayey soil. Selcuk University Journal of Engineering Sciences, 22 (1), 38-42.
5. Mugambi, M. L. (2023). Evaluation of physico-chemical properties of expansive soils stabilized by limestone calcined clay cement. [Doctoral dissertation, Meru University of Science and Technology].
6. Al-Rawas, A. A., Hago, A. W., & Al-Sarmi, H. (2005). Effect of lime, cement and Sarooj (artificial pozzolan) on the swelling potential of an expansive soil from Oman. Building and Environment, 40(5), 681-687. https://doi.org/10.1016/j.buildenv.2004.08.028
7. Cuisinier, O., Auriol, J. C., Le Borgne, T., & Deneele, D. (2011). Microstructure and hydraulic conductivity of a compacted lime-treated soil. Engineering Geology, 123(3), 187-193. https://doi.org/10.1016/j.enggeo.2011.07.010
8. Davoudi, M. H., & Kabir, E. (2011). Interaction of lime and sodium chloride in a low plasticity fine grain soils. Journal of Applied Sciences, 11(2), 330-335. https://doi.org/10.3923/jas.2011.330.335
9. Jauberthie, R., Rendell, F., Rangeard, D., & Molez, L. (2010). Stabilisation of estuarine silt with lime and/or cement. Applied Clay Science, 50(3), 395-400. https://doi.org/10.1016/j.clay.2010.09.004
10. Prusinski, J. R., & Bhattacharja, S. (1999). Effectiveness of Portland cement and lime in stabilizing clay soils. Transportation Research Record, 1652(1), 215-227. https://doi.org/10.3141/1652-28
11. Bell, F. G. (1996). Lime stabilization of clay minerals and soils. Engineering Geology, 42(4), 223-237. https://doi.org/10.1016/0013-7952(96)00028-2
12. Chen, F. H. (2012). Foundations on expansive soils, 12. Elsevier.
13. Raj, P. P. (1999). Ground improvement techniques (HB). Firewall Media.
14. Metelková, Z., Boháč, J., Sedlářová, I., & Přikryl, R. (2011). Changes of pore size and of hydraulic conductivity by adding lime in compacting clay liners. Geotechnical Engineering: New Horizons, 93-98. https://doi.org/10.3233/978-1-60750-808-3-93
15. Muhmed, A., & Wanatowski, D. (2013). Effect of lime stabilisation on the strength and microstructure of clay. IOSR Journal of Mechanical and Civil Engineering, 6(3), 87-94.
16. Di Sante, M., Fratalocchi, E., Mazzieri, F., & Pasqualini, E. (2014). Time of reactions in a lime treated clayey soil and influence of curing conditions on its microstructure and behaviour. Applied Clay Science, 99, 100-109. https://doi.org/10.1016/j.clay.2014.06.018
17. Song, Y., Geng, Y., Dong, S., Ding, S., Xu, K., Yan, R., & Liu, F. (2023). Study on mechanical properties and microstructure of basalt fiber-modified red clay. Sustainability, 15(5), 4411. https://doi.org/10.3390/su15054411
18. Ma, Q. Y., Cao, Z. M., & Yuan, P. (2018). Experimental Research on Microstructure and Physical-Mechanical Properties of Expansive Soil Stabilized with Fly Ash, Sand, and Basalt Fiber. Advances in Materials Science and Engineering, 2018(1), 9125127. https://doi.org/10.1155/2018/9125127
19. ASTM D698 (2012). Standard test methods for laboratory compaction characteristics of soil using standard effort. West Conshohocken, USA, ASTM International.
20. Ayothiraman, R., & Singh, A. (2017). Improvement of soil properties by basalt fibre reinforcement. In Proc., DFI-PFSF Joint Conf. on Piled Foundations & Ground Improvement Technology for the Modern Building and Infrastructure Sector, 403-412.
21. Jia, Y., Zhang, J. S., Wang, X., Ding, Y., Chen, X. B., & Liu, T. (2022). Experimental study on mechanical properties of basalt fiber-reinforced silty clay. Journal of Central South University, 29(6), 1945-1956. https://doi.org/10.1007/s11771-022-5056-z
22. Taguchi, G. J. Q. R. (1987). Taguchi techniques for quality engineering. Quality Resources, New York.
23. Sharma, R. K., & Hymavathi, J. (2016). Effect of fly ash, construction demolition waste and lime on geotechnical characteristics of a clayey soil: a comparative study. Environmental Earth Sciences, 75, 1-11. https://doi.org/10.1007/s12665-015-4796-6
24. Türköz, M., Savaş, H., & Tasci, G. (2018). The effect of silica fume and lime on geotechnical properties of a clay soil showing both swelling and dispersive features. Arabian Journal of Geosciences, 11, 1-14. https://doi.org/10.1007/s12517-018-4045-x
25. Tiwari, N., Satyam, N., & Patva, J. (2020). Engineering characteristics and performance of polypropylene fibre and silica fume treated expansive soil subgrade. International Journal of Geosynthetics and Ground Engineering, 6(2), 18. https://doi.org/10.1007/s40891-020-00199-x
26. Wang, D., Abriak, N. E., & Zentar, R. (2013). Strength and deformation properties of Dunkirk marine sediments solidified with cement, lime and fly ash. Engineering Geology, 166, 90-99. https://doi.org/10.1016/j.enggeo.2013.09.007
27. Tang, C., Shi, B., Gao, W., Chen, F., & Cai, Y. (2007). Strength and mechanical behavior of short polypropylene fiber reinforced and cement stabilized clayey soil. Geotextiles and Geomembranes, 25(3), 194-202. https://doi.org/10.1016/j.geotexmem.2006.11.002
28. Ma, Q. Y., Cao, Z. M., & Yuan, P. (2018). Experimental Research on Microstructure and Physical-Mechanical Properties of Expansive Soil Stabilized with Fly Ash, Sand, and Basalt Fiber. Advances in Materials Science and Engineering, 2018(1), 9125127. https://doi.org/10.1155/2018/9125127
29. Wang, D., Wang, H., Larsson, S., Benzerzour, M., Maherzi, W., & Amar, M. (2020). Effect of basalt fiber

inclusion on the mechanical properties and microstructure of cement-solidified kaolinite. Construction and Building Materials, 241, 118085. https://doi.org/10.1016/j.conbuildmat.2020.118085

Turkish Journal of Engineering

https://dergipark.org.tr/en/pub/tuje

e-ISSN 2587-1366

Digital twin of multi-model drone detection system on Airsim for RF and vision modalities

Yusuf Özben [1], Süleyman Emre Demir [1], Hüseyin Birkan Yılmaz [*1]

[1] Bogaziçi University, Department of Computer Engineering, Türkiye, yusufozben@gmail.com, SuleymanD41k9@gmail.com, birkan.yilmaz@bogazici.edu.tr

Cite this study: Özben, Y., Demir, S. E., & Yılmaz, H. B. (2024). Digital twin of multi-model drone detection system on Airsim for RF and vision modalities. Turkish Journal of Engineering, 8 (3), 572-582

https://doi.org/10.31127/tuje.1436757

Keywords
Drone detection
Signal processing
Image processing
Machine learning
Simulation

Research Article

Received: 13.02.2024
Revised: 11.03.2024
Accepted: 12.03.2024
Published: 15.07.2024

Abstract

Drones have become more prevalent in recent years and are used for both beneficial and malicious purposes. As a result, protecting restricted areas from unauthorized drone activities has become crucial. However, some researchers face challenges in developing drone detection systems due to the high costs of necessary equipment. This paper presents an innovative solution by creating an Airsim Graphical User Interface (GUI) tool compatible with the Unreal Engine. This tool enables the simulation of drone flights and creation of image and radio frequency (RF) datasets for drone detection in a simulation environment. Our approach involves modeling the measurement devices such as cameras to capture image data and software defined radio (SDR) receiver to capture RF signals as raw in-phase and quadrature (IQ) data. Moreover, users can manage automated route planning for drones, recording configurations, and different cameras and RF configurations. Researchers can now generate datasets with various images and RF configurations without the need for physical drones, cameras, or SDRs, enabling experimentation with different drone detection models. Furthermore, we proposed models for drone detection systems by using generated datasets from the proposed dataset generation system.

1. Introduction

Drone usage increased dramatically all over the world. As of 15 January 2024, there are 863,728 registered drones in the US, according to the Federal Aviation Administration (FAA). Moreover, in Turkey, the number of registered drones reached 68,426 in June 2023. Drones have many useful applications like 3D modelling of historical places [1-5] and communication technologies [6], but they can also be used for illegal activities such as transporting prohibited substances and carrying out attacks. Automated drone detection systems using technologies like image, radar, and AI algorithms are needed to prevent these activities in certain areas like airports and government buildings.

Automated drone detection systems can be built using different techniques. However, in the literature, there are four main methods: vision, RF, acoustic, and radar. These methods have several advantages and disadvantages, as shown in Table 1. Multi-model drone detection systems that use two or more detection methods are built to eliminate disadvantages.

There are several research for drone detection systems in the literature. For these drone detection systems, different types of equipment like cameras, drones, and radars are used. However, there are some researchers who are unable to access these pieces of equipment. They generally use publicly available drone detection datasets. However, these publicly available datasets are created for specific tasks like classifying drones and birds in images or classifying three drone brands by using RF signals. Also, created labels could be limited for researchers. For example, a researcher could need Cartesian coordinates of drones in image datasets, but in these datasets generally only bounding boxes are published. In addition, a researcher could need a dataset that uses multiple drone detection methods. For example, we could not find any combined image and RF dataset in the literature.

In this paper, we proposed a solution for creating datasets for drone detection systems in a simulation environment without any other drone detection hardware requirement. The proposed system supports only image and RF dataset generation for now, and we plan to extend the simulation capabilities by adding other types of measurement devices. We chose image and RF dataset generation because they are the most popular two drone detection methods in the literature. The

created solution uses Unreal Engine for the simulation environment and Airsim for drone simulation. In addition, the Airsim GUI is proposed for managing drone detection data generation steps on a single screen. From the Airsim GUI, drones, cameras, RF, recording, and autonomous drone path settings can be configured, and datasets can be generated based on settings.

The remainder of the paper is organized as follows, Section 2 provides the literature review about related topics, Section 3 details the created dataset generation simulation architecture for drone detection, Section 4 gives example of possible use cases and drone detection models for simulation datasets, and Section 5 concludes the article, identifies the contributions, and discuss the future directions.

Table 1. Drone detection methods and their advantages and disadvantages.

Method	Advantages	Disadvantages
Camera	Cost-effective, provide visual evidence, etc.	Different scale and shape of drones, similarities with birds, different weather and lighting conditions, requires line-of-sight, etc.
RF	Resistance to different weather and lighting conditions, long range detection, no need for line-of-sight, etc.	Hard to differentiate from other RF resources, cannot detect completely autonomous drones, etc.
Acoustic	Easy to deploy, cost-effective, no need for line-of-sight, etc.	A lot of background noise, short range, etc.
Radar	Long range, resistance to autonomous drones, localization, etc.	Hard to detect small drones, high cost, etc.

2. Literature review

In the literature, several publicly published datasets exist for drone detection systems and proposed drone detection models. We summarize the literature for image and RF based datasets and drone detection methods, and drone detection in simulation environments.

2.1. Image-based drone detection datasets

In the literature, the most accessible datasets are image-based drone detection datasets than the other methods. The main reason is creating an image dataset is easier than the other detection methods.

There are various workshops available that focus on image-based drone detection. One of the most popular workshops is the Anti-UAV Workshop & Challenge. This challenge happened three times and there are a publicly published datasets for these challenges. The first challenge was organized in conjunction with Computer Vision and Pattern Recognition Conference (CVPR) as the first International Workshop on Anti-UAV Challenge [7], and 160 RGB and infrared video sequences were published. The 2nd anti-UAV Workshop & Challenge [8] was organized in conjunction with International Conference on Computer Vision (ICCV) 2021, and they extended their dataset with 280 HQ videos which are captured with a thermal infrared camera and contain different scale multiple drones. Finally, the last challenge [9] was organized in conjunction with CVPR 2023 with an extended dataset.

The Drone vs. Bird Detection Challenge is another well-known challenge for drone detection. This challenge began during the International Workshop on Small-Drone Surveillance, Detection, and Counteraction Techniques (WOSDETC) [10] at Conference on Advanced Video and Signal-Based Surveillance (AVSS) 2017 and has occurred five times since [11-14]. During these workshops, image datasets were published. The main focus of these datasets was detecting small drones and differentiating them from birds.

Li et al. [15] published a public dataset for reconstructing 3D flight trajectories by using multiple cameras. The dataset includes five sub-datasets, each with variations in the number of cameras, locations, and synchronization configurations.

2.2. RF-based drone detection datasets

For RF-based drone detection, there are no such workshops as in image-based drone detection. However, in the literature, there are several public datasets.

Allahham et al. [16] has one of the most popular RF datasets for drone detection. It contains time series RF data recorded using universal software radio peripheral (USRP) in 2.4 GHz. Collected RF data contains Parrot Bebop, Parrot AR, and DJI Phantom 3 drones and only background RF activity. They also categorize the drone data into the drone modes as on, hovering action, flying, and video recording. The dataset serves 2 classes for drone detection, 3 classes for drone type classification, and 10 classes for drone mode classification.

Swinney et al. [17] and Glüge et al. [18] publicly shared a dataset that contains raw IQ data. Swinney et al. [17] made available the DroneDetect V2 dataset, which includes drone signals with other signal interference in the same frequency band as Bluetooth. They used a BladeRF software-defined radio (SDR) and three drone models. The created class types are similar to those in Allahham et al. [16], with 2 classes for drone detection, 4 classes for drone type classification, and 10 classes for drone mode classification. On the other hand, the dataset published in Glüge et al. [18] not only contains drone signals but also remote controller signals. After recording the signals, they applied Labnoise and Gaussian noise to the drone signals.

The dataset published in Medaiyese et al. [19] is known as the Cardinal RF (CardRF) dataset and frequently used in research. This dataset's primary contribution is that it includes labeled drone RF data from both line-of-sight (LoS) and non-line-of-sight (NLOS) scenarios. Additionally, the dataset contains various types of drones and 2.4 GHz transmitters.

2.3. Drone detection via RF dataset

Because creating RF dataset for drone detection from scratch is exhausting, mainly publicly available datasets are used in the literature for drone detection research.

The DroneRF dataset in Allahham et al. [16] has been used in various research studies [20-25]. Al-Sa'd et al. [20] outline the dataset and uses a deep neural network for analysis. For the 2-class problem, they reported the average accuracy as 99.7%, for the 4-class problem 84.5%, and for the 10-class problem 46.8%. Al-Emadi and Al-Senaid [21] and Allahham et al. [22] propose a solution using a 1D convolutional neural network, achieving 59.2% and 87.4% accuracy in the 10-class problem, respectively. In addition, Medaiyese et al. [23] uses XGBoost, a more resource-friendly machine learning system than a deep neural network, to achieve 70.1% accuracy for the 10-class classification. Nemer et al. [24] proposed a hierarchical model that improves the 10-class problem using concatenated classifiers. The first set of classifiers addresses a 2-class drone detection problem. Following that, if a drone is detected, they proceed to solve a 3-class drone-type detection problem. Finally, the last set of classifiers is employed to tackle the task of identifying drone modes. Furthermore, they used two different classification methods for their classifiers: XGBoost and K-Nearest Neighbors (KNN). Ensemble learning is used for decision. Their model achieved 99.2% accuracy for the 10-class problem.

Glüge et al. [18] not only shares the dataset but also compares the performance of drone detection using IQ data and 2D spectrogram. They utilized a very deep convolutional network (VGGNet) and showed that a 2D spectrogram is more reliable in detecting drones in lower SNR conditions.

Nguyen et al. [26] propose a drone detection system using two different software-defined radios and two different drones. Their objective was to locate the drone and its controller in 3D space by combining angle of arrivals (AoA) and distances. The proposed system achieved a 12.2-degree error in identifying the drone's direction, with a localization error of 12.71 meters. Moreover, the system obtained a 9.9-degree error in identifying the controller's direction, and an accuracy of 11.36 meters in locating the controller.

2.4. Drone detection via camera

In the literature, there are several drone detections via image exist. They use both publicly available datasets and custom created datasets.

The YOLO series is a popular choice for detecting objects, including drones. Unlu et al. [27] use two cameras for drone detection. One camera had a dynamic platform and lens for a low-angle view, while the other had a static platform and lens for a wide-angle view. They detected objects with the wide-angle camera and classified them using the low-angle camera with 35X zoom. This approach allowed them to capture drones as small as 6x6 pixels using YOLOv3. They achieved a 0% false alarm rate and a 91% true positive rate. Singha and Aydin [28] used YOLOv4 for drone detection and obtained a 74.36% mAP50 score using a dataset that contained both bird and drone images. Zhai et al. [29], the latest version of YOLO, YOLOv8s (YOLOv8 small model), was used. The researchers modified the YOLOv8 and achieved even higher performance metrics. They achieved an 85.2% mAP score using the original version of YOLOv8s (YOLOv8 small model), which was improved to 95.3% with their optimized version of YOLOv8.

Seidaliyeva et al. [30] proposed a solution for static cameras using the drone-vs-bird dataset. The researchers divided the drone detection problem into two phases: moving object detection and classification. They employed the background subtraction method to detect moving objects and used MobileNetV2 to classify drones, birds, and backgrounds. For their evaluation, they reported an F1-score of 0.742 for IoU at 0.3.

2.5. Drone detection in simulation environments

Game engines are useful for drone detection tasks due to their reality. However, game engines mainly used for drone detection via images in the literature.

Drone detection by utilizing another drone's camera is investigated as a problem in Carrio et al. [31]. To address this issue, the authors propose a solution that employs a depth map. They test this approach in real-world settings after producing synthetic data in Unreal Engine. The same team continues their research in Carrio et al. [32] and improves their approach by using the same synthetic data generation method.

During the Drone-vs-Bird challenge held at AVSS in 2021, the winning team utilized a simulation environment to improve their score [33]. They employed a 3D drone model and diverse 2D backgrounds, which were incorporated into the Unity game engine. Subsequently, they captured images from the game engine and merged them with the competition dataset. The team achieved better results for drone detection using YOLOv5 by only adding synthetic data.

3. Dataset generation system

3.1. The simulation component architecture

Our dataset generation system works based on Unreal Engine, Airsim, and pyphysim, as shown in Figure 1. As a simulation environment, Unreal Engine is used. It is one of the most popular game engines. It has high-quality vision, audio, and programmability features. Also, it has built-in Airsim compatibility. Airsim is developed by Microsoft and it is an open-source, cross-platform simulator for cars and drones. The dataset generation system mainly uses drone and camera functionalities of Airsim.

In addition to Unreal Engine, Airsim GUI is the main component of the proposed architecture. We developed Airsim GUI to manage the simulation environment and generate image and RF datasets for drone detection models. It contains Airsim client library to communicate with Airsim. This communication is required to send commands to Airsim and collect data from Airsim. Collected data is processed and written into a file system to create a dataset. Another component in the Airsim GUI is pyphysim library. It is a Python library that provides simulation for digital communication physical layer. It is used to generate RF datasets.

Figure 1. The simulation component architecture.

3.2. The simulation environment: Boğaziçi University north campus

Some parts of the Boğaziçi University north campus are selected as a simulation environment. The selected simulation environment covers the library, computer engineering building, basketball court, portal statue, the pyramid, and the Atatürk bust. This area covers approximately 10300 m². Figure 2 shows the selected area's map image.

The simulation environment was created through two main steps: creating individual buildings and creating the complete environment. In the first step, chosen buildings were created for the simulation environment. In the second step, these created buildings are put in the simulation environment.

In the process of creating individual buildings, the DJI Mavic Air 2 drone was used to capture aerial pictures of the selected buildings. All areas of the buildings were scanned by drone, combined in Blender, and created in a 1:1 scale for simulation. The buildings created in the first step were exported as material to import Unreal Engine. They were combined in Unreal Engine and the selected simulation environment was created with a 1:1 scale. Figure 3 shows images from the simulation environment.

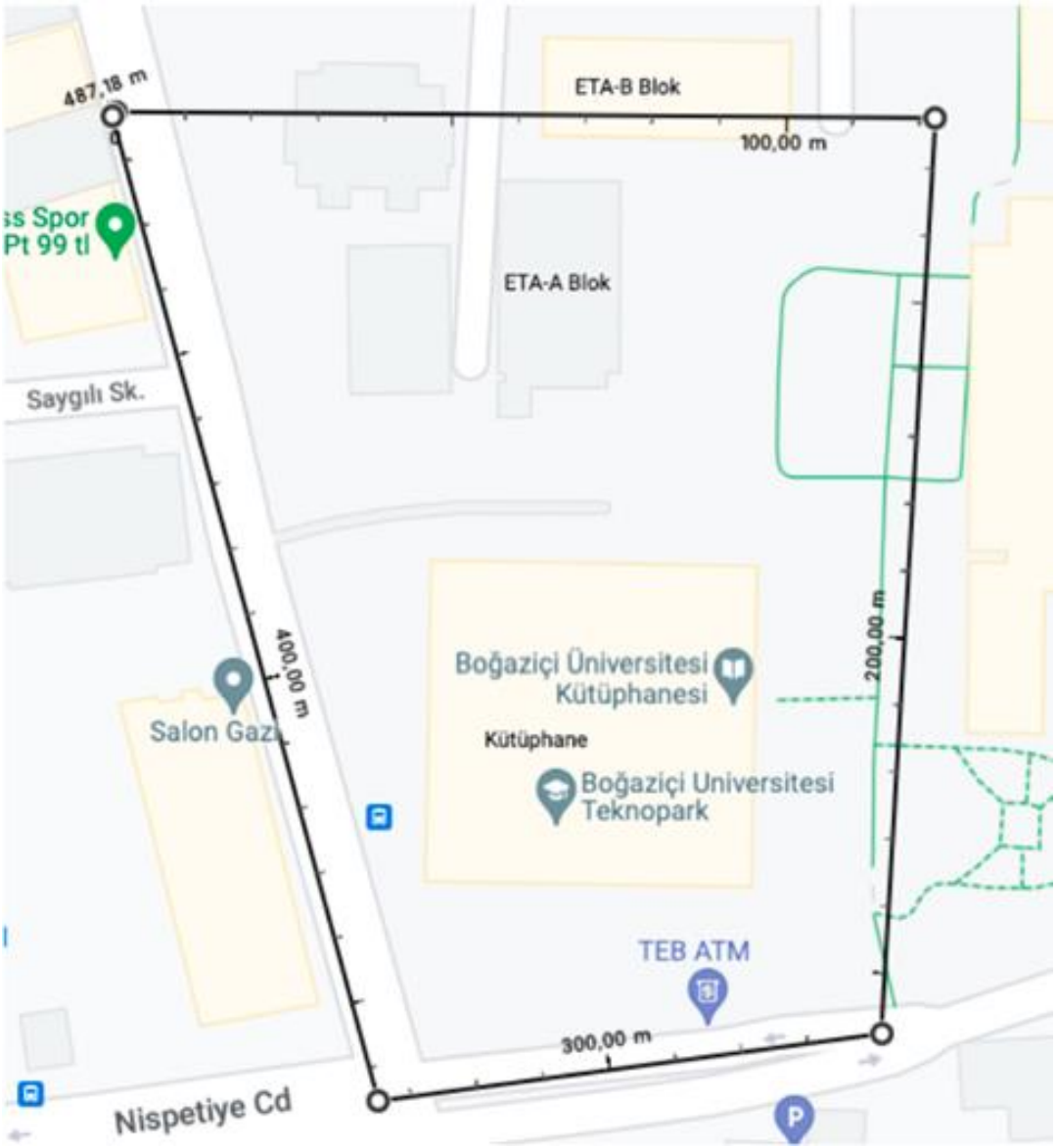


Figure 2. Map image of the selected simulation environment.

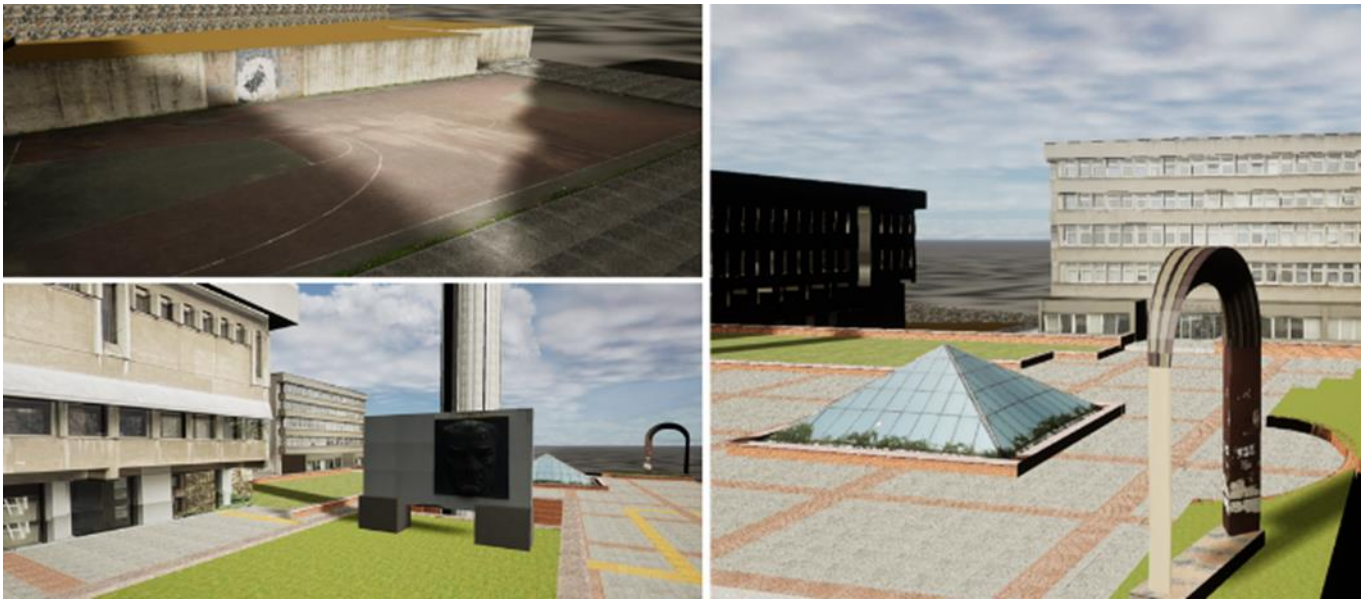


Figure 3. Screenshot images from the simulation environment. Top left image: The basketball court. Bottom left image: The Atatürk bust. Right image: The campus view from the Northeast.

3.3. Airsim GUI

For drone detection dataset generation, we created a GUI that communicates with the Airsim plugin in the Unreal Engine. We named this GUI as the Airsim GUI. This GUI was created with PyQt, a cross-platform GUI designer tool for Python. Communication between Airsim GUI and Airsim is performed by using the Airsim client library for Python. The Airsim client library has several APIs for several tasks.

There are several functions for managing the simulation environment and generating datasets in the Airsim GUI. Similar functions are put in the same tab to separate and group functionalities. Until now, six different tabs have been created: guide, drones, RF parameters, set drone path, recording, and offline data generation tabs.

- **Guide:** Directives are given in the guide tab. Also, it has an additional functionality to add a table component in a simulation environment. This table component provides locating multiple cameras with the exact relative location between them using only a single location and rotation information.
- **Drones:** The drones tab has spawning and retrieving drone functionalities. When spawning a drone, the drone's name and Cartesian coordinates for spawn location are taken as input. For retrieving a drone, the drone's name is selected from the list.
- **RF parameters:** RF parameters tab contains RF dataset-related inputs. If an RF dataset needs to be created, fields in this tab must be filled. These fields are mainly related to transmitter, channel, and SDRs. This tab will be described in Section 3.4 in detail.
- **Set drone path:** In the set drone path tab, drone paths can be defined. When creating a dataset, drones can follow these defined paths autonomously. There is a functionality to save and load created drone paths for generating another dataset with the same drone paths.
- **Recording:** The recording tab encompasses features for managing the recording process to generate datasets. Initially, a user chooses the cameras and their formats for recording. Then, the user can select or deselect the checkboxes for the recording functionalities. These checkboxes include "Save segmented images", "Follow path while recording", and "Generate RF after recording". By selecting the "Save segmented images" option, the drone segmentation can be recorded along with the camera images. These segmented images can later be used as a ground truth for drone detection models. Enabling the "Follow path while recording" ensures that the drones will follow the specified paths in the set drone path tab once the "Start Recording" button is clicked. If the user selects the "Generate RF after recording" option, the RF data generation process will start with the configured RF parameters after the simulation scenario finishes in the Unreal Engine. Once the image and RF datasets are generated, three main folders are created: "vehicle_logs", "camera_logs," and "rf_logs". The "vehicle_logs" folder contains individual files for each drone, named according to their drone names. These files contain logs of the drone's location, speed, and timestamps. The "camera_logs" folder includes separate files for each camera, named after their camera names. These files contain the image locations in the file system and timestamps of the captured images. The images themselves are stored in folders named after the respective cameras. The "rf_logs" folder stores the collected RF data from SDRs and the used RF parameters.
- **Offline data generation:** The final tab is the offline data generation. Since the RF generation step takes a long time, this process could be run later. Moreover, RF generation can be done multiple times with different parameters for the same scenario. This tab was created to generate RF datasets without running the same scenarios in Unreal Engine. Example images from the Airsim GUI are shown in Figure 4.

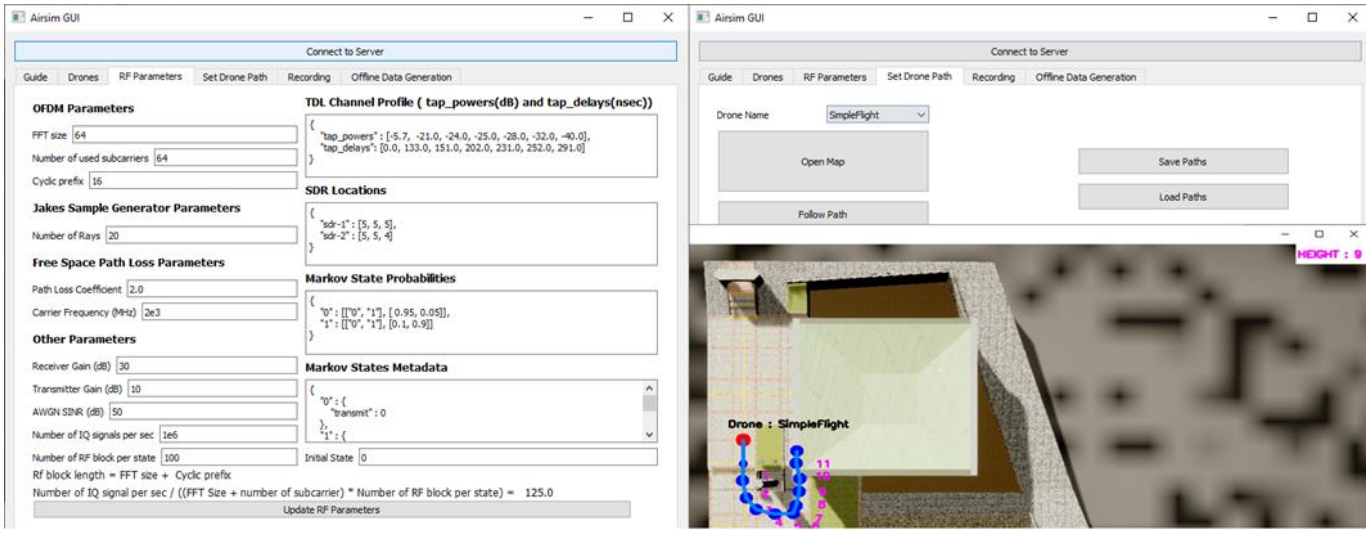


Figure 4. Example images from the Airsim GUI. (Left: RF parameters tab. Right:Set drone path tab.)

3.4. RF dataset generation

Figure 5 shows high-level RF dataset generation steps. RF dataset is generated with created files in the "vehicle_logs" folder. There is a file for every single drone which contains location and speed logs with timestamps. Until the AWGN step, RF data is generated separately for drones.

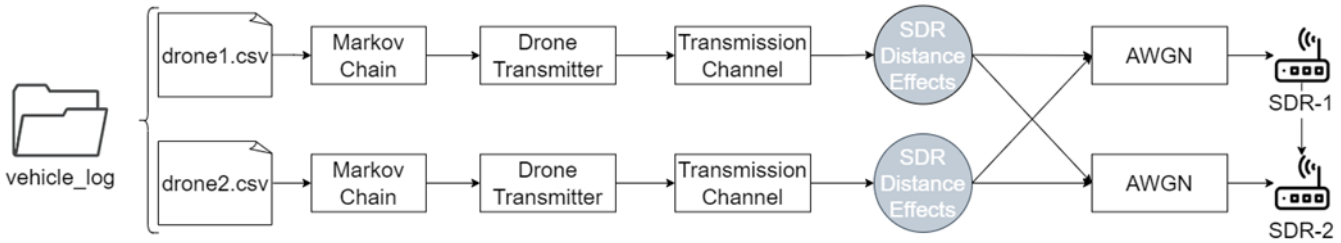


Figure 5. High-level RF dataset generation steps.

The pyphysim library was mainly used to generate RF dataset. The created RF datasets refer to collected IQ signals by SDRs in the simulation environment. SDR is the main component of RF signal-based drone detection systems. The RF dataset generation architecture supports multiple drones and SDRs. For this version of our data generation system, drone controllers are excluded since they are assumed to be out of coverage, as seen in Figure 6. Because of that, only drone-transmitted data is generated. SDR, drone transmitters, and other RF simulation configurations are set in the RF parameters tab of Airsim GUI. Then, the RF dataset can be created directly after recording or using previously created datasets with the offline data generation tab.

RF data generation steps start from reading drone log files, and at the end RF log files are created for every SDR. First of all, drone logs are read line by line. Between two lines, the time difference and the number of IQ signals that need to be generated between them is calculated. However, drones do not generate RF signals every time. To handle that Markov chain is used. Users can create a Markov chain to define idle (not transmitting) and busy (transmitting) states and probabilities of moving other states. Also, if the state represents busy, it contains transmitting power and one of the supported modulation techniques (Binary phase shift keying (BPSK) or quadrature phase shift keying (QPSK)). There is no limit to the created number of states. Figure 7 shows a Markov chain example containing data and probabilities of moving other states. A user defines the number of IQ signals in a single state from the Airsim GUI, and the number of generated states between two consecutive logs is calculated based on it. If the state is idle, there is no other process until the AWGN step.

If the state is busy, data transmission from a drone to SDR happens. The first block in this process is the drone transmitter block. It contains the processes that happen in the drone transmitter. First, the random number generator is run based on the modulation technique. If the modulation technique is BPSK, the generated numbers are zero and one. If the modulation technique is QPSK, the numbers range from zero to four. These generated numbers represent symbols. These represented symbols are transformed into complex signals in the IQ plane. Then, OFDM is applied to the signal with the FFT size, cyclic prefix size, and the number of subcarrier parameters taken from Airsim GUI. Finally, transmitting power and transmitter gain are applied to the signal transmitted to the transmission channel.

The transmission channel mainly contains two models: Jakes' channel model and tapped delay line model. In the transmission channel, only purely stochastic channel modeling is used. Jakes' model is used because it is famous for modeling mobile wireless communication systems realistic, contains effects of

scatters, and has a Doppler effect. Combining Jakes' model with the tapped delay line model provides a defining multipath tap delays and powers as an input.

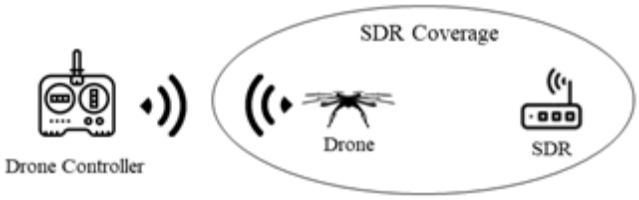


Figure 6. The SDR coverage in our simulation system shows that drone controllers are always out of coverage.

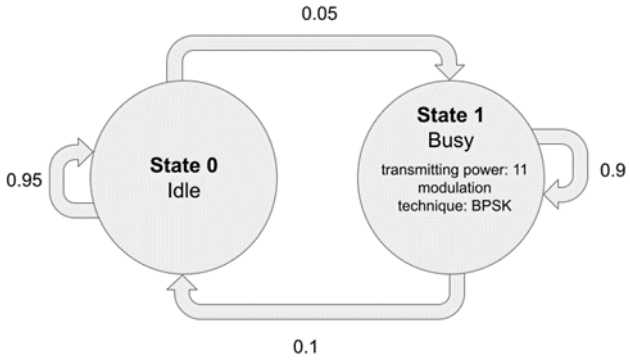


Figure 7. There are two states named idle and busy. In a busy state, transmission happens with 11 Watt and the BPSK modulation technique. Numbers on the arrows show the transition probabilities.

Because there is multiple SDR support, the same distance-related channel effects cannot be applied for every drone-SDR pair with a single calculation. So, for every SDR, these effects are applied separately for a drone transmission. These effects are path loss and phase shift. The free space path loss model is used as a path loss model. This model uses distance, carrier frequency, and path loss coefficient. Distance between a drone and SDR is calculated between two consecutive logs, and carrier frequency and path loss coefficient are taken from a user. Phase shift is applied based on the difference between the distances of drones and SDRs. After that, all drone signals generated simultaneously are summed together for SDR receivers.

At the end, for every SDR receiver, AWGN is added, and the receiver gain is applied. The generated signals are written as a byte stream to the "rf_logs" folder. Every SDR writes its data into a file with the name of the SDR. In addition to generated signals, the RF parameters used in dataset generation are also written in the "rf_parameters.pickle" file.

4. Example use cases

A dataset was created using the implemented simulation environment and the Airsim GUI to show example use cases. Different drone path scenarios, camera and SDR locations, and lighting conditions were chosen, and a dataset with image and RF was created with these configurations. In the simulation environment, two cameras, two SDRs, a drone, and ten birds were used. Between cameras and SDRs, there is 10 m. We proposed three different use cases: drone detection via image, drone detection via RF, and drone distance detection via RF.

The same hardware and software configurations were used in use cases as an 8GB Nvidia GTX 1070 graphic card, PyTorch 1.13.1 as the deep learning framework, Python version 3.8, CUDA version 11.0, and the operating system Windows 10.

4.1. Drone detection via image

Drone detection via image problem is a subset of object detection in computer vision. There are several object detection algorithms in the literature. YOLO is one of the most popular of them. It has different versions, and the state-of-the-art version is the YOLOv8 model. This model contains different sizes based on the number of parameters: nano(n), small(s), middle(m), large(l), and extra-large(x). Because the detection speed is crucial, achieving higher detection capability in smaller models is required. So, the nano version of YOLOv8 was chosen for drone detection.

Before training the model, the created dataset was prepared for the model. For ground truth, bounding boxes are required. However, Airsim GUI saves images with segmentation for ground truth. Because of that, segmented ground truths were transformed into bounding boxes. Then, the dataset was divided into training, validation, and test datasets as 60%-20%-20%, respectively.

The transfer learning was used for the training with a pre-trained model on the COCO dataset [34]. The model was trained to detect only drones. 1024 x 576 pixel image format was used as an input of the model. After 14 epochs, the training process was finished. The training metrics are given in Table 2.

Table 2. YOLOv8 nano model training process via image.

Epoch	Precision	Recall	mAP50	mAP50-95
1	0.76755	0.53322	0.55266	0.2614
2	0.61498	0.51329	0.53621	0.27792
3	0.74676	0.48494	0.54547	0.27878
4	0.80832	0.5604	0.61579	0.35433
5	0.68501	0.52492	0.5474	0.29289
6	0.80637	0.62261	0.67955	0.40068
7	0.85189	0.64014	0.68377	0.40726
8	0.82561	0.65275	0.72545	0.44725
9	0.84713	0.65615	0.70576	0.41265
10	0.91571	0.59552	0.71363	0.4251
11	0.87754	0.6711	0.73256	0.4228
12	0.85971	0.62292	0.707	0.41085
13	0.8656	0.63953	0.71531	0.44009
14	0.90008	0.65947	0.73839	0.45815

Used metrics in Table 2 are precision, recall, and mean average precision (mAP) scores. The mean average precision score is the primary performance evaluation metric for object detection tasks. The mAP50 is calculated by computing the average precision (AP) at a 50% intersection over union (IoU) threshold. The mAP50-95 scores are calculated by taking the average of precision values across a range of IoU thresholds starting from 0.5 and ending at 0.95, with a step size of 0.05. This calculation is performed as in Equation 1. The average precision is calculated based on the precision-recall

curve that corresponds to a specific IoU threshold. Equation 2 and 3 provide the formulas for precision and recall, in terms of False Negative (FN), True Positive (TP), and False Positive (FP). The IoU score is calculated by calculating the ratio of the intersection areas between the two bounding boxes and the area of their union, as shown in Equation 4.

$$mAP50-95 = \frac{1}{11}\sum_{t=0.5}^{0.95} mAPt \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$IoU = \frac{Area\ of\ Intersection}{Area\ of\ Union} \quad (4)$$

4.2. Drone detection via RF

As in Section 4.1, YOLOv8 was used for drone detection via RF. The RF data was transformed into images because YOLO was developed for vision tasks. First of all, SDR data is split into 0.25-second sequences. These sequences contain IQ signals. For every sequence, a short-time Fourier transform (STFT) is calculated. Because there were two different SDR in the simulation environment, both STFT results are concatenated vertically, as shown in Figure 8. In the created dataset, there are only cases that have drones. Because of that, cases that do not contain drones were generated using Gaussian noise, and the same processes were applied to these signals. All datasets are labeled with drone and background noise signals. Then, the dataset is split into training, validation, and test datasets as 60%-20%-20%, respectively.

The nano version of YOLOv8 is used for the classification task for this problem. Rather than an object detection task, the classification task classifies the whole image. Again, transfer learning was used for this model's training, and the pre-trained model with the ImageNet dataset [35] was used. After the first epoch, the accuracy metric remained relatively consistent, as indicated in Table 3. The third epoch provides the most accurate calculation; the corresponding accuracy is 0.92531. The accuracy indicates how often the network predicts the correct label with the highest probability.

Table 3. YOLOv8 nano model training process for drone detection via RF.

Epoch	Accuracy
1	0.92304
2	0.92485
3	0.92531
4	0.92531

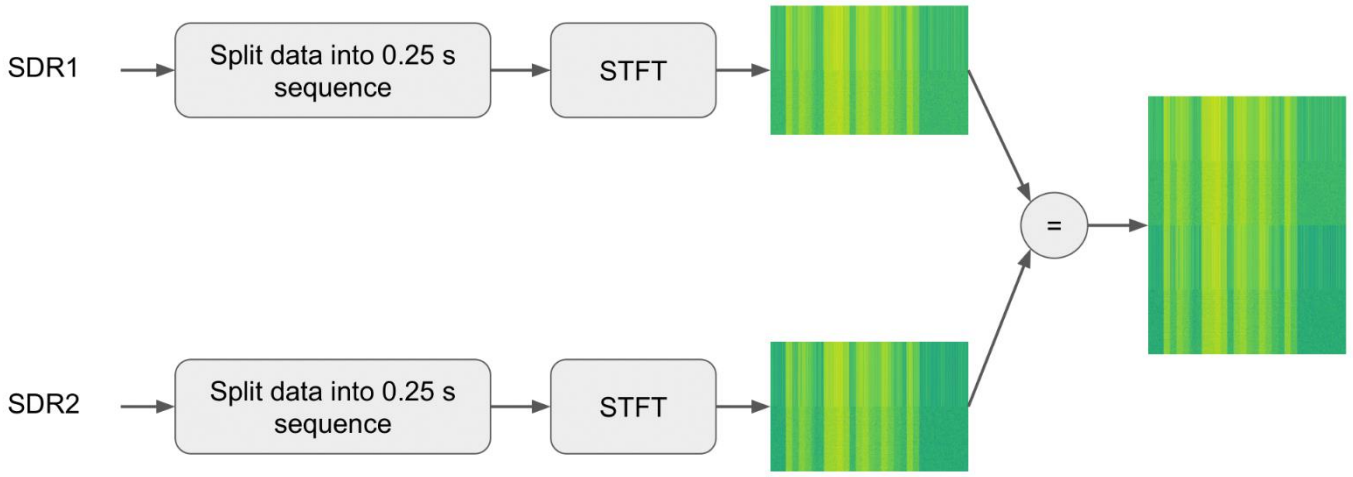


Figure 8. Data preprocess for RF data.

4.3. Drone distance detection via RF

In the final use case, drone distance detection via RF dataset problem was chosen. For this task, a custom convolutional neural network (CNN) was built to predict the distance between drones and SDRs. Because there are two SDRs in the simulation environment, the middle of the SDRs is selected as a SDR location. Two different sequence lengths, 0.25 seconds and 0.50 seconds, are used to compare performances.

For data preprocessing, the SDR signals were divided into 0.25-second and 0.50-second sequences and labeled with the average distance in the sequences. The sampling frequency was 10^6 in the dataset. This means that each 0.25-second sequence contains 250,000 time points, and a 0.50-second sequence contains 500,000 time points. I and Q components are split for both SDRs and merged vertically to provide input for the CNN model. 4x250000-shaped and 4x500000-shaped arrays were created as input for the custom CNN models. The datasets for different sequence lengths were divided into training, validation, and test datasets as 60%-20%-20%, respectively.

The proposed CNN models are shown in Figure 9. The proposed models contain three 2D convolution layers and two 2D average pooling layers. To prevent

overfitting, two dropout layers with a 20% probability are added. After the convolution layers, fully connected layers were added to the end. Because 0.50 second sequences contain more data, one more dense layer is added to its model.

Both models are trained with their datasets separately with the same hyperparameters. For the training, the batch size is chosen as 10. Rectified linear units are used as an activation layer. To converge the optimal solution, Adam optimizer is used, and the learning rate is chosen as 0.001. The best results for 0.25 second sequence were achieved in the 8th epoch and 0.50 second sequence were achieved in the 10th. The test result statistics are given in Table 4. In the given statistics, the errors in min 10% and max 10% were trimmed. The root mean square error for the test dataset is 7.29 meters for 0.25 second sequence and 6.07 meters for 0.50 second sequence. As a result, broader sequences provide better performance for the drone distance detection task.

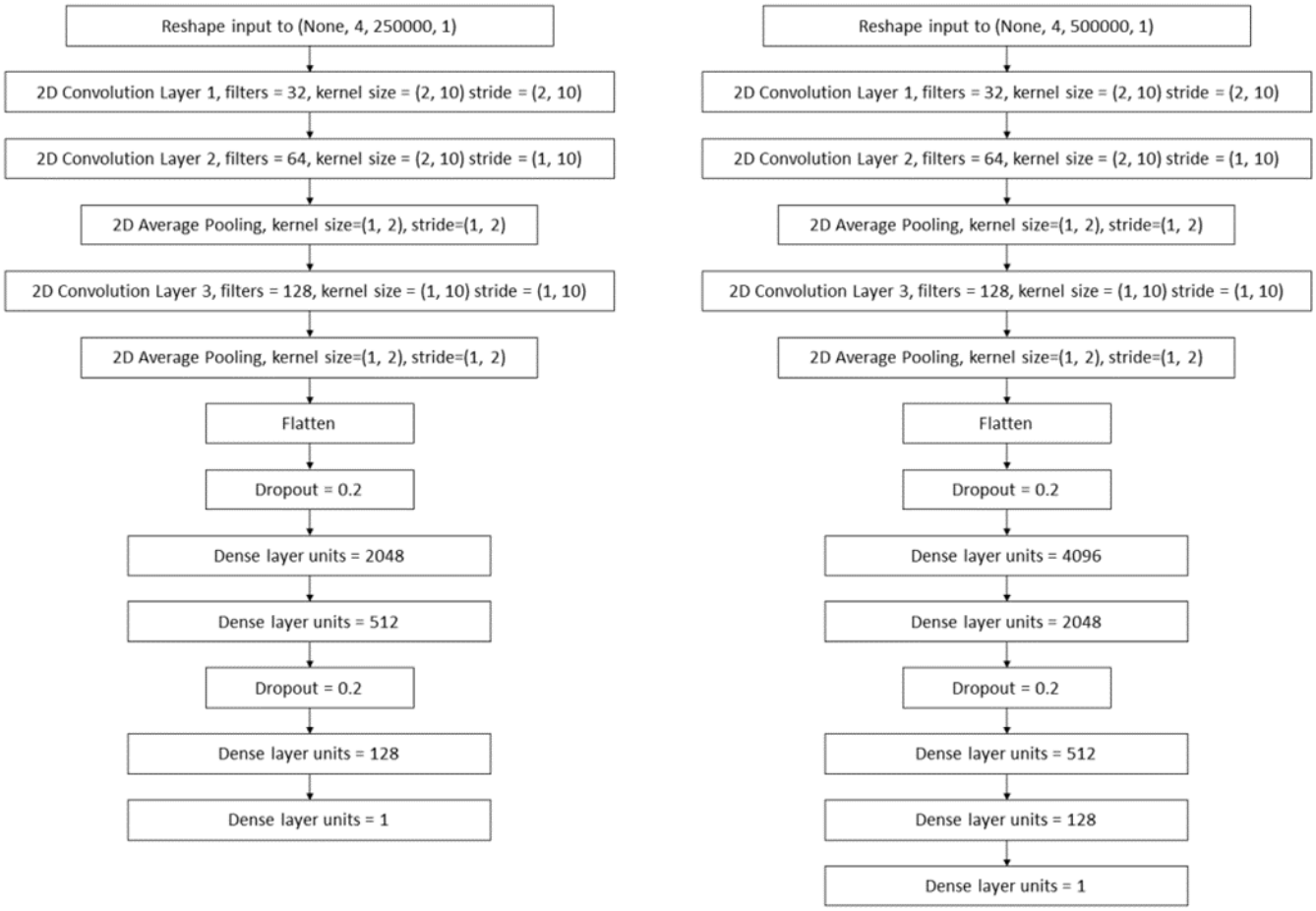


Figure 9. Custom CNN models to predict the distance. (Left is for 0.25 second and right is for 0.50 second sequences).

Table 4. Drone distance detection statistics for the absolute difference and percentage error between ground truth and estimations. For statistics, we trimmed rows that are in the min 10% and max 10% percentage error.

0.25 second sequences (RMSE=7.29) (Epoch 8)	Absolute difference (m)	mean	std	min	25%	50%	75%	max
		6.26	3.73	0.93	3.16	5.70	8.54	17.98
	Percentage error (%)	mean	std	min	25%	50%	75%	max
		14.82	7.73	3.33	8.20	14.10	20.22	34.23
0.50 second sequences (RMSE=6.07) (Epoch 10)	Absolute difference (m)	mean	std	min	25%	50%	75%	max
		5.32	2.93	0.64	2.94	4.65	7.16	14.96
	Percentage error (%)	mean	std	min	25%	50%	75%	max
		12.50	5.53	3.17	7.92	11.76	16.72	25.71

5. Conclusion

In this study, we have proposed a comprehensive solution for generating datasets for drone detection models in a simulation environment. We described the proposed architecture and created the Airsim GUI tool. Then, the implemented structure for creating RF dataset is explained. Moreover, three different drone detection models were tested on the generated dataset from the proposed architecture.

As a future work, we aim to enhance the RF dataset generation architecture in the Airsim GUI. We want to incorporate LOS and NLOS channel models. Also, we want to add dataset generation features for other modalities like acoustics and radar.

Acknowledgement

This work has been supported by Boğaziçi University BAP under the grant BAP-SUP-17862

Author contributions

Yusuf Özben: Methodology, Software, Tests, Writing-Original draft preparation. **Süleyman Emre Demir:** Methodology, Software, Tests. **Hüseyin Birkan Yılmaz:** Methodology, Visualization, Investigation, Writing-Reviewing and Editing.

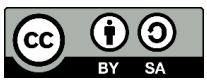
Conflicts of interest

The authors declare no conflicts of interest.

References

1. Şasi, A., & Yakar, M. (2017). Photogrammetric modelling of Sakahane Masjid using an unmanned aerial vehicle. Turkish Journal of Engineering, 1(2), 82-87. https://doi.org/10.31127/tuje.316675
2. Karataş, L., Alptekin, A., Kanun, E., & Yakar, M. (2022). Tarihi kârgir yapılarda taş malzeme bozulmalarının İHA fotogrametrisi kullanarak tespiti ve belgelenmesi: Mersin Kanlıdivane ören yeri vaka çalışması. İçel Dergisi, 2(2), 41-49.
3. Karataş, L., Alptekin, A., Karabacak, A., Yakar, M. (2022). Detection and documentation of stone material deterioration in historical masonry buildings using UAV photogrammetry: A case study of Mersin Sarisih Inn. Mersin Photogrammetry Journal, 4(2), 53-61. https://doi.org/10.53093/mephoj.1198605
4. Şasi, A., & Yakar, M. (2018). Photogrammetric modelling of Hasbey Dar'ülhuffaz (Masjid) using an unmanned aerial vehicle. International Journal of Engineering and Geosciences, 3(1), 6-11. https://doi.org/10.26833/ijeg.328919
5. Çolak, A., Aktan, N., & Yılmaz, H. M. (2022). Modelling of its surroundings and Selime Cadhetral by UAV data. Advanced UAV, 2(1), 24-28.
6. Banafaa, M., Pepeoğlu, Ö., Shayea, I., Alhammadi, A., Shamsan, Z., Razaz, M. A., ... & Al-Sowayan, S. (2024). A comprehensive survey on 5G-and-beyond networks with UAVs: Applications, emerging technologies, regulatory aspects, research trends and challenges. IEEE Access, 12, 7786 – 7826. https://doi.org/10.1109/ACCESS.2023.3349208
7. Jiang, N., Wang, K., Peng, X., Yu, X., Wang, Q., Xing, J., ... & Han, Z. (2021). Anti-uav: a large-scale benchmark for vision-based uav tracking. IEEE Transactions on Multimedia, 25, 486-500. https://doi.org/10.1109/TMM.2021.3128047
8. Zhao, J., Wang, G., Li, J., Jin, L., Fan, N., Wang, M., ... & Guo, Y. (2021). The 2nd anti-uav workshop & challenge: Methods and results. Computer Vision and Pattern Recognition. https://doi.org/10.48550/arXiv.2108.09909
9. Zhao, J., Li, J., Jin, L., Chu, J., Zhang, Z., Wang, J., ... & Shengmei, J. S. (2023). The 3rd anti-uav workshop & challenge: Methods and results. Computer Vision and Pattern Recognition. https://doi.org/10.48550/arXiv.2305.07290
10. Coluccia, A., Ghenescu, M., Piatrik, T., De Cubber, G., Schumann, A., Sommer, L., Klatte, J., Schuchert, T., Beyerer, J., Farhadi, M., Amandi, R., Aker, C., Kalkan, S., Saqib, M., Sharma, N., Daud, S., Makkah, K., & Blumenstein, M. (2017). Drone-vs-bird detection challenge at IEEE AVSS2017. IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) Drone-vs-Bird detection challenge, 1-6, Lecce, Italy.
11. Coluccia, A., Fascista, A., Schumann, A., Sommer, L., Dimou, A., Zarpalas, D., ... & Mantegh, I. (2017). Drone-vs-bird detection challenge at IEEE AVSS2017. 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 1-8. https://doi.org/ 10.1109/AVSS.2017.8078464
12. Coluccia, A., Fascista, A., Schumann, A., Sommer, L., Dimou, A., Zarpalas, D., ... & Rajashekar, S. (2021). Drone vs. bird detection: Deep learning algorithms and results from a grand challenge. Sensors, 21(8), 2824. https://doi.org/10.3390/s21082824
13. Coluccia, A., Fascista, A., Schumann, A., Sommer, L., Dimou, A., Zarpalas, D., ... & Mantegh, I. (2021). Drone-vs-bird detection challenge at IEEE AVSS2021. 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 1-8. https://doi.org/10.1109/AVSS52988.2021.9663844
14. Coluccia, A., Fascista, A., Schumann, A., Sommer, L., Dimou, A., Zarpalas, D., ... & Pavleski, D. (2022). Drone-vs-bird detection challenge at ICIAP 2021. In International Conference on Image Analysis and Processing, 410-421. https://doi.org/10.1007/978-3-031-13324-4_35
15. Li, J., Murray, J., Ismaili, D., Schindler, K., & Albl, C. (2020). Reconstruction of 3D flight trajectories from ad-hoc camera networks. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 1621-1628. https://doi.org/10.1109/IROS45743.2020.9341479
16. Allahham, M. S., Al-Sa'd, M. F., Al-Ali, A., Mohamed, A., Khattab, T., & Erbad, A. (2019). DroneRF dataset: A dataset of drones for RF-based detection, classification and identification. Data in Brief, 26, 104313. https://doi.org/10.1016/j.dib.2019.104313
17. Swinney, C. J., & Woods, J. C. (2021). RF detection and classification of unmanned aerial vehicles in environments with wireless interference. International Conference on Unmanned Aircraft Systems (ICUAS), 1494-1498. https://doi.org/10.1109/ICUAS51884.2021.9476867
18. Glüge, S., Nyfeler, M., Ramagnano, N., Horn, C., & Schüpbach, C. (2023). Robust drone detection and classification from radio frequency signals using convolutional neural networks. 15th International Joint Conference on Computational Intelligence (IJCCI), 496-504. https://doi.org/10.5220/0012176800003595
19. Medaiyese, O., Ezuma, M., Lauf, A., & Adeniran, A. (2022). Cardinal RF (CardRF): An outdoor UAV/UAS/drone RF signals with Bluetooth and WiFi signals dataset.
20. Al-Sa'd, M. F., Al-Ali, A., Mohamed, A., Khattab, T., & Erbad, A. (2019). RF-based drone detection and

identification using deep learning approaches: An initiative towards a large open source drone database. Future Generation Computer Systems, 100, 86-97. https://doi.org/10.1016/j.future.2019.05.007
21. Al-Emadi, S., & Al-Senaid, F. (2020). Drone detection approach based on radio-frequency using convolutional neural network. 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), 29-34. 10.1109/ICIoT48696.2020.9089489
22. Allahham, M. S., Khattab, T., & Mohamed, A. (2020). Deep learning for RF-based drone detection and identification: A multi-channel 1-D convolutional neural networks approach. IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), 112-117. https://doi.org/10.1109/ICIoT48696.2020.9089657
23. Medaiyese, O. O., Syed, A., & Lauf, A. P. (2021). Machine learning framework for RF-based drone detection and identification system. 2nd International Conference on Smart Cities, Automation & Intelligent Computing Systems (ICON-SONICS), 58-64. https://doi.org/10.1109/ICON-SONICS53103.2021.9617168
24. Nemer, I., Sheltami, T., Ahmad, I., Yasar, A. U. H., & Abdeen, M. A. (2021). RF-based UAV detection and identification using hierarchical learning approach. Sensors, 21(6), 1947. https://doi.org/10.3390/s21061947
25. Zhao, Z., Du, Q., Yao, X., Lu, L., & Zhang, S. (2023). A Two-Dimensional Deep Network for RF-based Drone Detection and Identification Towards Secure Coverage Extension. In 2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall), 1-5. https://doi.org/10.1109/VTC2023-Fall60731.2023.10333485
26. Nguyen, P., Kim, T., Miao, J., Hesselius, D., Kenneally, E., Massey, D., ... & Vu, T. (2019). Towards RF-based localization of a drone and its controller. Proceedings of the 5th Workshop on Micro Aerial Vehicle Networks, Systems, and Applications, 21-26. https://doi.org/10.1145/3325421.3329766
27. Unlu, E., Zenou, E., Riviere, N., & Dupouy, P. E. (2019). Deep learning-based strategies for the detection and tracking of drones using several cameras. IPSJ Transactions on Computer Vision and Applications, 11, 1-13. https://doi.org/10.1186/s41074-019-0059-x
28. Singha, S., & Aydin, B. (2021). Automated drone detection using YOLOv4. Drones, 5(3), 95. https://doi.org/10.3390/drones5030095
29. Zhai, X., Huang, Z., Li, T., Liu, H., & Wang, S. (2023). YOLO-Drone: an optimized YOLOv8 network for tiny UAV object detection. Electronics, 12(17), 3664. https://doi.org/10.3390/electronics12173664
30. Seidaliyeva, U., Akhmetov, D., Ilipbayeva, L., & Matson, E. T. (2020). Real-time and accurate drone detection in a video with a static background. Sensors, 20(14), 3856. https://doi.org/10.3390/s20143856
31. Carrio, A., Vemprala, S., Ripoll, A., Saripalli, S., & Campoy, P. (2018). Drone detection using depth maps. In 2018 IEEE/RSJ international conference on intelligent robots and systems (IROS), 1034-1037. https://doi.org/10.1109/IROS.2018.8593405
32. Carrio, A., Tordesillas, J., Vemprala, S., Saripalli, S., Campoy, P., & How, J. P. (2020). Onboard detection and localization of drones using depth maps. IEEE Access, 8, 30480-30490. https://doi.org/10.1109/ACCESS.2020.2971938
33. Akyon, F. C., Eryuksel, O., Ozfuttu, K. A., & Altinuc, S. O. (2021). Track boosting and synthetic data aided drone detection. 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 1-5. https://doi.org/10.1109/AVSS52988.2021.9663759
34. Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. Computer Vision–ECCV 2014: 13th European Conference, Zurich, 13, 740-755. https://doi.org/10.1007/978-3-319-10602-1_48
35. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248-255. https://doi.org/10.1109/CVPR.2009.5206848



Turkish Journal of Engineering

https://dergipark.org.tr/en/pub/tuje

e-ISSN 2587-1366

A predictive machine learning framework for diabetes

Danjuma Maza [*1], **Joshua Olufemi Ojo** [1], **Grace Olubumi Akinlade** [1]

[1] *Obafemi Awolowo University Ile-Ife, Department of Physics and Engineering Physics, Nigeria, mazadd@oauife.edu.ng, jojo@oauife.edu.ng, gakinlade@oauife.edu.ng*

Cite this study: Maza, D., Ojo, J. O., & Akinlade, G. O. (2024). A predictive machine learning framework for diabetes. Turkish Journal of Engineering, 8 (3), 583-592

https://doi.org/10.31127/tuje.1434305

Keywords
Classification
Diabetes
Prediction
Accuracy
Recall

Research Article

Received: 09.02.2024
Revised: 16.04.2024
Accepted: 17.04.2024
Published: 15.07.2024

Abstract

Diabetes, a non-communicable disease, is associated with a condition indicative of too much glucose in the bloodstream. In the year 2022, it was estimated that about 422 million were living with the disease globally. The impact of diabetes on the world economy was estimated at $ 1.31 trillion in the year 2015 and implicated in the death of 5 million adults between the ages of 20 and 79 years globally. If left untreated for an extended time, could result in a host of other health complications. The need for predictive models to supplement the diagnostic process and aid the early detection of diabetes is therefore important. The current study is an effort geared toward developing a machine learning framework for the prediction of diabetes, expected to aid medical practitioners in the early detection of the disease. The dataset used in this investigation was sourced from the Kaggle database. The dataset consists of 100,000 entries, with 8,500 diabetics and 91,500 non-diabetics, indicating an imbalanced dataset. The dataset was modified to achieve a more balanced dataset consisting of 8,500 entries each for the diabetic and non-diabetic classes. Gradient Boosting classifier (GBC), Adaptive Boosting classifier (ADA), and Light Gradient Boosting Machine (LGBM) were the best three performing classifiers after comparing fifteen classifiers. The proposed framework is a stack model consisting of GBC, ADA, and LGBM. The ADA classifier was utilized as the meta-model. This model achieved an average accuracy, area under the curve (AUC), recall, precision, and f1-score of 91.12 ± 0.75 %, 97.83 ± 0.29 %, 92.03 ± 1.55 %, 90.40 ± 1.01 %, and 91.12 ± 0.77 %, respectively. The selling point of the proposed framework is the high recall of 92.03 ± 1.55 %, indicating that the model is sensitive to both the diabetic and the non-diabetic classes.

1. Introduction

An estimated 422 million people globally are believed to be living with diabetes in the year 2022, according to the World Health Organization [1]. The International Diabetes Federation (IDF) projected this figure to rise to 643 million and 783 million by the years 2030 and 2045, respectively [2]. As of 2015, the impact of diabetes on the global economy was estimated to be US $1.31 trillion. Furthermore, in 2015 diabetes was implicated in the death of an estimated 5 million adults aged between 20 and 79 years [3]. Studies have shown that the risk of getting infections is enhanced in people living with diabetes compared to the normal population, with an attendant increase in their morbidity and mortality [3].

Diabetes, most often, is associated with hyperglycemia, a condition indicative of too much sugar (glucose) in the bloodstream. This happens when the body has too little insulin (Type I diabetes) or if the body can't properly utilize the available insulin (Type II diabetes), leading to insulin resistance [4]. If left untreated for an extended period, hyperglycemia can damage the nervous system, blood vessels, tissues, and critical organs [5,6], which may result in a host of complications, namely: renal failure, retinal failure, coma, cardiovascular dysfunction, cerebral vascular dysfunction, peripheral vascular disorders, sexual dysfunction, joint failure, weight loss, ulcer, ocular diseases, kidney failure, and loss of immunity against pathogens [5–9]. Additionally, diabetes has been implicated in adverse pregnancy outcomes, including increased risk of maternal and fetal morbidity and mortality [10]. Furthermore, hyperglycemia has been known to increase the risk for adverse events and outcomes for patients undergoing cancer treatment [11,12].

Conventionally, the diabetes diagnosis process requires that multiple blood sugar tests be taken both before and after a meal. This presents practitioners with a difficult task in the diagnosis of diabetes. However, the diabetes diagnostic regime could be made simpler with the use of computational methods. Over the years, predictive models have been developed to aid in the diagnosis and consequently the treatment of diabetes.

The earlier diabetes predictive models [13–17] were developed based on simple scoring or logistic regression. These models used data derived from population-based studies that included risk factors that can easily be measured, such as age, BMI, waist circumference, physical activity, daily consumption of fruits, vegetables, or berries, history of antihypertensive drug treatment, and history of high blood glucose. In these early models different scores, ranging from 0 to 10 or 0 to 20, were assigned to each risk factor based on the ß-coefficients of a regression model (Equation 1).

$$x = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \quad (1)$$

The emergence of new technologies, however, ushered in new techniques and methods that could improve the performance and applicability of these earlier predictive models. To this end, cutting-edge algorithms such as: Machine learning (ML) and Deep Learning (DL) could and have been employed to utilize the massive amount of data being churned out by both research cohorts and medical practitioners [4,18–23]. Machine learning is fast becoming a tool of choice among researchers in the medical field for building predictive models to supplement the diagnostic process and treatment of various diseases [24]. This tool has demonstrated its usefulness in handling large numbers of variables, detecting, and interpreting complex relationships between variables [9,24–27].

This research effort is geared towards the development of a predictive model using machine learning tools implemented using Python programming language. In this study, we propose a stacked model consisting of a gradient boosting classifier (GBC), an adaptive boost classifier (ADA), and a light gradient boosting machine (LGBM), with the ADA boost as the meta_model. The proposed framework is intended to achieve not only high accuracy but also high sensitivity (recall) towards both the diabetic and non-diabetic classes. The proposed model is expected to aid medical practitioners in the clinical diagnosis of diabetes. Furthermore, this model could be useful and helpful to numerous undiagnosed diabetic patients who, very often, are unaware of their condition.

In a study conducted by Maniruzzaman [28] and co-researchers, a framework was developed where missing values and outliers were replaced by the group median and median values, respectively. Using a combination of Radom Forest feature selection and Random Forest classifier, the authors achieved accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and area under the curve of 92.26 %, 95.96 %, 79.72 %, 91.14 %, 91.20 %, and 93.0 %, respectively.

In a related development, a research effort undertaken by [29] utilized the Pima Indians Diabetes Dataset, which consisted of 768 entries, to develop a predictive model for diabetes. The authors applied feature selection techniques along with five classification algorithms, namely: Support Vector Machine (SVM), Multi-Layer Perception (MLP), Logistic Regression (LR), Random Forest (RF), and Decision Tree (DT), to achieve the highest classification accuracy of 78.7% with MLP.

In another research effort [7], the authors used an artificial neural network that was implemented in the Jacobian Neural Network (JNN) environment, for the prediction of diabetes. The dataset used in the study consists of 1004 entries and 9 features. Their effort resulted in a predictive model for diabetes with 87.3 % accuracy. On the other hand, Hasan and co-researchers [8] proposed a framework for the prediction of diabetes using the Pima Indian Diabetes Dataset. A weighted ensemble of different machine learning classifiers (K-Nearest Neighbors, Decision Tree, Random Forest, Naïve Bayes, Ada Boost, XGBOOST Multi-Layer Perception) was employed. The Area Under the Curve (AUC) was the metric chosen for evaluating the performance of the mode. The ensemble classifier achieved a sensitivity, specificity, false omission rate, diagnostic odds ratio, and AUC of 0.789, 0.934, 0.092, 66.234, and 0.950, respectively.

Butt and collaborators [30], proposed a model for early-stage detection and prediction of diabetes. The authors compared three classifiers, Random Forest, Multi-Layer Perception, and Logistic Regression, where the Multi-Layer Perception was the best-performing model with an accuracy of 86.08%. This was improved to 87.26 % by using Deep Learning based Long Short-Term Memory.

In another study conducted by Roy and collaborators [31], a diabetes predictive model was developed, based on an Artificial Neural Network (ANN) algorithm. In the study, median value data imputation was used to handle missing data, while the imbalance in the dataset was handled using a combination of SMOTETomek and random oversampling. This model achieved an accuracy of 98%, however, like other models using the Pima Indian dataset, it was plagued with concerns over the dataset. The dataset consisted of all females aged 21 years and above, with the population limited to Native Americans. Using the Pima Indian Diabetes Dataset, [4] developed a framework to conduct a comparative study based on the effectiveness of a three-category categorization model. In the first category, the authors considered the model's performance with and without data pre-processing. In the second category, using the Recursive Feature Elimination (RFE) feature selection method, the performance of five different algorithms was compared. While in the third category, data augmentation was done employing the SMOTE oversampling method, and model performances were compared with and without.

With dataset collected from the Murtala Mohammed Specialist Hospital, Kano, Nigeria [32] developed a supervised predictive model based on Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbours (KNN), Radom Forest (RF), Naïve Baiyes (NB), Gradient Boosting Classifier (GBC). In that study RF achieved the highest accuracy of 88.76 %, while RF and GBC had better receiver operating characteristic curve (ROC) of 86.28 %. Lai and cohorts [33] used a dataset of 13,309 Canadian patients aged between 18 and 90 years. In their research effort, RF was the best-performing model with sensitivity and area under the receiver operating characteristic curve (AROC) of 84% and 73.4 %, respectively.

Abnoosian and co-investigators [34] proposed a multi-classification model framework to predict diabetes in three classes: namely; diabetic, non-diabetic, and pre-diabetics. Their framework achieved accuracy, precision, recall, and F1-score values of 0.9887, 0.9861, 09792, 0.9851, and 0.999, respectively.

2. Method

2.1. Dataset

The dataset used in this study was sourced from the Kaggle database [35]. This dataset donated by Mohammed Mustafa, is a combination of demographic and medical data of patients, including their diabetes status. Pre-processing analysis showed that the dataset consists of 100,000 entries with 41,430 males, 58,552 females, and 18 others, with no missing values; It has eight features, namely: gender, age, hypertension, heart disease, smoking history, body mass index (BMI), haemoglobin A1c (HbA1c) level, and blood glucose level; gender and smoking history were categorical features while the rest of the features were numerical; the smoking history feature has three categories: current, never, and no info. Further pre-processing analysis shows that there were 8500 diabetics and 91500 non-diabetics in the dataset, indicating an imbalance in the dataset. Two approaches were used to handle the imbalance in the dataset before experimentation. First, the dataset was augmented using the SMOTE oversampling method. This resulted in the dataset being augmented to 158100 entries. Secondly, the dataset was reshuffled and the size was reduced to 17,000 entries, with 8500 diabetics and 8500 non-diabetics.

2.2 Data analysis

To determine the inter-relationship between the different features of the dataset, a correlation analysis was carried out on the original unbalanced and the reduced-balanced datasets. Figure 1 and 2 display the correlation heat map of the unbalanced and reduced-balanced datasets, respectively. These heat maps show the correlation coefficients (R-value) between respective features, and also with the target feature (diabetes). These coefficients indicate the strength and direction of the relation. These coefficients indicate the strength and direction of the relation.

For further analysis, features with continuous data fields such as age, BMI, HbA1c level, and blood glucose levels were modified into categorical fields. To this end, the age feature was split into four fields (< 40, 40 – 49, 50 – 60, and > 60 yrs). On the other hand, the BMI feature was split into three fields (< 18.5, 18.5 – 25, 25 – 30, > 30 kg/m^{-2}), the HbA1c level was split into three fields (< 5.7, 5.7 – 6.4, > 6.4 mmol/mol), and the blood glucose level was also split into three fields (< 126, 126 – 200, > 200 µmol/dl).

Figure 3 - 9 show the distribution of diabetes across the various categorical fields of these features. These distributions show the predisposition to diabetes of people across the categorical field groupings, for the respective features.

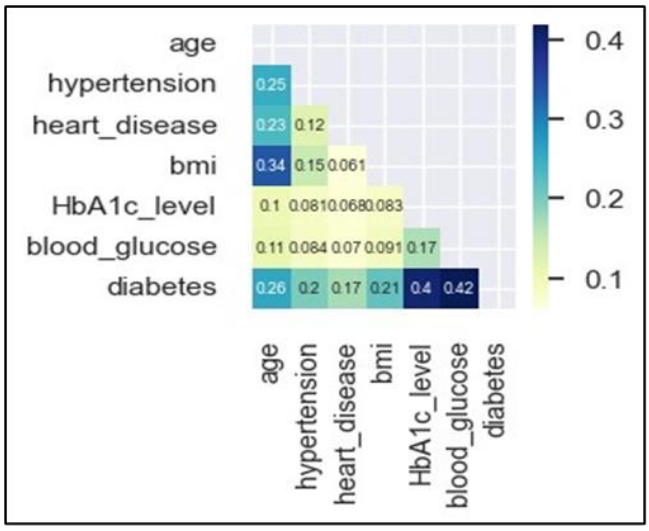


Figure 1. The heat map of the unbalanced dataset.

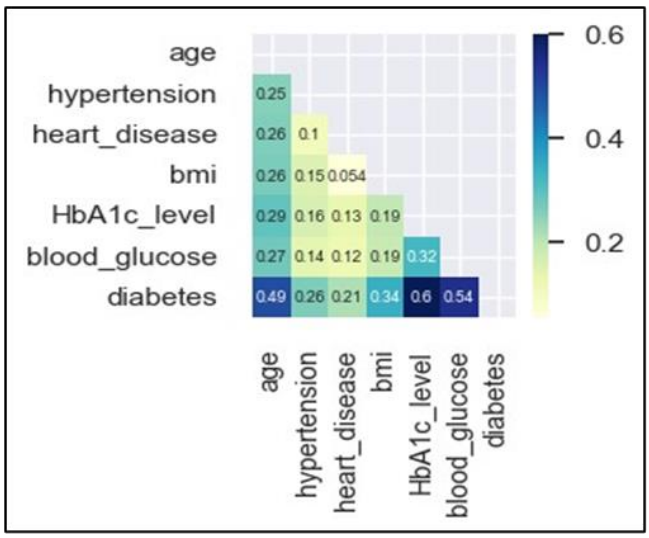


Figure 2. The correlation heat map of the balanced dataset.

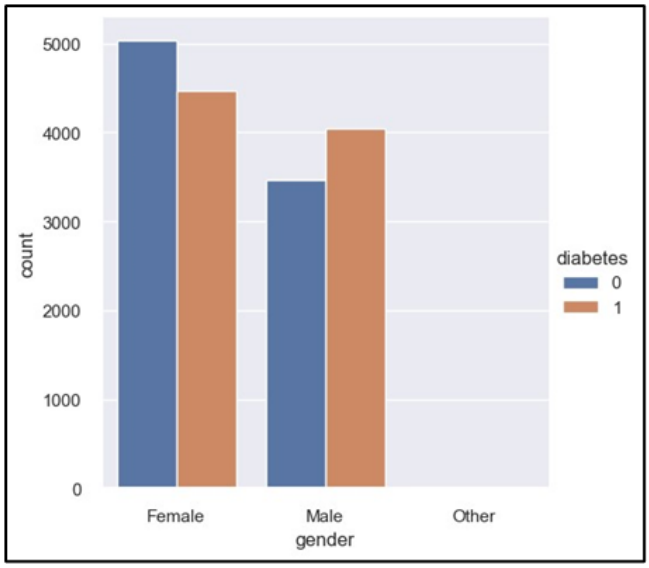


Figure 3. Diabetes distribution across gender groups.

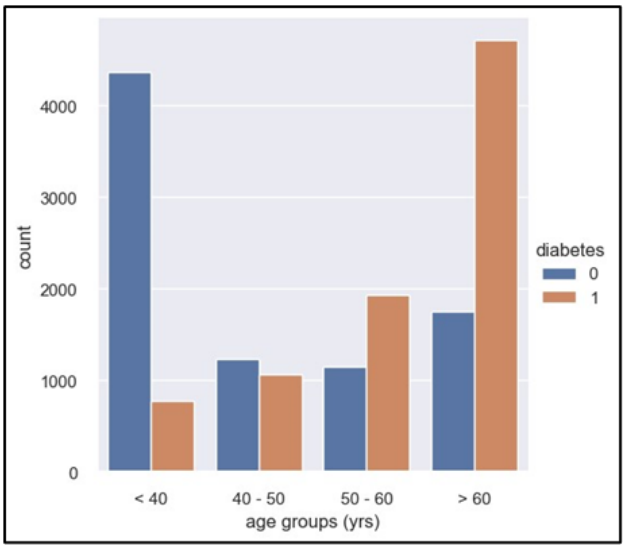


Figure 4. Diabetes distribution across age groups.

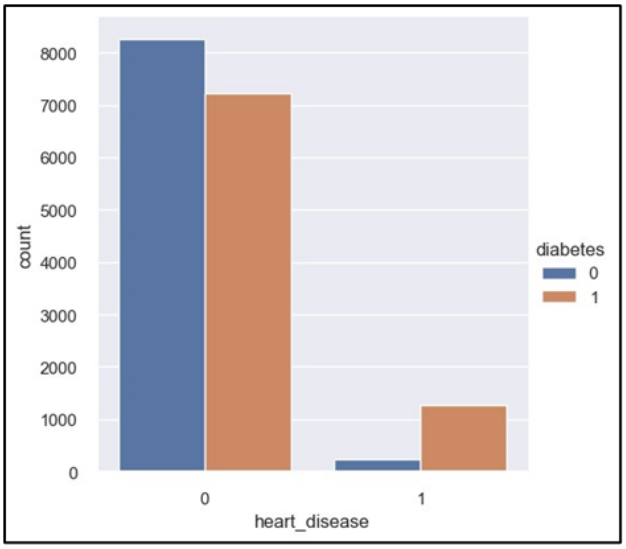


Figure 5. Diabetes distribution across heart disease status (On the x-axis: 0 = no heart disease, 1 = heart disease).

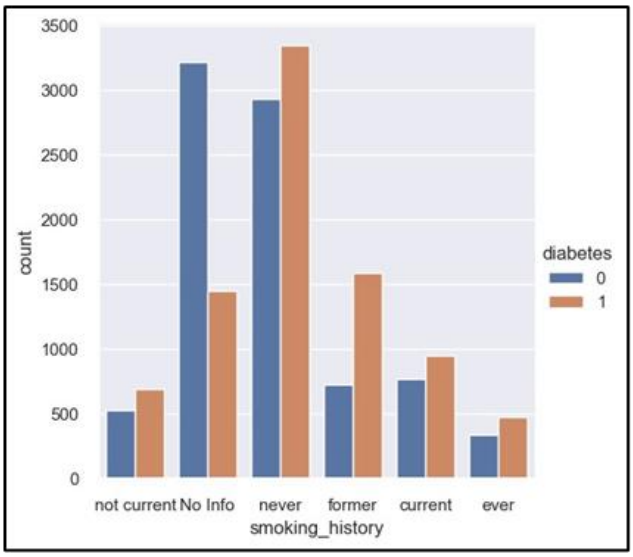


Figure 6: Diabetes distribution according to smoking history (x-axis: not current = not currently smoking, no info = no information, never = never smoked, former = former smoker, current = current smoker, ever = ever smoking).

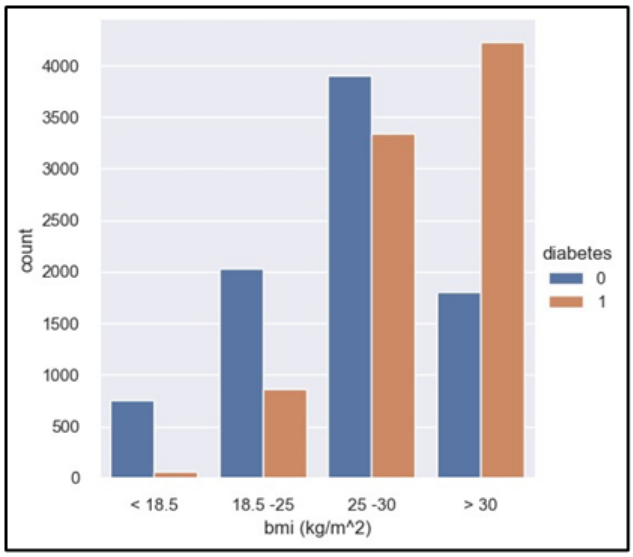


Figure 7. Diabetes distribution across bmi groupings.

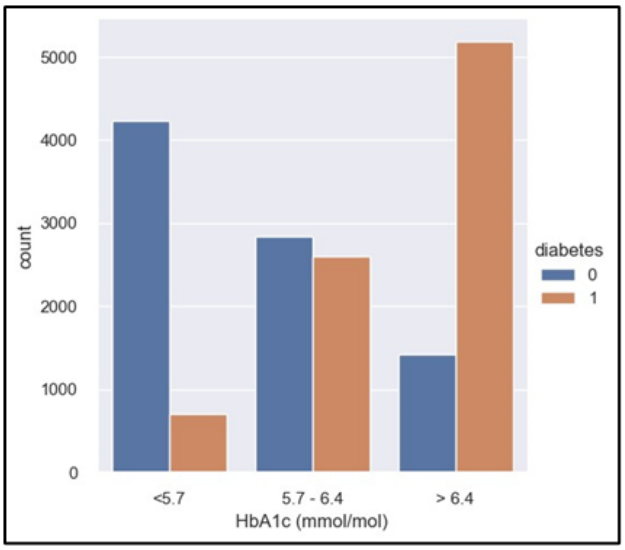


Figure 8. Diabetes distribution across HbA1c groups.

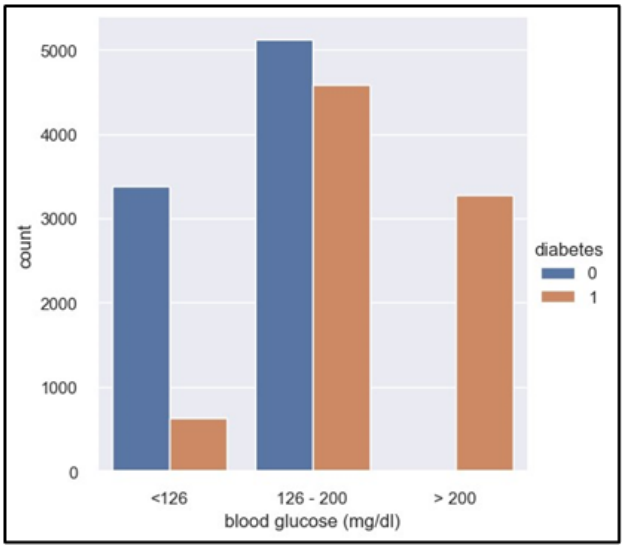


Figure 9. Diabetes distribution according to blood glucose level.

2.3 Experimentation

In this study, the PyCaret (version 3.0.2), a Python library, was employed for the investigation. The PyCaret

library is preloaded with fifteen (15) models (classifiers). These are: gradient boosting classifier (GBC), light gradient boosting machine (LGBM), adaptive boosting classifier (ADA), extreme gradient boosting (xgboost), random forest classifier (RF), extra trees classifier (ET), logistic regression (LR), linear discriminant analysis (LDA), k-nearest neighbors' classifier (KNN), decision tree classifier (DT), SVM-linear kernel (SVM) ridge classifier (RIDGE), dummy classifier (dummy), naïve bayes (NB), and quadratic discriminant analysis (QDA). The dataset was split into training and test sets using the PyCaret setup. The default setting of 0.3 was retained in splitting the dataset, thus, allocating 70 % of the dataset for training the model and 30 % for testing or evaluation. The setup was first initialized with the original, unbalanced dataset of 100,000 entries. The 15 classifiers were compared, where GBC was the best-performing model with an accuracy, AUC, recall, precision, and F1-score of 97.23 %, 97.91 %, 68.37 %, 98.60 %, and 80.74 %, respectively, followed by ADA, and LGBM, with accuracies of 97.22 % and 97.19 %, respectively. The best-performing model, GBC, was further tuned with 10-fold cross-validation to achieve an average accuracy of 97.20 %, with an average recall of 68.25 %. Next, the setup was initialized with the augmented dataset of 158100 entries. This was an attempt to improve on the recall. With the augmented dataset, the LGBM was the best-performing classifier with an average accuracy of 97.15 % and recall of 68.86 %, following a 10-fold cross-validation. There was no significant improvement in the recall with the augmented dataset. Thereafter, the setup was initialized with the reduced-balanced dataset of 17,000 entries, having 8,500 diabetics and 8,500 non-diabetics. With this setup (utilizing the reduced-balanced dataset), the 15 models were again compared. The best four (4) performing models from this setup were GBC, ADA, LGBM, and XGBOOST, achieving accuracies of 91.25 %, 91.21 %, 90.83 %, and 90.71 %, respectively with recalls of 92.30 %, 92.12 %, 91.78 %, and 91.28 %. The recalls with this setup were significantly improved. Table 1 shows the best four (4) classifiers with the different versions of the dataset.

Table 1. Best four models with different versions of the dataset.

Dataset	Model	Accuracy	AUC	Recall	Precision	F1
imbalanced	gbc	0.9723	0.9791	0.6837	0.9860	0.8074
	ada	0.9722	0.9790	0.6929	0.9715	0.8088
	lgbm	0.9719	0.9788	0.6887	0.9734	0.8066
	xgboost	0.9715	0.9779	0.6961	0.9568	0.8058
augmented	lgbm	0.9715	0.9787	0.6886	0.9659	0.8039
	xgboost	0.9709	0.9779	0.6939	0.9501	0.8020
	gbc	0.9683	0.9757	0.7151	0.8907	0.7931
	rf	0.9677	0.9613	0.6983	0.8995	0.7861
reduced-balanced	gbc	0.9125	0.9788	0.9230	0.9042	0.9134
	ada	0.9121	0.9786	0.9212	0.9049	0.9129
	lgbm	0.9083	0.9774	0.9178	0.9008	0.9092
	xgboost	0.9071	0.9761	0.9128	0.9027	0.9077

Table 2. Performance of tuned GBC, ADA, LGBM, xgboost, and NB classifiers.

Model		Accuracy	AUC	Recall	Precision	F1-score
gbc	Mean	0.9119	0.9780	0.9212	0.9046	0.9127
	Std	0.0105	0.0034	0.0143	0.0131	0.0105
ada	Mean	0.9115	0.9791	0.9224	0.9030	0.9125
	Std	0.0091	0.0029	0.0116	0.0129	0.0088
lgbm	Mean	0.9097	0.9758	0.9240	0.8985	0.9110
	Std	0.0069	0.0034	0.0133	0.0103	0.0070
xgboost	Mean	0.8974	0.9769	0.9751	0.8441	0.9048
	Std	0.0083	0.0035	0.0036	0.0117	0.0070
naive	Mean	0.8303	0.9069	0.8173	0.8390	0.8280
	Std	0.0137	0.0107	0.0170	0.0133	0.0142

Table 3. Performance of the different stacked models.

		Accuracy	AUC	Recall	Precision	F1-score
Stacked model of tuned gbc, ada, lgbm, xgboost, with untuned nb as meta_model	Mean	0.9067	0.9741	0.9513	0.8736	0.9107
	Std	0.0083	0.0043	0.0101	0.0117	0.0077
Stacked model of tuned gbc, ada, lgbm, with tuned nb as meta_model	Mean	0.9098	0.9724	0.9168	0.9043	0.9104
	Std	0.0080	0.0047	0.0138	0.0101	0.0082
Stacked model of tuned gbc, ada, lgbm, with tuned gbc as meta_model	Mean	0.9112	0.9783	0.9203	0.9040	0.9120
	Std	0.0075	0.0029	0.0155	0.0118	0.0077
Stacked model of tuned gbc, ada, lgbm, with tuned ada as meta_model	Mean	0.9120	0.9779	0.9249	0.9018	0.9131
	Std	0.0091	0.0028	0.0180	0.0089	0.0096
Stacked model of tuned gbc, ada, lgbm, with tuned lgbm as meta_model	Mean	0.9113	0.9782	0.9222	0.9028	0.9123
	Std	0.0088	0.0028	0.0154	0.0117	0.0090

The best 3 performing models were then separately created and tuned (Table 2). These tuned models were then stacked to produce a more robust model. After experimenting with GBC, ADA, LGBM, XGBOOST, and NB

as the meta-model (Table 3), the ADA classifier was adapted as the meta-model, since it achieved a higher recall without sacrificing much accuracy and precision. Although with the NB classifier as the meta-model, the stacked model achieved the highest recall of 95.13 %, accuracy, and precision were sacrificed. The final model therefore was a stack of GBC, ADA, and LGBM, with ADA as the meta-model. Figure 10 shows the framework for the model development.

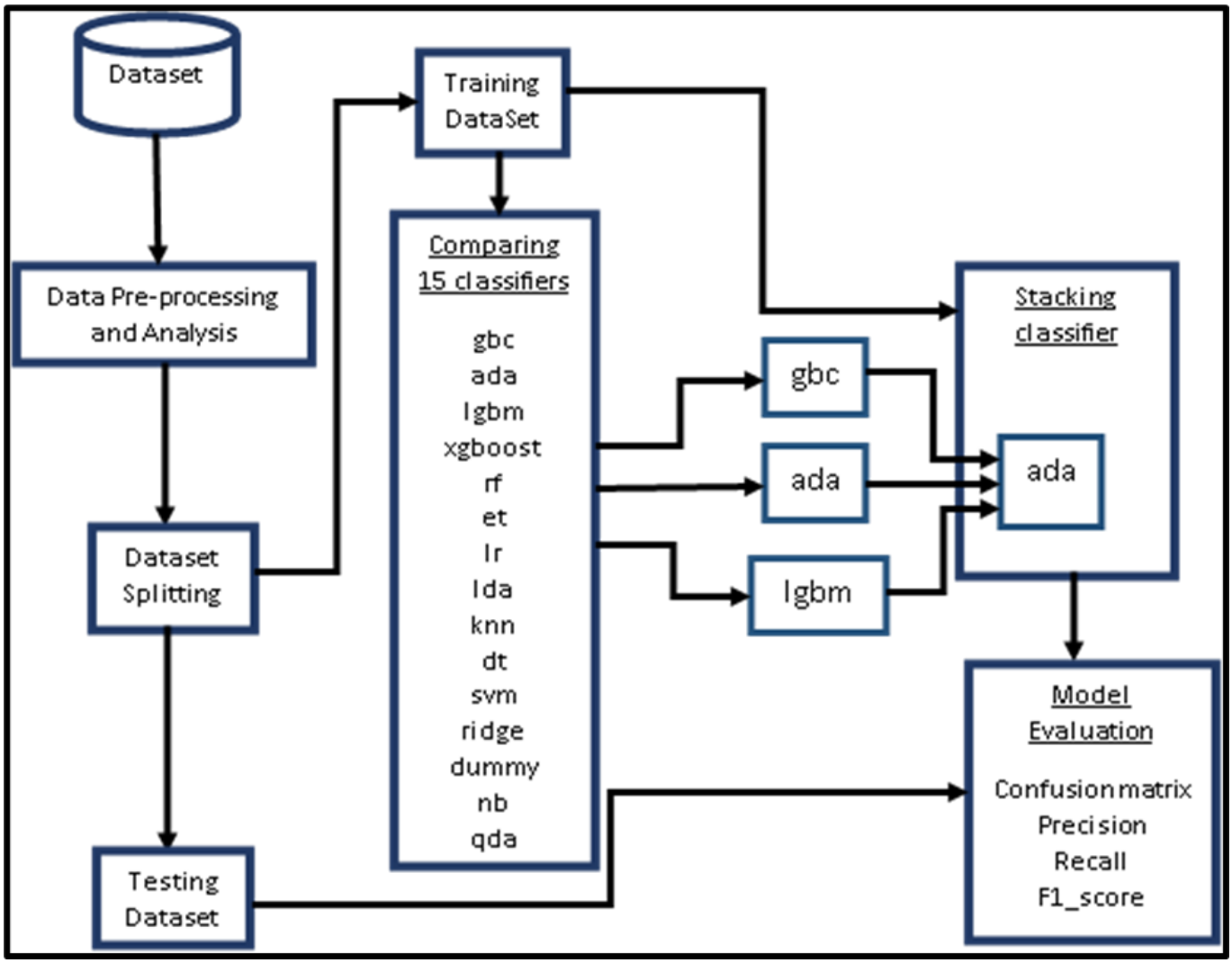


Figure 10. Framework for the model development.

2.4 Evaluation metrics

The outcomes of predictions from the model developed were assessed using the following metrics: Confusion matrix, Accuracy, AUC, Recall, Precision, and F1-score. These are respectively defined as follows:

2.4.1 Confusion matrix

This is a matrix (Figure 11) used for the evaluation of a model's overall performance, highlighting, True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) predictions of a classification model [36 - 38]. It compares the actual target values with the values predicted by the classification model. It is a matrix summarizing the number of correct and incorrect predictions by the classifier. A confusion matrix forms the basis for the calculation of other performance metrics such as accuracy, precision, recall, and F1-score.

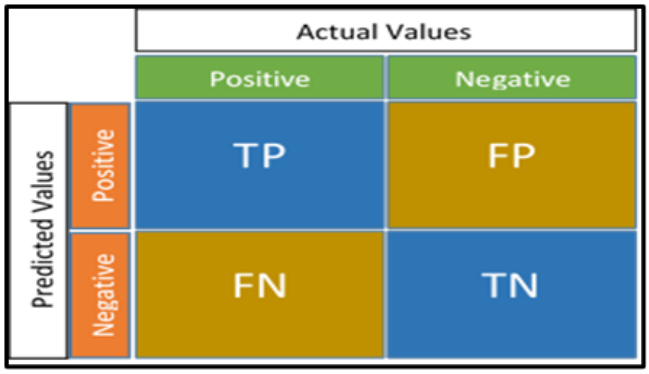


Figure 11. Confusion matrix highlighting the positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions of a classification model.

2.4.2 Accuracy

The accuracy of a classification model is the ratio of

correct predictions to the total number of predictions (Equation 2).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

It is a measure of how often a model makes correct predictions. The accuracy metric, however, is not suited for imbalanced data. For instance, if the model predicts that all the predicted values are in the majority class, the accuracy will be high but the model itself is not accurate.

2.4.3 Precision

The precision of a classification model is the ratio of the total number of correctly predicted positive classes to the total number of predicted positive classes (Equation 3).

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

In a nutshell, it gives us a measure of how many predictions are actually positive out of all the positive predictions.

2.4.4 Recall

Recall, also referred to as the sensitivity of a classification model is a measure of actual observations that are correctly predicted (Equation 4). It is thus a measure of how many positive classes are predicted as positive, given as:

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

Recall is an important metric in the present study where actual positive cases mustn't go undetected.

2.4.5 F1-Score

F1-Score is the harmonic mean of the precision and recall of a classification model. It is defined in Equation 5.

$$F1 - Score = 2 * \frac{Recall * Precision}{Recall + Precision} \quad (5)$$

F1-Score maintains the balance between Recall and Precision of a classification model.

3. Results and discussion

Figure 1 and 2 show that with the unbalanced dataset, all the features have a weak but positive correlation with one another and with diabetes. However, with the reduced-balanced dataset, all the features have a weak positive correlation with one another, except for HbA1c/diabetes and blood glucose level/diabetes which have R-values of 0.6 and 0.54, respectively. Pre-modeling analysis of the reduced-balanced dataset Figures 3 to 9 show the predisposition to diabetes as follows: People above 50 years of age; individuals with heart disease; BMI greater than 30 kg/m^2; HbA1c level above 6.4 mmol/mol; and blood glucose level above 200 mg/dl; respectively are predisposed to diabetes.

The final model in this study is a stack model consisting of GBC, ADA, and LGBM with ADA as the meta-model. Figure 12 depicts the confusion matrix of the final model based on the reduced-balanced dataset of 17,000 entries. On the other hand, Figure 13 shows the classification report, while Figure 14 and 15 show the class prediction error, and precision-recall curve, respectively, of the stack classifier. The confusion matrix (Figure 12) shows that out of the 5,100 entries in the test set, the diabetic and non-diabetic classes have 2,550 entries each. Of the 2,550 diabetics, the model predicted 2,374 correctly and only 176 incorrectly. On the other hand, 2,290 of the 2,550 non-diabetics were correctly predicted by the model while 260 were incorrectly predicted. This gives an overall accuracy of 91.45 % for the model. A comparison of the performance of the model developed in this study with other studies is shown in Table 4.

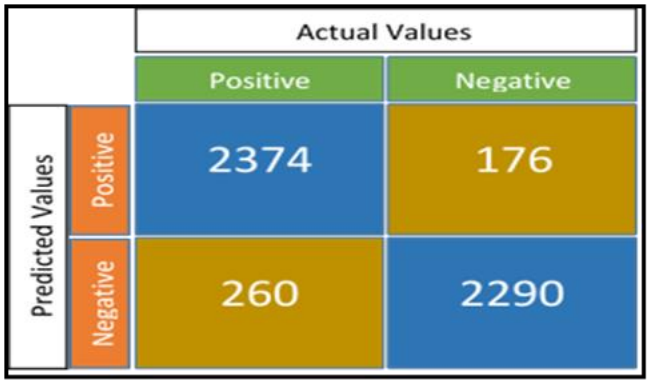


Figure 12. The confusion matrix of the stacked classifier used for predicting diabetes.

Table 4. Comparison of present study with similar studies.

Accuracy (%)	AUC (%)	Recall (%)	Precision (%)	Reference
91.12	97.83	92.03	90.40	Present study
92.26	93.0	95.26	79.72	28
78.7	-	-	-	29
87.3	-	-	-	7
-	95.0	78.9	93.4	8
87.26	-	-	-	30
98.17		97.00	99.00	31
88.76	-	-	-	32
-	-	73.4	-	33
98.87	-	97.92	98.61	34

Furthermore, the classification report (Figure 13) shows that the model has a sensitivty (ability to correctly predict positive cases) of 92.9 % and specificity (ability to predict true negatives) of 92.6 %. This high values for both sensitivity and specificity are a pointer to the model's great potential and efficacy to assist medical practioners in early diagnosis and treatment of diabetic patients. It is important to stress that the sensitivity (recall) of a classification model is an important metric that shows the sensitivity of the model, particularly towards the positive class. A high accuracy, or precision alone does not speak to the sensitivity or otherwise of a classification model. The values of both the sensitivity

and specificity of the model shows that the model is "sensitive" to both diabetics and non-diabetics. For a disease like diabetes where early detection is crucial, a predictive model designed for the detection of the disease in patients must be sensitive to the positive class in particular. It is not enough for a predictive model of this nature to have high accuracy and precision; it must also have high sensitivity towards the positive class. This ensures that positive cases are not easily missed which will aid in the early diagnosis and treatment of patients. The authors believe that a recall of 92.9 % for the positive class is a good starting point while seeking improvements.

Furthermore, the prediction error plot (Figure 14) shows that fewer diabetic cases were falsely classified as non-diabetic than non-diabetic cases that were falsely classified as diabetic. This again highlights the sensitivity of the model to the diabetic class. Further still, Figure 15, precision-recall curve, shows that the AUC (the blue shaded area) is large, which is a mark of a good model.

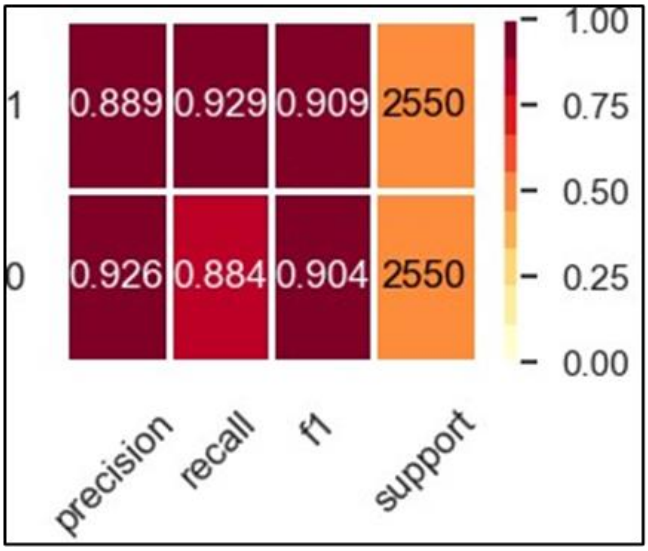


Figure 13. Classification report for the stacked classifier of GBC, ADA, and LGBM with ADA as the meta-model.

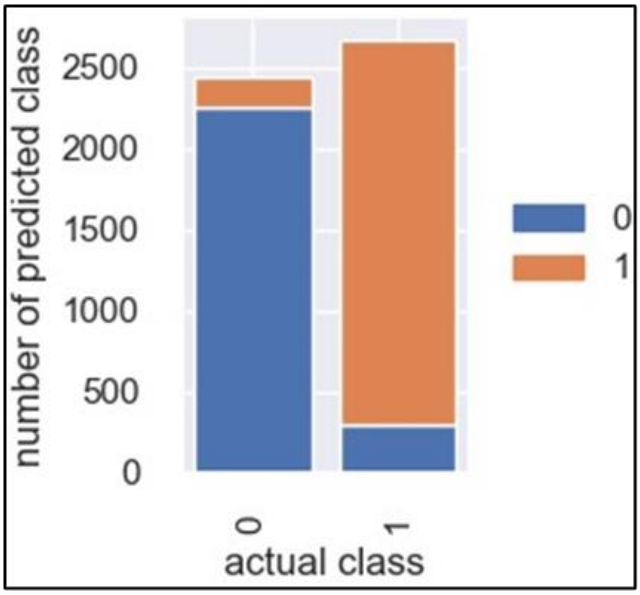


Figure 14. Class prediction error of the stacked classifier.

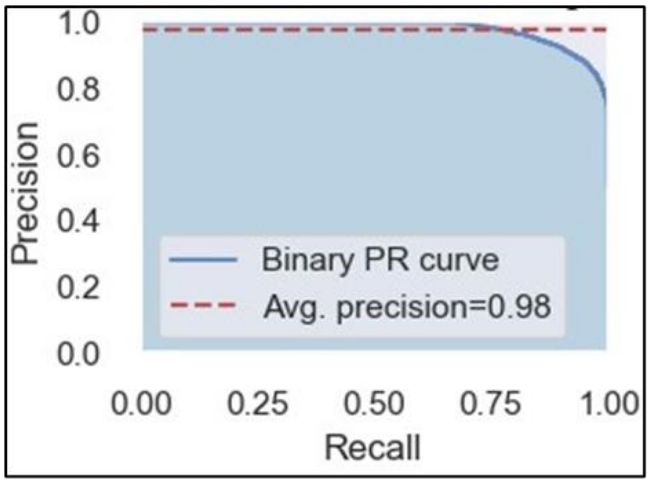


Figure 15. Precision-Recall curves of the stacked classifier.

4. Conclusion

In this study, a model was developed for classifying diabetes in humans, using a modified diabetes dataset sourced from the Kaggle database. The dataset was modified to have a balanced dataset which ensures a good recall for the model. The model is intended to aid medical practitioners in the diagnostic process of diabetes. The model developed in this study was a stack model consisting of GBC, ADA, and LGBM, with the ADA classifier as the meta-model. This stack model achieved an average accuracy, AUC, recall, precision, and F1-score of 91.12 ± 0.75 %, 97.83 ± 0.29 %, 92.03 ± 1.55 %, 90.40 ± 1.01 %, and 91.12 ± 0.77 %, respectively. The current model achieved a relatively high recall (sensitivity) without sacrificing accuracy and precision. Of particular note is the high AUC. The high AUC, high recall, and precision of the model highlight its potential clinical value and efficacy in assisting medical practitioners in diagnosing diabetes. While its high AUC highlights its overall good performance, its high recall means that very few positive cases will be wrongly classified as negative (low false negative rate), ensuring that positive cases are not easily misdiagnosed, which in turn will enhance early detection and treatment. Equally important is the high precision of the model, indicative of its low false positive rate, meaning that non-diabetics will not be easily misdiagnosed as diabetics. For a predictive model of this nature, the model must be sensitive to the positive class, ensuring that positive cases are easily captured.

Acknowledgement

The authors acknowledge the support and encouragement of ENVIRON LAB., Department of Physics and Engineering Physics Obafemi Awolowo University, Ile-Ife, Nigeria.

Author contributions

Danjuma Maza: Study conception, Design, Methodology, Data Analysis, Experimentation, Writing-Original draft preparation. **Joshua Olufemi Ojo:** Writing-Reviewing and Editing. **Grace Olubumi Akinlade:** Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. WHO. (2023). Diabetes, Diabetes Report. https://www.who.int/health-topics/diabetes#tab=tab_1
2. IDF (2021). Facts & figures. https://idf.org/about-diabetes/diabetes-facts-figures/
3. Woldaregay, A. Z., Årsand, E., Botsis, T., Albers, D., Mamykina, L., & Hartvigsen, G. (2019). Data-driven blood glucose pattern classification and anomalies detection: machine-learning applications in type 1 diabetes. Journal of medical Internet research, 21(5), e11030. https://doi.org/10.2196/11030
4. Sabitha, E., & Durgadevi, M. (2022). Improving the diabetes Diagnosis prediction rate using data preprocessing, data augmentation and recursive feature elimination method. International Journal of Advanced Computer Science and Applications, 13(9), 921-930. https://doi.org/10.14569/IJACSA.2022.01309107
5. Choubey, S., Agrahari, S., Shaw, A., Dhar, S., Sarma, R. R., Singh, S. K., Das, P., & Saha, B. (2023). Diabetes Prediction Using ML. International Journal for Research in Applied Science and Engineering Technology, 11(6), 4209-4212. https://doi.org/10.22214/ijraset.2023.54415
6. Marcovecchio, M. L. (2017). Complications of acute and chronic hyperglycemia. US Endocrinol, 13(1), 17-21. https://doi.org/10.17925/USE.2017.13.01.17
7. El_Jerjawi, N. S., & Abu-Naser, S. S. (2018). Diabetes prediction using artificial neural network. International Journal of Advanced Science and Technology, 121, 54-64. http://dx.doi.org/10.14257/ijast.2018.121.05
8. Hasan, M. K., Alam, M. A., Das, D., Hossain, E., & Hasan, M. (2020). Diabetes prediction using ensembling of different machine learning classifiers. IEEE Access, 8, 76516-76531. https://doi.org/10.1109/ACCESS.2020.2989857
9. Temurtas, H., Yumusak, N., & Temurtas, F. (2009). A comparative study on diabetes disease diagnosis using neural networks. Expert Systems with Applications, 36(4), 8610-8615. https://doi.org/10.1016/j.eswa.2008.10.032
10. Bashir, M., Naem, E., Taha, F., Konje, J. C., & Abou-Samra, A. B. (2019). Outcomes of type 1 diabetes mellitus in pregnancy; effect of excessive gestational weight gain and hyperglycaemia on fetal growth. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 13(1), 84-88. https://doi.org/10.1016/j.dsx.2018.08.030
11. Hammer, M., Storey, S., Hershey, D. S., Brady, V. J., Davis, E., Mandolfo, N., Bryant, A. L., & Olausson, J. (2019). Hyperglycemia and Cancer: A State-of-the-Science Review. Oncology Nursing Forum, 46(4), 459-472. https://doi.org/10.1188/19.ONF.459-472
12. Storey, S., Von Ah, D., & Hammer, M. (2017). Measurement of hyperglycemia and impact on the health outcomes in people with cancer: challenges and opportunities. Oncology Nursing Forum, 44(4), E141. https://doi.org/10.1188/17.ONF.E141-E151
13. Griffin, S. J., Little, P. S., Hales, C. N., Kinmonth, A. L., & Wareham, N. J. (2000). Diabetes risk score: towards earlier detection of type 2 diabetes in general practice. Diabetes/metabolism Research and Reviews, 16(3), 164-171. https://doi.org/10.1002/1520-7560(200005/06)16:3<164::AID-DMRR103>3.0.CO;2-R
14. Park, P. J., Griffin, S. J., Sargeant, L., & Wareham, N. J. (2002). The performance of a risk score in predicting undiagnosed hyperglycemia. Diabetes Care, 25(6), 984-988. https://doi.org/10.2337/diacare.25.6.984
15. Lindstrom, J., & Tuomilehto, J. (2003). The diabetes risk score: a practical tool to predict type 2 diabetes risk. Diabetes Care, 26(3), 725-731. https://doi.org/10.2337/diacare.26.3.725
16. Heikes, K. E., Eddy, D. M., Arondekar, B., & Schlessinger, L. (2008). Diabetes risk calculator: a simple tool for detecting undiagnosed diabetes and pre-diabetes. Diabetes Care, 31(5), 1040-1045. https://doi.org/10.2337/dc07-1150
17. Stern, M. P., Williams, K., & Haffner, S. M. (2002). Identification of persons at high risk for type 2 diabetes mellitus: do we need the oral glucose tolerance test?. Annals of Internal Medicine, 136(8), 575-581. https://doi.org/10.7326/0003-4819-136-8-200204160-00006
18. Kodama, S., Fujihara, K., Horikawa, C., Kitazawa, M., Iwanaga, M., Kato, K., ... & Sone, H. (2022). Predictive ability of current machine learning algorithms for type 2 diabetes mellitus: A meta-analysis. Journal of Diabetes Investigation, 13(5), 900-908. https://doi.org/10.1111/jdi.13736
19. Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. Computational and Structural Biotechnology Journal, 15, 104-116. https://doi.org/10.1016/j.csbj.2016.12.005
20. Nai-Arun, N., & Moungmai, R. (2015). Comparison of classifiers for the risk of diabetes prediction. Procedia Computer Science, 69, 132-142. https://doi.org/10.1016/j.procs.2015.10.014
21. Olisah, C. C., Smith, L., & Smith, M. (2022). Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective. Computer Methods and Programs in Biomedicine, 220, 106773. https://doi.org/10.1016/j.cmpb.2022.106773
22. Singh, A., Halgamuge, M. N., & Lakshmiganthan, R. (2017). Impact of different data types on classifier performance of random forest, naive bayes, and k-nearest neighbors algorithms. International Journal of Advanced Computer Science and Applications, 8(12), 1-10.
23. Tejedor, M., Woldaregay, A. Z., & Godtliebsen, F. (2020). Reinforcement learning application in diabetes blood glucose control: A systematic review. https://doi.org/10.1016/j.artmed.2020.101836
24. Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and

perspective. Artificial Intelligence in Medicine, 23(1), 89-109. https://doi.org/10.1016/S0933-3657(01)00077-X
25. Asfaw, T. A. (2019). Prediction of diabetes mellitus using machine learning techniques. International Journal of Computer Engineering and Technology, 10(4), 145-148. https://doi.org/10.34218/ijcet.10.4.2019.004
26. Yu, W., Liu, T., Valdez, R., Gwinn, M., & Khoury, M. J. (2010). Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. BMC Medical Informatics and Decision Making, 10, 1-7. https://doi.org/10.1186/1472-6947-10-16
27. MacMahon, H., Naidich, D. P., Goo, J. M., Lee, K. S., Leung, A. N., Mayo, J. R., ... & Bankier, A. A. (2017). Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017. Radiology, 284(1), 228-243. https://doi.org/10.1148/radiol.2017161659
28. Maniruzzaman, M., Rahman, M. J., Al-MehediHasan, M., Suri, H. S., Abedin, M. M., El-Baz, A., & Suri, J. S. (2018). Accurate diabetes risk stratification using machine learning: role of missing value and outliers. Journal of Medical Systems, 42, 92. https://doi.org/10.1007/s10916-018-0940-7
29. Ahuja, R., Sharma, S. C., & Ali, M. (2019). A diabetic disease prediction model based on classification algorithms. Annals of Emerging Technologies in Computing (AETiC), 3(3), 44-52. https://doi.org/10.33166/AETiC.2019.03.005
30. Butt, U. M., Letchmunan, S., Ali, M., Hassan, F. H., Baqir, A., & Sherazi, H. H. R. (2021). Machine learning based diabetes classification and prediction for healthcare applications. Journal of Healthcare Engineering, 2021(1), 9930985. https://doi.org/10.1155/2021/9930985
31. Roy, K., Ahmad, M., Waqar, K., Priyaah, K., Nebhen, J., Alshamrani, S. S., ... & Ali, I. (2021). An enhanced machine learning framework for type 2 diabetes classification using imbalanced data with missing values. Complexity, 2021(1), 9953314. https://doi.org/10.1155/2021/9953314
32. Muhammad, L. J., Algehyne, E. A., & Usman, S. S. (2020). Predictive supervised machine learning models for diabetes mellitus. SN Computer Science, 1(5), 240. https://doi.org/10.1007/s42979-020-00250-8
33. Lai, H., Huang, H., Keshavjee, K., Guergachi, A., & Gao, X. (2019). Predictive models for diabetes mellitus using machine learning techniques. BMC Endocrine Disorders, 19, 1-9. https://doi.org/10.1186/s12902-019-0436-6
34. Abnoosian, K., Farnoosh, R., & Behzadi, M. H. (2023). Prediction of diabetes disease using an ensemble of machine learning multi-classifier models. BMC Bioinformatics, 24(1), 337. https://doi.org/10.1186/s12859-023-05465-z
35. Mustafa, M. (2023). A Comprehensive Dataset for Predicting Diabetes with Medical & Demographic Data. https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset
36. Morris, A., & Misra, H. (2002). Confusion matrix based posterior probabilities correction.
37. Allen, G. D., & Goldsby, D. (2014). Confusion theory and assessment. International Journal of Innovative Science, Engineering & Technology, 1(10), 436-443.
38. Tharwat, A. (2021). Classification assessment methods. Applied Computing and Informatics, 17(1), 168-192. https://doi.org/10.1016/j.aci.2018.08.003