JOURNAL
Science and Technology

Volume 30 - No. (9) - 2025

Journal of Science and Technology is an Open Access Peer-Reviewed Journal Published Monthly by Faculty of Engineering and Computing – University of Science & Technology - Aden - Yemen.

Online ISSN: 2410-5163 Print ISSN: 1607-2073

جامعة العلوم والتكنولوجيا

University of Science & Technology

المركز الرئيس - عـــدن

https://doi.org/10.20428/jst.v30i9

Journal of Science and Technology

Vol. 30 No. 9 (2025)

Journal of Science and Technology

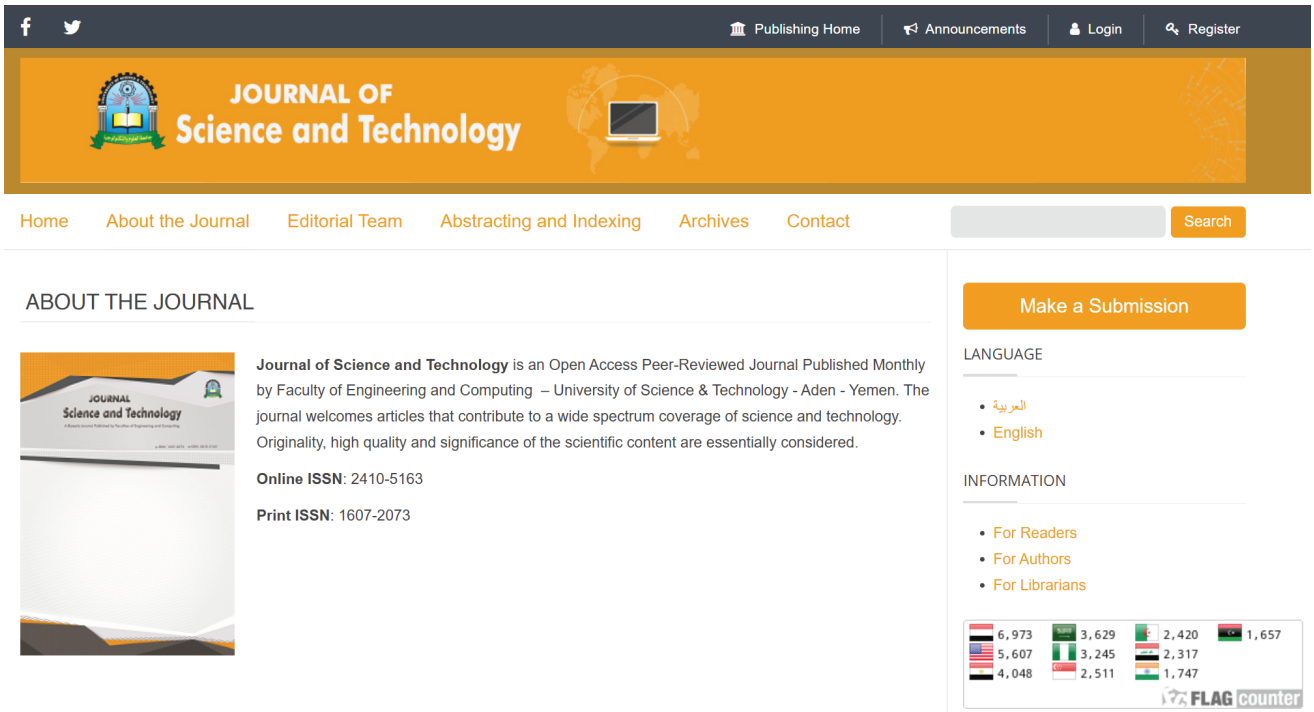
https://doi.org/10.20428/jst.v30i9

Editorial Team

For Contact
Journal of Science and Technology
Faculty of Engineering and Faculty of Computing and Information Technology, University of Science and Technology, Main Campus - Aden - Yemen
+967718032009
jst@ust.edu

https://doi.org/10.20428/jst.v30i9

Website:



https://journals.ust.edu/index.php/JST

https://doi.org/10.20428/jst.v30i9

Publication Rules and Procedures

General Rules:

Research papers sent for publication in the Journal of Science and Technology at the University of Science and Technology – Aden, Yemen, should meet the following requirements:

1. The journal publishes research papers in Arabic and English, in the following categories (but not limited to): Physics, Chemistry, Mathematics, s Engineering, and Computer Science. if they comply with the following conditions:

a. Title:

A title should summarize the main idea of the manuscript. It should identify the variables under investigation and the relationship between them. It should be concise and fully explanatory for readers when standing alone. It is recommended to be no more than 12 words and with no abbreviations. It should be typed in uppercase and lowercase letters, centered in the upper half of the page.

b. Author's name and institutional affiliation:

Every manuscript should include the name of the author and the institutional affiliation of the author when the research was conducted.

Author's name:

The preferred form of an author's name is first name, middle initial(s), and last name, with no titles (e.g., Dr., Professor) or degrees (e.g., PhD, PsyD, EdD). If the manuscript is done by more than one author, the names of the authors should be in the order of their contributions, centered between the side margins.

Institutional affiliation:

The affiliation (i.e. institution) should be centered under the author's name on the next line. If an author has no institutional affiliation, list the city and state of residence below the author's name. The emails for all authors should be provided.

c. Abstract:

An abstract in both Arabic and English must be included, with the former version **in the language of the manuscript.** It should be in a form of a brief, comprehensive summary of the contents of the manuscript written in a single paragraph with no paragraph indentation. It must not exceed 200 words.

The abstract should clearly describe the problem under investigation, in one sentence if possible; identify the purpose of the research, the essential features of study method; the participants' relevant characteristics such as age, sex, and ethnic and/or racial group; the basic findings including statistical significance levels; and the conclusions and the implications or applications. Keywords (3 - 5) should follow the abstract to increase user's ability to find useful information in the manuscript.

Introduction:

The body of a manuscript starts with an introduction that frames the problem under study and explores the importance of the problem (why the problem deserves new research). The statement about importance might involve the need to resolve any inconsistency in results of past work and/or extend the reach of a theoretical formulation and/ or investigate a practical problem that people suffer. The introduction ends in concluding the statement of the problem with a brief but formal statement of the purpose of the research that summarizes the material preceding it.

The author needs to discuss the relevant related literature in the introduction. A scholarly description of the earlier work will provide a summary of the most recent directly related work and recognize the priority of the work of others. The description of relevant literature will present what other aspects of this study have been investigated in the previous studies and how the current study differs from the earlier ones. For summarizing earlier works, focus should be on the topic (research syntheses of the topic), methodological issues, relevant findings, and main conclusions.

The discussion of related literature should demonstrate the logical continuity between previous and present work (**demonstration of gaps**); and the development of the problem should have enough breadth and clarity that make it easy to understand by a wide range of professionals.

After developing the theoretical background and the problem of the study, the author has to state the objectives and the hypotheses or specific questions. The introduction should be 10 to 15% of the manuscript. It starts on a separate page (i.e. p. 3).

https://doi.org/10.20428/jst.v30i9

d. Method:
This section describes in detail how the study was conducted, including conceptual and operational definitions of the variables used in the study. A comprehensive description of the methods used enables the reader to evaluate the appropriateness of the methods and the reliability and the validity of the results. It may include participant characteristics, sampling procedures, sample size, measures and research design.

e. Results:
Results section summarizes the collected data and the analysis done on those data. It should provide sufficient detail about the data to justify the study conclusions. The results should also include details that may not match the study expectation; or even small effect sizes (or statistically non-significant findings) particularly when theory predicts large (or statistically significant) ones. Uncomfortable results should not be omitted. The data can be presented in tables or figures (data presented in tables should not be represented in figures). Tables should be numbered in order of mention in the text. Tables can be single-spaced and should not contain any lines. Asterisks may be used to indicate significant findings. Symbols, acronyms or abbreviations should be used sparingly. Explanatory footnotes should be used whenever possible rather than overlong titles. Images should be submitted as high-resolution files (300 dpi or higher) in TIFF format (LZW compression) or JPEGs.

f. Discussion:
After presenting the results, their implications should be evaluated and interpreted, especially with respect to the original hypotheses. The author needs to examine, interpret, and draw inferences and conclusions from the results emphasize any theoretical or practical consequences of the results (Results and discussion can be combined in one section). Similarities and differences between the results and the work of others should be used to contextualize, confirm, and clarify the conclusions. Each new statement should contribute to the interpretation and to the reader's understanding of the problem.
The interpretation of the results should discuss the limitations or weaknesses of the study, and address alternative explanations of the results. It also discusses the generalizability of the findings. This critical analysis should take into account differences between the target population and the accessed sample.

g. Conclusion:
This concluding section presents a brief, reasoned and justifiable commentary on the importance of the findings. It is tightly reasoned, self-contained, and not overstated. In this section, the importance of the problem (as stated in the introduction) should be discussed; what larger issues might depend on the findings; and what propositions are confirmed or disconfirmed.

Acknowledgements:
This **section** refers to the aid received by the author from other relevant parties. Reference should also be made to any financial assistance received to conduct the research. Any extraordinary assistance received by the author in word processing, data collection, data analysis, and so on, should be acknowledged. The acknowledgements should not exceed 60 words.

References:
References start on a separate page.
Authors should acknowledge the work of previous scholars by citing references to document statements in their manuscripts. JST uses IEEE standard format. For accurate, complete, and useful citations, authors can use a reference management tool.

Publication Procedures:

Manuscript acceptance, rejection, or acceptance with revision:
The editor decides whether the manuscript is accepted, rejected or needs to be revised based on the reviewers' reports.

Manuscript acceptance: Accepted manuscripts will undergo copy-editing and production phases of publication process. The authors will not be allowed to make further changes to the manuscript except for those recommended by the copyeditors. The authors remain responsible for the completion of any amendments required by the journal.

Manuscript Rejection: A manuscript is rejected if it falls outside the domain of the journal, has serious defects in design, methodology, analysis or interpretations, lack of contribution to the field, or has a low-quality.

Manuscript acceptance with revision:
A manuscript may be conditionally accepted. This takes place when the manuscript has a high potential for final acceptance and publication in the journal, and the author adheres to all the essential modifications required by the journal (e.g. gathering essential data, conducting new experiments, reanalyzing the data, etc.). The author has to attend to the editor's recommendations for revision. The revised manuscript should be resubmitted with an enclosed cover letter that contains a table explaining in detail how and where (in the manuscript) amendments have been done based on the reviewers' comments.

Publication Ethics:
The **JST** complies with the recommendations of the Committee on Publication Ethics (COPE) to promote the integrity of its published articles. The **JST** considers the following topics during the publishing process:

- **Originality & Source Acknowledgement:** The **JST** scans all submitted manuscripts before the peer-reviewing process using Turnitin®. The **JST** is zero-tolerant to plagiarism, self-plagiarism, copyright infringement, dual publication, text recycling and salami slicing. When any of these is identified after publishing, an announcement of retraction of the published material is highlighted in the journal's website. The authors are asked to provide appropriate references for published/unpublished cited texts. The corresponding author should confirm that the submission has not been previously published and is not being considered for publication elsewhere.
- **Research Misconduct:** The **JST** editorial team struggles to counter any possibility for data fabrication, manipulation and falsification. In case of suspected misconduct, the **JST** editors act in accordance to the COPE guidelines with this respect.
- **Conflicts of interest:** Authors should disclose potential conflicts of interest and indicate financial agreements or affiliations with any product or services used in the manuscript (as well as any potential bias against another product or service).
- **Authors:** Authors should disclose (in an author note) activities and relationships that if known to others might be viewed as a conflict of interest, even if the authors do not believe that any conflict or bias exists (e.g. an author has his stock in a company that manufactures a drug used in his study).
- **Reviewers:** Reviewers should also reveal their potential conflicts of interest (if any) to the action editor. They have an ethical obligation to be open and fair in assessing a manuscript without bias. They should not review a manuscript from a colleague or collaborator, a close personal friend, or a recent student. Reviewers should maintain the confidentiality of a manuscript. They should not discuss the manuscript with other individuals.
- **Using materials under copyrights:** The author should obtain letters of permission from copyright holders to reproduce (or adapt) copyrighted material and enclose copies of these letters with the accepted manuscript. Examples of material that require permission include reprinted figures and tables, tests and scale items, questionnaires, vignettes, etc.

Correction notices: If an error is detected in the published manuscript, the author can submit a proposed correction notice to the journal's editor. The notice should indicate the full title of the journal, the year of publication, the volume no., issue no., and the page nos. of the article, the precise location of the error(s) (e.g. page, line, column, exact quotation of the error, or paraphrasing of lengthy errors).

Publication Fees:
JST does not require any article processing charges or any article submission charges

Sponsorship:
JST is sponsored by the University of Science and Technology.

https://doi.org/10.20428/jst.v30i9

Content:

Journal of Science and Technology

https://doi.org/10.20428/jst.v30i9

An Enhanced Gray-Scale Digital Watermarking Approach Utilizing Discrete Wavelet Transform and Reed-Solomon Error Correction

M. F. Abdullah [2]
N. M. A. Munassar [1,2]
M. A. A. Al-Hobishi [1,*]

Received: 07/05/2025
Revised: 08/07/2025
Accepted: 09/07/2025

[1] College of engineering and computing, IT department, University of Science and Technology, Aden, Yemen.
[2] Electronic & Distance College, University of Science and Technology, Aden, Yemen.
*Corresponding Author's Email: Eng.Monia85@gmail.com, n.munassar@ust.edu, m.albadwi@ust.edu

An Enhanced Gray-Scale Digital Watermarking Approach Utilizing Discrete Wavelet Transform and Reed-Solomon Error Correction

Monia Abdullah Ahmed Al-Hobishi
College of engineering and computing, IT department, University of Science and Technology,
Aden, Yemen.
Eng.Monia85@gmail.com

Nabil Mohammed Ali Munassar
College of engineering and computing, IT department, University of Science and Technology,
Electronic & Distance College, University of Science and Technology,
Aden, Yemen.
n.munassar@ust.edu

Muhammed Fadhl Abdullah
Electronic & Distance College, University of Science and Technology,
Aden, Yemen.
m.albadwi@ust.edu

Abstract— Digital watermarking has become essential for protecting intellectual property and ensuring content authenticity in the digital age. However, a significant challenge remains in developing watermarking techniques that are both robust against various attacks (such as compression, noise, and cropping) and imperceptible to the human eye, which is crucial for maintaining the quality of the original content. This paper addresses these challenges by proposing an advanced digital image watermarking technique that combines a three-level Discrete Wavelet Transform (DWT) using the Haar wavelet family with Reed-Solomon (RS) error-correcting codes. The three-level DWT decomposes the image into multiple frequency components, allowing the watermark to be embedded in the most significant parts of the image, thereby enhancing robustness. The integration of Reed-Solomon codes over finite fields further increases the watermark's resilience, enabling recovery even when parts of the watermark are damaged or lost due to attacks. Experimental results demonstrate that the proposed approach significantly improves watermark performance, with Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) showing substantial gains. The combination of Haar-based DWT and Reed-Solomon codes improves the watermark's robustness by up to 27 times compared to traditional methods without error correction. This approach provides a promising solution for secure, efficient, and reliable digital watermarking in applications requiring high robustness and content integrity.

Keywords— Discrete Wavelet Transform (DWT), Gray image watermarking, Reed Solomon (R.S.), watermarking embedding, watermarking extraction, Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM),Attack(Salt and paper, Gaussian ,Speckle).

الملخص ـ أصبح تضمين العلامات المائية الرقمية أمرًا أساسيًا لحماية الملكية الفكرية وضمان أصالة المحتوى في العصر الرقمي. ومع ذلك، تظل التحديات الرئيسية تتمثل في تطوير تقنيات العلامات المائية التي تكون قوية بما يكفي لمقاومة الهجمات المختلفة (مثل الضغط، والضوضاء، والقص) وفي نفس الوقت غير ملحوظة للعين البشرية، وهو أمر ضروري للحفاظ على جودة المحتوى الأصلي. يعالج هذا البحث هذه التحديات من خلال اقتراح تقنية متقدمة لتضمين العلامات المائية في الصور الرقمية تعتمد على التحويل الموجي المتقطع ثلاثي المستويات (DWT) باستخدام عائلة موجات HAAR إلى جانب أكواد تصحيح الأخطاء REED-SOLOMON (RS). يقوم التحويل الموجي ثلاثي المستويات بتقسيم الصورة إلى مكونات تردد متعددة، مما يسمح بتضمين العلامة المائية في الأجزاء الأكثر أهمية في الصورة، مما يعزز من قوة العلامة المائية. تعزز إضافة أكواد REED-SOLOMON فوق الحقول المنتهية من مقاومة العلامة المائية، مما يمكنها من التعافي حتى في حالة تعرض أجزاء من العلامة للتلف أو الفقد بسبب الهجمات. أظهرت النتائج التجريبية أن المنهجية المقترحة تحسن أداء العلامة المائية بشكل كبير، حيث أظهرت مقاييس نسبة الذروة للإشارة إلى الضوضاء (PSNR) ومؤشر التشابه البنيوي (SSIM) زيادات ملحوظة. ويعزز الجمع بين تحويل HAAR الموجي وأكواد REED-SOLOMON من قوة العلامة المائية بمقدار يصل إلى 27 ضعفًا مقارنة بالطرق التقليدية التي لا تتضمن تصحيح الأخطاء. تقدم هذه المنهجية حلًا واعدًا لتضمين العلامات المائية الرقمية بشكل آمن وفعال وموثوق في التطبيقات التي تتطلب درجة عالية من القوة وسلامة المحتوى.

الكلمات المفتاحية: *تحويل الموجات المنفصلة (DWT)، العلامات المائية للصورة الرمادية، ريد سولومون (R.S.)، تضمين العلامات المائية، استخراج العلامات المائية، نسبة ذروة الإشارة إلى الضوضاء (PSNR)، مقياس مؤشر التشابه البنيوي (SSIM)، الهجوم (الملح والورق، الغاوسي، البقع).*

I. Introduction

Digital watermarking embeds a digital image into a host image, enabling detection and proving ownership rights, tracking content usage, ensuring authorized access, facilitating content authentication, and preventing unauthorized replication.
[3,1]. Digital watermarking research covers various branches, Figure 1 [2, 7, and 10].
using classification methods like Discrete Fourier Transform, Discrete Cosine Transform, and Discrete Wavelet Transform. Techniques include DFT, DCT, and DWT.
Researchers investigate non-blind watermarking in the transform domain using invisible watermarks and evaluate DWT's effectiveness with Reed-Solomon cyclic error-correcting codes against various attacks. [6, 5, 8, 11, 9]. The remainder of this paper is organized as follows:

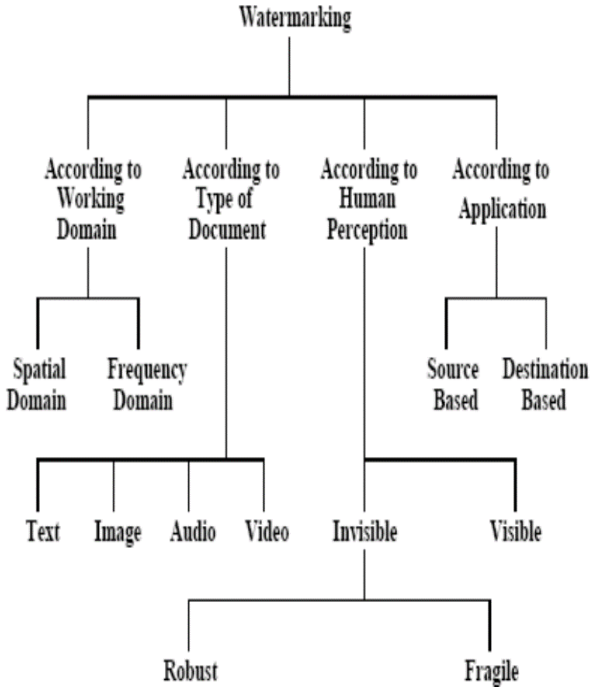


Figure 1: Watermarking Classification

II. Literature Review

Research in digital watermarking is related to many subjects in the area of computer science, such as image processing, communication theory, and encryption. In addition to data hiding and steganography science, all these related subjects will be covered in this research. As a result of the importance of digital watermarking in solving the problems of protecting the copyrights of owners or publishers in digital media, there are numerous studies that have suggested using error control coding techniques to correct the possible errors caused by the attacks on the digitally marked image [1, 2, 3, 4, 5, 6].

It has motivated the researcher to contribute to this field by investigating the effects of the well-known error correction codes, the Reed-Solomon code, and DWT domain embedding and extracting algorithms. The study aims to increase the robustness of image watermarks.

As mentioned previously, the transformation of products from physical to digital requires the development of methods and algorithms to address intellectual perceptual problems, in particular the copyright problem. The good solution is a combination of digital watermarking (visible/invisible mark) methods and error correction coding (ECC) methods, which protect copyright [7, 8, 9, 10, 11, 12].
Many models and algorithms have been designed to address copyright problems. Numerous papers and research about the digital watermarking techniques are proposed to solve some watermarking problems, and other research focuses on the error correction code concept [15, 16, 17, and 18].
More specific research articles that are very closely related to the research topic are given below for completeness.
Details of This study presents invisible, blind, and robust color image watermarking algorithms using wavelet transform and error-correcting codes. These algorithms are designed to be robust against attacks like JPEG compression, noise, and color. Reed-Solomon codes are used to improve robustness, with Sudan's algorithm showing good performance against hue and saturation attacks [4, 18, 19, 20].
A new digital image watermarking algorithm uses Reed-Solomon codes and rank metric codes for attack resistance, complementing other random error watermarking strategies like JPEG compression. [24, 25, 27].

The study examines the impact of various Reed-Solomon codes, including 15 and 31, on the robustness of image watermarks, focusing on five of them and their embedding method.
The study investigates Discrete Wavelet Transform (DWT) sensitivity against various attacks on watermarked images, revealing that the Reed-Solomon code outperforms other codes.
[26, 18, 30].
The proposed digital watermarking method utilizes DWT and DCT techniques to embed and extract copyright protection, resulting in a hybrid approach for two-dimensional images.
[18].

Mathematical Preliminary & State of the Art

A. Discrete Wavelet Transform

The wavelet transform is a sophisticated mathematical tool widely used across various applications, particularly in image processing and watermarking [20,6]. The core concept of the Discrete Wavelet Transform (DWT) is to decompose an image into its frequency components using wavelets of varying frequencies and limited durations. During each level of decomposition, the DWT separates the image into four sub-bands:
The wavelet transform creates four sub-bands:
1-LL (Low-Low),
2-HL (High-Low),
3-LH (Low-High),
and 4-HH (High-High).
The LL sub-band captures lower-resolution approximations of the image, while the LH sub-band focuses on vertical detail. The LL sub-band can be further decomposed to achieve additional detail levels. [13,9,23] .

B. Reed Solomon:

Reed-Solomon (RS) codes are error-correcting codes used in digital communication systems and storage devices, notably for correcting burst errors in CDs, DVDs, QR codes, and satellite communications [15,18.20].

Key Concepts of Reed-Solomon Codes:

1. Block Codes: Reed-Solomon codes are block codes, meaning that data is divided into fixed-size blocks, and each block is encoded separately. In each block, a certain number of bits are added as redundancy to help detect and correct errors.
2. Symbols and Field Arithmetic: Unlike binary codes (which work with individual bits), Reed-Solomon codes work with symbols. A "symbol" is typically a group of bits (e.g., 8 bits = 1 byte, or 1 symbol = 8 bits in a common implementation). The key part is that Reed-Solomon codes are based on arithmetic over finite fields, specifically Galois fields.

- For an RS code, the finite field typically used is the field of polynomials modulo an irreducible polynomial. The field can have size q = 2mq = 2^mq = 2m, where mmm is an integer and q is the number of possible symbols in the field.
- In simpler terms, instead of just using binary arithmetic (0 or 1), the symbols are drawn from a larger set of values.

3. Parameters: A Reed-Solomon code is usually denoted as RS(n, k), where:
   - n is the total number of symbols in the encoded block (length of the codeword).
   - k is the number of data symbols (original message).
4. Encoding and Decoding:
   - **Encoding**: The data is transformed into a codeword by adding redundancy (parity symbols) based on a set of mathematical operations in the finite field.
   - **Decoding**: When the data is received, the decoder uses the parity symbols to detect and correct errors. The most common decoding algorithm is the **Berlekamp-Massey algorithm**, which efficiently finds errors in received codewords.
5. Error Detection and Correction:

- Because the code is designed to detect patterns of errors in the data, even if multiple symbols are corrupted, it can correct a certain number of errors without needing to request retransmission. This is particularly useful for applications like satellite communication or optical media, where retransmission is costly or impractical. [30.26.25]

Advantages of Reed-Solomon Codes:

- Error correction: Can correct multiple errors in data blocks, especially burst errors.
- Flexibility: The parameters n and k can be chosen depending on the application, allowing for customizable error-correction strengths.
- Efficiency: Well-suited for practical implementation in hardware and software. [31].

III. Proposed Methodology

The overall Digital Watermarking System, consists of the following:

A. Digital Watermarking Embedding:

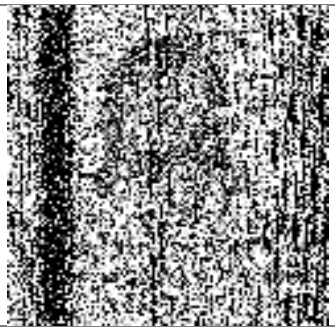
The study uses a non-blind watermarking approach, embedding a digital watermark into a host image using RS block error correcting codes and multilevel DWT. The embedding is performed using the alpha blending technique, transforming the host image directly.

$$WMI = k \times (LL_j) + q \times (EWM_j) \quad (1)$$

Where :

- EWM_j is the j^{th}-level The text describes a method for obtaining a low frequency approximation of the encoded watermark image.
- LL_j is the j^{th}-level The text describes a method for obtaining a low frequency approximation of the host image. HI.
- WMI is the watermarked image.
- k & q are the scaling factors. :
- Scaling factor k= r controls the strength of the watermark.
- Scaling factor q influences the intensity or distortion caused by the watermark embedding.
- The optimal values of r = 0.99 and q=0.001 are chosen after experiments to balance imperceptibility and robustness, making sure the watermark survives attacks while maintaining the quality of the original image.
- $j = 1,2,3$ levels in DWT.

Table.1: Determination of Extensive experiments were conducted to determine the optimal values for the scaling factors r and q.

r and q variable	Watermark Image	Watermarked Image	Note
r= 0.99 **q=0.001**			Optimal Value
r= 0.2 **q=0.01**			Worst case

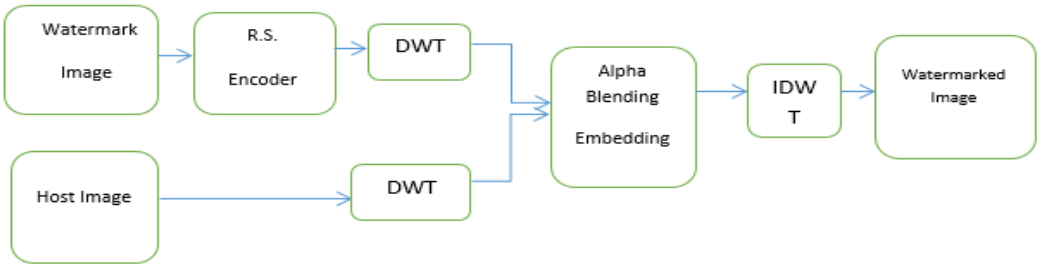


Figure 2: Watermarking Embedding

As mentioned above, two steps need more explanation:
1- RS encoding of the digital watermark image and the DWT level for both the encoded digital watermark and the host image. RS Encoding of the Digital Watermark
The encoding using length n=255 RS block codes depends on the arithmetic of finite fields GF (256) , so the digital watermark image is converted to binary digits and then to Galois field elements to be able to encode the information symbols to codewords. This is illustrated in figure (2).
2-Multilevel DWT of the encoded digital watermark & host image
The alpha blending technique for embedding the encoded digital watermark in the host image is done in the DWT transform domain.

Table 2.Digital Watermark Image Encoding Steps

Read the digital watermark image (WM)
Convert to binary bits stream
Form the GF(256) Galois field symbols
Divide into blocks of size k symbols
Call RS Encoder (n, k) to obtain the n length codeword for each k symbols
Convert to binary stream
Obtain the encoded digital watermark image (EWM)

This table outlines the steps for Watermarking Image Encoding (WME) using Reed-Solomon, further illustrated in the following figure.

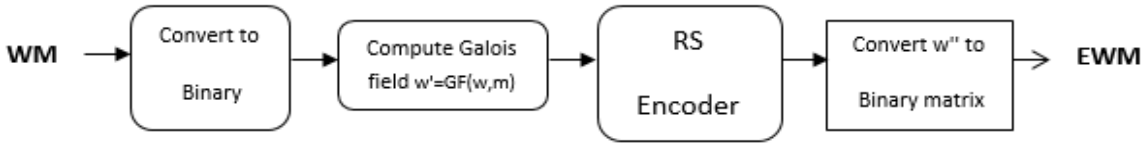


Figure 3: Watermark Image Encoding

B. Performance Evaluation & Measures

Non-blind digital watermarking systems avoid distortion, where the digital watermark image(WMI) doesn't affect the host image's appearance(HI), as per the Peak Signal to Noise ratio.

$$SNR = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} HI^2(i,j)}{\sum_{i=1}^{M}\sum_{j=1}^{N} [HI(i,j)-WMI(i,j)]^2} \quad (2)$$

And ,

$$PSNR_{dB} = 10 * \log_{10} \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} HI^2(i,j)}{\sum_{i=1}^{M}\sum_{j=1}^{N} [HI(i,j)-WMI(i,j)]^2} \quad (3)$$

Digital watermarking systems' copy write protection requires successful extraction despite severe attacks, using similarity between inserted and extracted watermarks, as outlined in Structured Similarity Image Value (SSIM).

$$SSIM = \frac{(2\mu_x \mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \quad (4)$$

Where "μ", "σ", & "σ_{xy}" The mean, variance, and covariance of the digital watermark image and the recovered digital watermark image are being analyzed.

. "c_1", "c_2" The stabilizing constants in images range from 0 to 1, with similar images having an SSIM value near 1.

C. Attacks on the Watermarked Image:

In this subsection, the various types of attacks on the watermarked image will be presented. Three main categories of attacks are considered:

a. Addition of noise attack

a-Salt & Pepper noise: J = imnoise(I,'salt & pepper',d) .The command adds salt and pepper noise to image I, affecting approximately d*numel(I) pixels, with a default of 0.05.

b-Gaussian noise: J = imnoise(I,'gaussian',M,V). The command adds Gaussian white noise with a mean of m and variance of v to the image I, defaulting to zero mean noise and 0.01 variance.

c-Speckle noise: J = imnoise(I,'speckle',v). The command adds multiplicative noise to an image using the equation J = I + n * I, where n is uniformly distributed random noise with mean 0 and variance v.

D. Digital Watermark Extraction

In this research, various types of attacks on the watermarked image are considered. Addition of noise was the focus of this research; the noise types considered are salt & pepper, Gaussian, and speckle noise. In this subsection, the extraction of the digital watermark method is presented and illustrated. Watermark Extraction Using Alpha Blending Extraction Technique: Let denote the attacked digital watermarked image. The formula of the alpha blending extraction technique is as follows:

$$ERW = (WMI_A - k \times LL_j)/q \quad (5)$$

Where :

- ERW is the low frequency approximation of the recovered encoded watermark.
- LL_j is the j^{th}-level low frequency approximation of the host image HI.
- WMI_A is the attacked watermarked image.

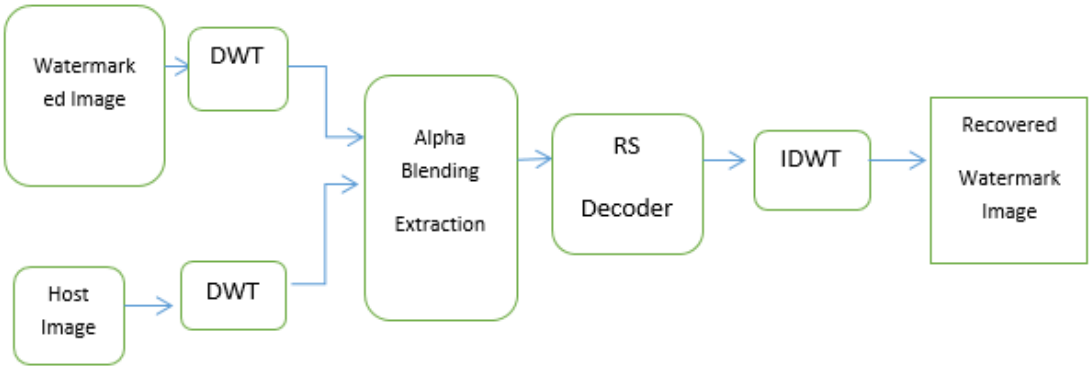


Figure 4: Watermark Extraction

In figure (5), the decoded digital watermark is obtained, so in order to obtain the recovered digital watermark image, a decoding step are to be performed as given below:

Table 3. Digital Watermark Image Decoding Steps

Read the encoded digital watermark image (EWM)
Convert to binary bits stream
Form the GF(256) Galois field symbols
Divide into blocks of size n symbols
Call RS Decoder (n, k) to obtain the k length information symbols
Convert to binary stream
Obtain the decoded digital watermark image (DWM)

This table shows the steps for Watermarking Image Decoding (WMD), and these steps are shown by the figure following.

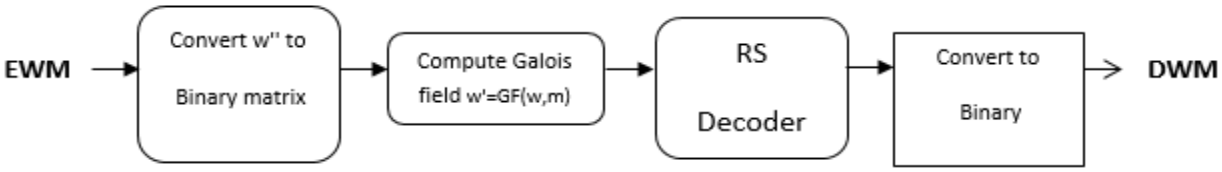


Figure 5: Watermark Image Decoding

IV. Experimental setup & Demonstration of Watermark Embedding

This table shows the watermark used in this paper along with its host image and the process of embedding it to obtain the watermarked image.

Table 4: Kinds of images that using in this paper:

Host image	Watermark image	Watermarked image
		

DWT (Discrete Wavelet Transform) is often used to embed watermarks into images at various levels of decomposition. The "DWT level" refers to the number of times the wavelet transform is applied to break down the image into different frequency sub-bands (low- and high-frequency components).

Types of attacks refer to methods that attempt to degrade, remove, or alter the watermark, such as noise addition, compression, cropping, or filtering. The effectiveness of these attacks varies depending on the DWT level used for watermark embedding. Higher DWT levels can provide more robustness against attacks, as the watermark is embedded deeper within the image's frequency components, making it harder for attackers to remove or distort it without significant degradation of the host image.

Table 5: Type of attack with DWT level

DWT _level SALT & PEPPER ATTACK	**DWT _level - GAUSSIAN NOISE ATTACK**	**DWT _level SPECKLE NOISE ATTACK**

A. Peak Signal Nosie Ratio (PSNR)

PSNR (Peak Signal-to-Noise Ratio) is a commonly used metric to measure the quality of an image before and after any processing, such as the application of attacks in watermarking. It compares the original image to the modified one (e.g., after watermark embedding or after an attack) by calculating the ratio between the maximum possible pixel value and the noise introduced due to the alteration. A higher PSNR value generally indicates that the modified image retains more similarity to the original, suggesting better quality. In the context of watermarking, PSNR is used to assess the distortion in the host image before any attacks are applied, with higher PSNR values indicating that the watermarking process has caused minimal visual distortion to the image. As shownin Table 6, the image before the attack likes Salt and Paper, Gaussian, and Speckle.

Table 6: PSNR for before Attack types

Noise	**Images Without R.S. before attacks**		
	PSNR before Salt and Paper Attack	PSNR before Gaussian Attack	PSN before Speckle Attack
	PSNR	PSNR	PSNR
0.001	46.5680	46.5680	46.5680
0.01	46.5680	46.5680	46.5680
0.1	46.5680	46.5680	46.5680

PSNR (peak signal-to-noise ratio) after an attack type measures the quality of the image after it has undergone some form of tampering or distortion, such as noise addition (salt and pepper, Gaussian, or speckle). In the context of image watermarking, this metric helps evaluate the impact of the attack on both the host image and the embedded watermark. A lower PSNR value after an attack indicates more degradation in image quality, meaning the watermark or the host image has been significantly affected. It reflects how resistant the watermark is to the applied attack, with higher PSNR values after an attack suggesting that the watermark remains intact and the image quality is still good. As shown in the table 7 image after the attack with R.S.

Table 7: PSNR for after Attack type

Noise	**Images With R.S. after attacks**		
	With R.S. after Salt and paper attack	With R.S. after Gaussian attack	With R.S. after Speckle attack
	PSNR	PSNR	PSNR
0.001	33.324	33.6678	33.6678
0.01	32.345	33.541	33.541
0.1	32.123	32.345	32.345

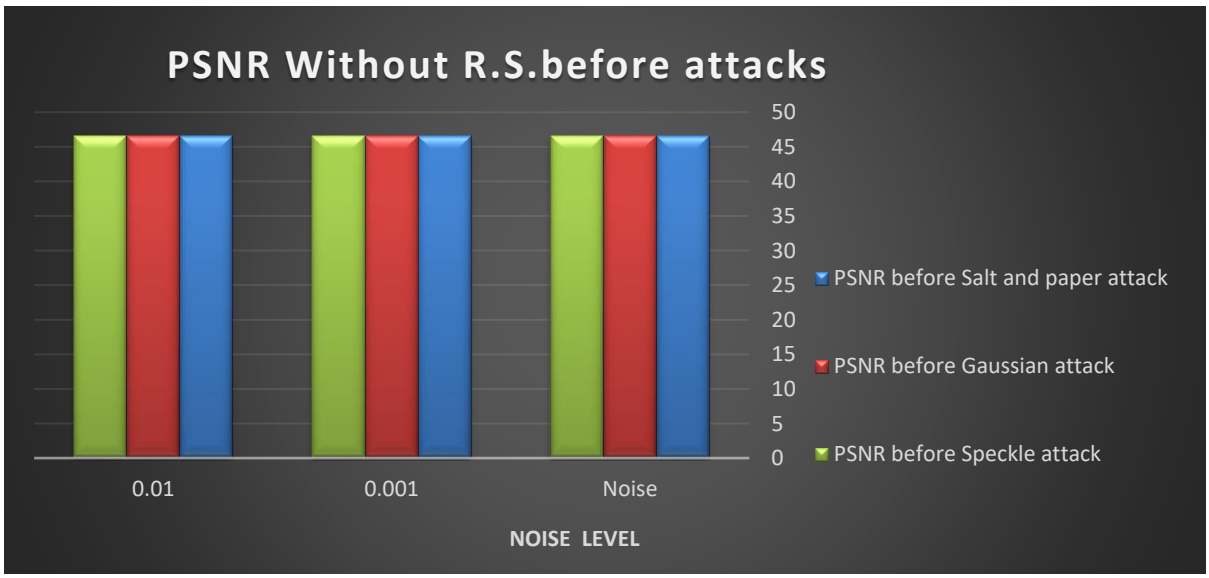


Figure 6: PSNR without R.S. before Attacks

This Figure shows PSNR before an attack this measures the quality of the original image before it undergoes any alterations or attacks.

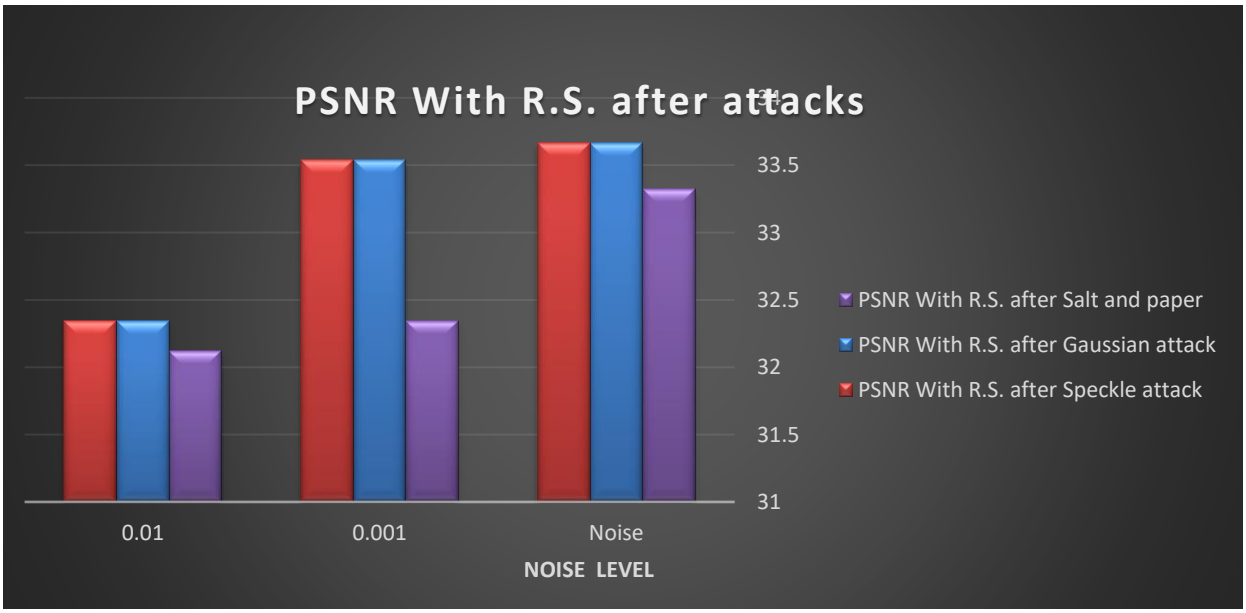


Figure 7: PSNR with R.S. after attacks

This figure shows PSNR with R.S. (reversible watermarking scheme). After attacks, it measures the quality of the watermarked image and the effectiveness of the watermarking technique after it has been subjected to various attacks. A lower PSNR value indicates that the image quality and watermark integrity have been significantly degraded, while a higher PSNR value suggests that the watermark is more resistant to the attack and the image quality has been largely maintained.

B. Similarity Structure Index Measurement (SSIM)
SSIM (Structural Similarity Index) after an attack measures the perceived quality of the image after it has undergone some form of tampering or distortion, such as noise, compression, or cropping. Unlike PSNR, which focuses on pixel-wise differences, SSIM evaluates changes in structural information, luminance, contrast, and texture, which are more aligned with human visual perception. After an attack, SSIM provides a better understanding of how the attack has impacted the overall visual quality of the image and the watermark. A lower SSIM value indicates significant degradation, meaning the attack has caused visible damage to the image or watermark, while a higher SSIM value suggests that the image and watermark remain visually similar to the original, even after the attack. In this case, a noise of type Salt & Pepper, Gaussian, or Speckle is added to the watermarked image with different levels. Table 8 and Figure 8 show the values of SSIM after extracting the watermark image for each noise DWT level for the case with RS block code.

It can be observed that using Reed-Solomon (RS) codes results in approximately a threefold improvement in performance at the noise levels of Salt and Pepper (0.001, 0.1), with values of (0.9977, 0.434). For Gaussian noise (0.001, 0.1), the values are (0.9865, 0.5235), and for Speckle noise (0.001, 0.1), the values are (0.9725, 0.5532). This illustrates, through bar plots, the significant increase in Structural Similarity Index Measure (SSIM) when using RS codes across various noise levels. Show that in table 8.

Table 8: SSIM for after Attack type

Noise Level	**Images With R.S. after Attacks**		
	SSIM after Salt And Paper attacks	SSIM after Gaussian attacks	**SSIM after Speckle attacks**
0.001	0.9977	0.9865	**0.9725**
0.01	0.8733	0.8763	**0.8588**
0.1	0.434	0.5235	**0.5532**

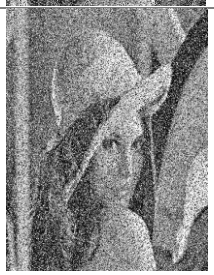
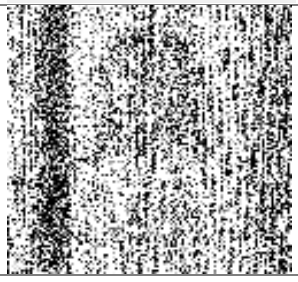
Figure 8: SSIM with R.S. after attacks

SSIM (Structural Similarity Index) after an attack evaluates the visual quality of an image after it has been subjected to distortions such as noise, compression, or cropping. Unlike PSNR, which measures pixel-based differences, SSIM focuses on how structural elements like luminance, contrast, and texture are perceived by the human eye. After an attack, SSIM provides a more accurate assessment of the image's overall visual integrity and the preservation of the watermark. A lower SSIM value indicates that the attack has significantly degraded the image, making it appear noticeably distorted, while a higher SSIM value suggests that the image and watermark have been well-preserved, retaining a high degree of similarity to the original, as shown in figure 8.

V. Experimental Setup & Demonstration of Attacks on the Watermarked Image

In the following, we demonstrate the watermarked image method for attack without R.S. The one case of that is DWT level is used and three cases of attacks (Salt and Paper, Gaussian, and Speckle). The results indicate that when noise is present without R.S., the images are unclear and distorted. However, when R.S. is present, the images are clearer and of higher quality. Yet, at a value of 0.1, they begin to lose clarity by 60%. The results are as shown below.

Table 9: Gray images in Salt and Paper attack

Gray image **Salt and Paper Attack**				
Noise level	Image without Reed Solomon		**Image with Reed Solomon**	
	Host Image	Watermark image	Host image	**Watermark image**
0.001				
0.01				

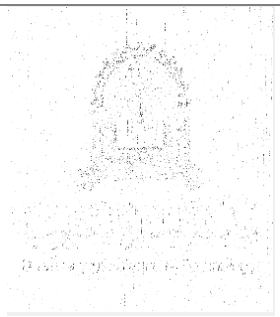
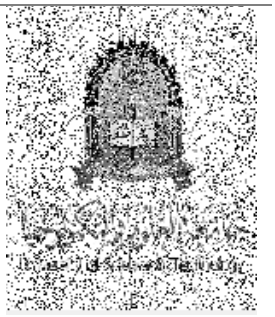
0.1				

Table 10: Gray images in Gaussian attack

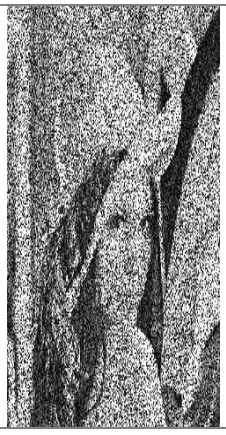
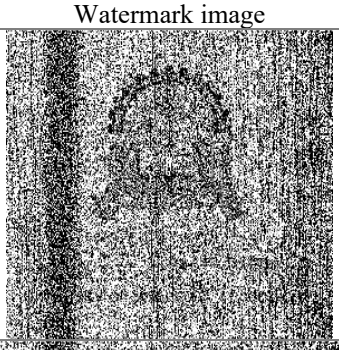
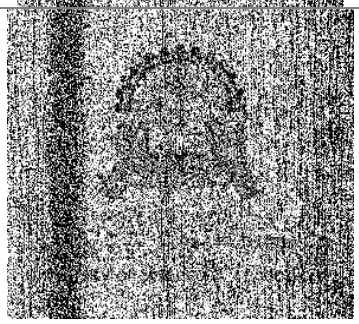
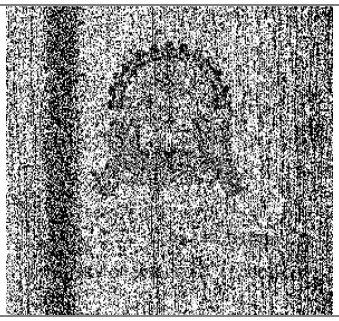
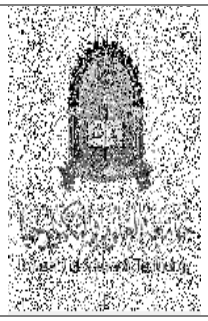
Gray image **Gaussian Attack**				
Noise level	Image without Reed Solomon		**Image with Reed Solomon**	
	Host Image	Watermark image	Host image	**Watermark image**
0.001				
0.01				
0.1				

Table 11: Speckle images in Speckle attack

Gray image Speckle Attack				
Noise level	Image without Reed Solomon		**Image with Reed Solomon**	
	Host Image	Watermark image	Host image	**Watermark image**
0.001				
0.01				
0.1				

VI. Conclusion

This paper presents an enhanced digital watermarking technique that combines Discrete Wavelet Transform (DWT) with Reed-Solomon (R.S.) codes to improve the robustness and quality of watermarking in digital images. The proposed methodology effectively embeds watermarks in significant frequency components, ensuring minimal distortion to the host image. Experimental results demonstrate substantial improvements in performance metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), particularly in scenarios involving various noise attacks. The integration of error-correcting codes, specifically Reed-Solomon codes, significantly enhances the resilience of the watermark against attacks, achieving robustness improvements up to 27 times compared to systems without such codes. The findings highlight the effectiveness of using a three-level DWT in conjunction with R.S. codes, making this approach a promising solution for secure digital watermarking applications.

7- Future work could explore further optimizations of the embedding process and the application of this technique to a broader range of multimedia content, enhancing copyright protection and content authentication in digital media.

References

[1] S. Gupta, & R. Meh, "Combining DWT and Reed-Solomon Codes for Secure Color Image Watermarking,"Journal Multimedia Tools and Applications. Vol 31(5), pp.17-23,2023.

[2] N. Patel, & V. Singh, "Dynamic Watermarking of Color Images Using DWT and Reed-Solomon for Robustness". Journal of Imaging. Vol 9(6), pp.1-13,2023.

[3] F. Ali, & A. Khan, "An Efficient Watermarking Scheme for Color Images Using DWT and Reed-Solomon Code", journal Multimedia Tools and Applications, Volume 31 (2), pp.123-135,2023.

[4] Y. Chen, & L. Zhao,"DWT-Based Color Image Watermarking with Reed-Solomon for Enhanced Robustness" journal Multimedia Tools and Applications ,Vol 31(3),pp.201-215,2023.

[5] S. Sharma, & R. Verma, Robust "Color Image Watermarking Based on DWT and Reed-Solomon Error Correction1", journal Multimedia Tools and Applications, Vol 31(4), pp.301-316 ,2023.
[6] R. Patel, & M. Joshi, " Robust Image Watermarking Using DWT and Reed-Solomon Codes: A Comparative Study", journal Multimedia Tools and Applications, Vol 31(5), PP.451-465,2023.
[7] A. Kumar, &R. Singh, "Robus Color Image Watermarking Using DWT and Reed-SolomoCodes", journal Multimedia Tools and Applications ,Vol 31(6), pp.451-465,2023.
[8] P. Sharma, & S. Gupta, "Adaptive Watermarking for Color Images Using DWT and Error Correction Techniques", journal Multimedia Tools and Applications, Vol 31(7), pp.501-515, 2023.
[9] L. Zhang, & J. Liu," A Novel Approach to Image Watermarking Using DWT and Reed-Solomon Coding", journal Multimedia Tools and Applications, Vol30(5), pp.345-360, 2022.
[10] M. Ali, & T. Khan "Enhancing Watermarking Techniques for Color Images: A DWT and Reed-Solomon Perspective" journal Multimedia Tools and Applications, Vol 30(6),pp.421-438,2022.
[11] A. Khan, & A. Kumar, "A Comparative Study of Wavelet Transform Approaches for Robust Digital Watermarking." Multimedia Tools and Applications. Vol 34(4), pp.127-130, May 2022.
[12] A. Saffor, A. Ramli, "Comparative study of image compression between JPEG and Wavelet". Malaysian Journal of computer science.; Vol 14(1), pp. 39-45. Jully2022.
[13] D. Mishra, K. Singh, R. Singh. "Wavelet-based deep auto encoder-decoder (wdaed)-based image compression". IEEE Transactions on Circuits and Systems for Video Technology. Vol 31(4), pp.1452-1462. February 2020.
[14] D. Onufriienko, Y. Taranenko. " Filtering and compression of signals by the method of discrete wavelet decomposition into one-dimensional series". Cybernetics and Systems Analysis.; Vol 19(4) pp.1-8.Nov2023.
[15] D. Silva ," Use of daubechies wavelets in the representation of analytical functions". In Wavelet Theory. Intech Open, Vol12(3),pp.234-240,Oct2020.
[16] D. Wang, L. Zhang, V. Vincent," Improvement of JPEG2000 using curved wavelet transform. In Proceedings".(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 2(3), pp. 350-365, June2020.
[17] I. Ince, F. Bulut, I. Kilic, M. Yildirim, O. Ince. "Low dynamic range discrete cosine transform (LDR-DCT) for high performance JPEG image compression". The Visual Computer,; Vol 38(5), pp.1845-1870, April 2022 .
[18] İ. ÖZ, A. Ankara, B. Yıldırım," Comparative Analysis of Wavelet Families in Image Compression, Featuring the Proposed New Wavelet", Eleco International Conference on Electrical and Electronics Engineering; Turkey. Vol3(4),pp.44,50, May2024.
[19] I.Oz , C.Oz, N. Yumusak. "Image compression Using 2-D multiple-level discrete wavelet transform (DWT)" . Eleco International Conference on Electrical and Electronics Engineering; Turkey. Vol3(4),pp.44,50, May2020.
[20] J. Bhatia, & A. Kumar, "Robust Image Watermarking Using Multilevel DWT and Genetic Algorithm." Soft Computing, vol26(5), pp.245-259, March 2022.
[21] J. Bulut, "Low dynamic range histogram equalization (LDR-HE) via quantized Haar wavelet transform", The Visual Computer, Vol 23(4) pp.2239-2255, April 2022
[22] M. Hussain, & A. Javaid, "Wavelet Transform Based Watermarking Techniques: A Survey." Journal of Visual Communication and Image Representation vol96(4), pp.116-129, November 2021.
[23] M. Martin, M. Bell. "New Image compression techniques using multiwavelets and multiwavelet packets". in IEEE Transactions on Image Processing, vol 10. (4), pp. 500-510.April 2020.
[24] M. Sajjad, & S. Shah, "Enhanced Digital Watermarking Using Multi-Level DWT and Hybrid Techniques." Journal of Visual Communication and Image Representation, vol88(4), pp.103-115, March 2023.
[25] N. Taujuddin, R. Ibrahim, S. Sari. "An improved technique to wavelet thresholding at details subbands for image compression". ARPN Journal of Engineering and Applied Sciences.; Vol 11(18), pp.10721-10726.November 2019.
[26] O. Yilmaz, M. Aksoy, Z Kesilmiş. "Misalignment fault detection by wavelet analysis of vibration signals". International Advanced Researches and Engineering Journal, Vol3(3), pp.156-163, Jully 2022.
[27] P. Viswanthan, P. Kalavathi," Sub band Thresholding for Near-Lossless Medical Image Compression", International Journal of Computing and Digital Systems.; Vol 14(1), pp.1-11, March 2023.
[28] R. Boujelbene. Y. Jemaa, M. Zribi. "A comparative study of recent improvements in wavelet-based image coding schemes". Multimedia Tools and Applications. Vol 78(2), pp.1649-1683.April 2019.
[29] R. Gupta, & R. Singh, "Comparative Study of Various Wavelet Families for Image Watermarking." International Journal of Image Processing, vol 15(3), pp.1-10, October 2022
[30] R.Ranjan, P. Kumar. "An Improved Image Compression Algorithm Using 2D DWT and PCA with Canonical Huffman Encoding." Entropy, Vol 25(10), pp.1382, March 2023.

[31] R. Starosolski “Hybrid adaptive lossless image compression based on discrete wavelet transform". Entropy.; Vol 22(7), pp.751-760, April 2.

Effect of Depreciation Calculation Methods on the Evaluation of Petroleum Projects

S. O. Baarimah [1]
K. S. Ba-Jaalah [1,*]

Received: 06/07/2025
Revised: 29/07/2025
Accepted: 30/07/2025

[1] Associate professor, Faculty of Engineering and Petroleum, Hadhramaut University, Hadhramaut, Yemen.
*Corresponding Author's Email: kbajaalah@hotmail.com , soob2005@hu.edu.ye

Effect of Depreciation Calculation Methods on the Evaluation of Petroleum Projects

Khaled S. Ba-Jaalah
Associate professor, Faculty of Engineering and Petroleum, Hadhramaut Unversity, Hadhramaut, Yemen
kbajaalah@hotmail.com

Salem O. Baarimah
Associate professor, Faculty of Engineering and Petroleum, Hadhramaut Unversity, Hadhramaut, Yemen
soob2005@hu.edu.ye

Abstract— The development of oil reserves requires critical decision-making due to the high investment costs involved. Given the substantial investment costs in petroleum projects, these decisions can significantly impact profitability. A key component of these costs is the depreciation, depletion, and amortization (DD&A) of assets over their useful life. This study provides guidelines for managing asset costs in petroleum projects by employing cash flow and financial models. Two cases were analyzed: Case 1 (fixed asset costs) applied S.L. and DDB, while Case 2 (production-linked costs) employed UOP. Although both models yield the same total profit, the cash flow model reflects profits only after several years, whereas the financial model accounts for profit from the first year by distributing the investment cost over the project's estimated life. The study compares the implications of these methods—Straight-Line, Double Declining Balance, and Unit of Production—on profit recognition and investment recovery.

Keywords— Total profit, DD&A calculation, Cash flow model, Financial model, Petroleum projects

I. Introduction

Profit, in its simplest concept, is the excess of revenue over all implicit costs of conducting business within a specified time period [1-9, 11-17]. In the oil production business, profit is directly determined by three factors, including 1) prices, 2) costs, and 3) volume.

For petroleum projects, two models are used to estimate profit:

A. Cash flow model
B. Financial model

The cash flow model is an operational concept that sums all project revenues and deducts all operating expenditures, with total asset costs incurred at the project's start as shown in Figure 1.

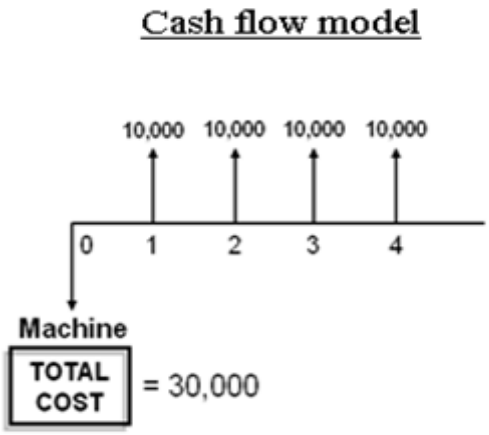


Fig. 1. Cash flow model

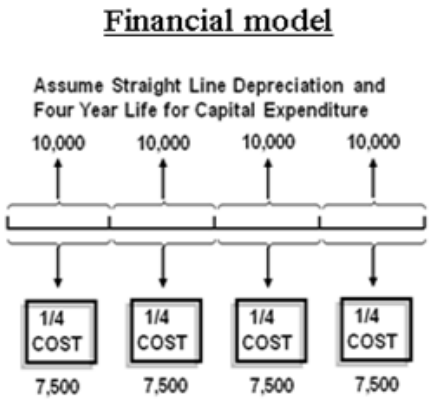


Fig. 2. Financial Model

Depreciation, Depletion, and Amortization (DD&A)

DD&A is defined as the allocation of the cost of an asset over its useful life. It is important to note that these methods are calculated by subtracting the asset salvage value from its original cost.

Depreciation refers to prorating the cost of tangible assets, such as vehicles, pipelines, wellheads, and buildings, over their useful lives. For example, an office building used for

many years before being sold will have its cost spread across its expected lifespan, with a portion expensed each year.
Depletion allocates the cost of natural resource assets (e.g., oil wells) over their productive lives. The initial costs are distributed over the total estimated recovery period.
Amortization typically involves spreading the cost of intangible assets, such as signature bonuses, royalties, and lease acquisition costs, over their useful lives.[10]
This study aims to compare the effects of three DD&A methods—straight-line, double-declining balance, and unit of production—on profit recognition in petroleum projects, evaluating the implications of using cash flow versus financial models for project valuation and decision-making. By analyzing these aspects, the research seeks to provide practical guidelines for selecting appropriate cost allocation methods based on project-specific characteristics, ensuring more accurate financial assessments and informed decision-making in the petroleum sector.

DD&A CALCULATION
The straight-line method (S.L.), the double declining balance (DDB) method, and the unit-of-production (UOP) method are the most commonly used approaches for DD&A calculations. Each requires the following parameters: (a) asset cost, (b) useful life, and (c) salvage.

A. Straight line method (S.L.):
It is the most common method used in the industry. It assumes that the value of an asset decreases at a constant rate over time. The depreciation is computed by dividing the depreciable amount of the asset by the expected number of accounting periods of its useful life, as shown in Equation 1. Useful economic life is not equal to physical life, but it is the period over which the present owner intends to use the asset [10, 14].

$$D_{SL} = \frac{C - S}{t_L} \quad \text{---------(1)}$$

Where:
D_{SL} = Straight line depreciation per year
C = Cost
S = Salvage value at the end of the life of the project
t_L = Depreciation life

B. Double decline balance method (DDB):
This method of depreciation is also known as 200% decline balance, and it is a form of accelerated depreciation. Compared to other methods, the depreciation will vary by year, but the total amount will still remain the same. Essentially, the double-decline balance means double the rate of straight line. However, the double-decline balance method ignores salvage value [10, 14].
For the first year, Equation 2 is used for calculating DDB1:

$$DDB_1 = (C - S) \times \left(\frac{M}{t_L}\right) \quad \text{----------(2)}$$

While Equation 3 is used to estimate DDB for any year (n):

$$DDB_n = DDB_{n-1} \times \left[1 - \left(\frac{M}{t_L}\right)\right] \quad \text{--------(3)}$$

To fully depreciate a capital cost by the end of its life using the double-declining balance method, a switch to the straight-line method is required. The year at which this is made can be determined using Equation 4:

$$N = t_L - \left(\frac{t_L}{M}\right) + 1 \quad \text{----------(4)}$$

C. Unit of production method (UOP):
This method is more suitable for production assets. Thus, another piece of information that will be needed is the number of units produced each year. In this case, the useful life of an asset is related more directly to units of work performed by the asset than to the passage of time. In such a case, depreciation can be calculated at the same rate for each unit of output. Equation 5 is used for determining UOP. [10, 14].

$$UOP = (C - S) \times \left(\frac{Q_n}{U}\right) \quad \text{--------(5)}$$

Where:
Q_n = Annual production of year n
U = Reserves or ultimate recovery
n = Depreciation year

II. DATA DESCRIPTION
The two cases were chosen to demonstrate the application of different depreciation calculation methods (DD&A) in petroleum projects and to highlight their practical relevance in financial decision-making. The first case compares Straight-Line (S.L.) and Double Declining Balance (DDB) for time-based depreciation, relevant for long-lived assets like pipelines. S.L. spreads costs evenly, while DDB front-loads expenses for faster cost recovery, switching to S.L. in year 9 to meet regulatory timelines. The second case uses the Unit-of-Production (UOP) method, tying depreciation to extraction rates—crucial for oil and gas fields where asset utility declines with reserves. DD&A calculations were used to evaluate the investment, and then the total profit was determined using the cash flow and financial models. Tables 1 and 2 presented input data for each case.

TABLE 1. Input Data for Case 1

Investment = $ 5 ×109					
Year	Cost (MM$)	Income (MM$)	Year	Cost (MM$)	Income (M$)
1	150	1280	9	150	1200
2	150	1280	10	150	1150
3	150	1280	11	150	1100
4	150	1280	12	150	1040
5	150	1280	13	150	980
6	150	1280	14	150	930
7	150	1280	15	150	840
8	150	1280	16	150	745

TABLE 2. Input Data for Case 2

Investment = 4000 M$ **Reserves = 8000 BSCF**			
Year	q (BSCF/Y)	Cost (MM$)	Income (MM$)
1	585	120	1025
2	585	120	1025
3	585	120	1025
4	585	120	1025
5	585	120	1025
6	585	120	1025
7	585	120	1025
8	585	120	1025
9	555	120	960
10	530	120	920
11	500	120	880
12	480	120	830
13	450	120	780
14	425	120	750
15	380	120	670

In Tables 1 & 2:
M$** : A thousand dollars, **MM$: One million dollars , **q** : Flow rate

III. Result and Discussion

In Case 1, both the straight-line and the double-declining balance methods were applied to allocate investment costs over the project's estimated lifespan, while the unit-of-production method was used in Case 2. Table 3 summarizes the DD&A calculations for Case 1, indicating when it is necessary to transition from DDB to the straight-line method, while Table 4 summarizes the DD&A calculations for Case 2. Tables 5 and 6 and Figures

TABLE 3. DD&A calculation for Case 1

Year	**Investment (MM$)**	**DS.L (MM$)**	**DDB (MM$)**	**DDB & S.L (MM$)**
0	5000			
1		312.5	625	625
2		312.5	546.875	546.875
3		312.5	478.515625	478.515625
4		312.5	418.7011719	418.701172
5		312.5	366.3635254	366.363525
6		312.5	320.5680847	320.568085
7		312.5	280.4970741	280.497074
8		312.5	245.4349399	245.43494
9		312.5	214.7555724	214.755572
10		312.5	187.9111258	214.755572
11		312.5	164.4222351	214.755572
12		312.5	143.8694557	214.755572
13		312.5	125.8857738	214.755572
14		312.5	110.150052	214.755572
15		312.5	96.38129553	214.755572
16		312.5	84.33363359	214.755572

Total (MM$)	5000	4409.664565	5000

TABLE 4. DD&A calculation for Case 2

Year	Investment (MM$)	q (BSCF/Y)	UOP (MM$)
0	5000		
1		585	365.625
2		585	365.625
3		585	365.625
4		585	365.625
5		585	365.625
6		585	365.625
7		585	365.625
8		585	365.625
9		555	346.875
10		530	331.25
11		500	312.5
12		480	300
13		450	281.25
14		425	265.625
15		380	237.5
	Total (MM$)	8000	5000

TABLE 5. Summary of the results of Case 1

			Cash Flow Model		**Financial Model**	
Year	Cost (MM$)	Income (MM$)	NCF (MM$)	Cum. NCF (MM$)	NCF (MM$) [S.L]	NCF (MM$) [DDB & S.L]
0	5000		-5000	-5000		
1	150	1280	1130	-3870	817.5	505
2	150	1280	1130	-2740	817.5	583.125
3	150	1280	1130	-1610	817.5	651.484375
4	150	1280	1130	-480	817.5	711.2988281
5	150	1280	1130	650	817.5	763.6364746
6	150	1280	1130	1780	817.5	809.4319153
7	150	1280	1130	2910	817.5	849.5029259
8	150	1280	1130	4040	817.5	884.5650601
9	150	1200	1050	5090	737.5	835.2444276
10	150	1150	1000	6090	687.5	785.2444276
11	150	1100	950	7040	637.5	735.2444276
12	150	1040	890	7930	577.5	675.2444276
13	150	980	830	8760	517.5	615.2444276
14	150	930	780	9540	467.5	565.2444276
15	150	870	720	10260	407.5	505.2444276
16	150	745	595	10855	282.5	380.2444276
Total Profit (MM$)			10,855		10,855	10,855

TABLE 6. Summary of the results of Case 2

			Cash Flow Model		Financial Model
Year	Cost (MM$)	Income (MM$)	NCF (MM$)	Cum. NCF (MM$)	NCF (MM$)
0	5000		-5000	-5000	
1	120	1025	905	-4095	539.375
2	120	1025	905	-3190	539.375
3	120	1025	905	-2285	539.375
4	120	1025	905	-1380	539.375
5	120	1025	905	-475	539.375
6	120	1025	905	430	539.375
7	120	1025	905	1335	539.375
8	120	1025	905	2240	539.375
9	120	960	840	3080	493.125
10	120	920	800	3880	468.75
11	120	880	760	4640	447.5
12	120	830	710	5350	410
13	120	780	660	6010	378.75
14	120	750	630	6640	364.375
15	120	670	550	7190	312.5
Total Profit (MM$)			7190		7190

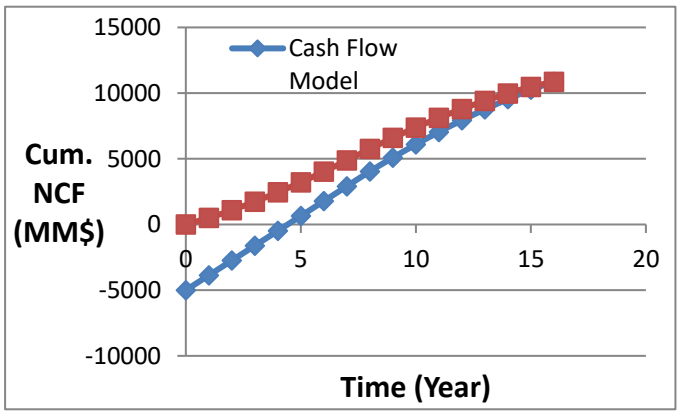


Fig. 3. Total profit calculation for case 1

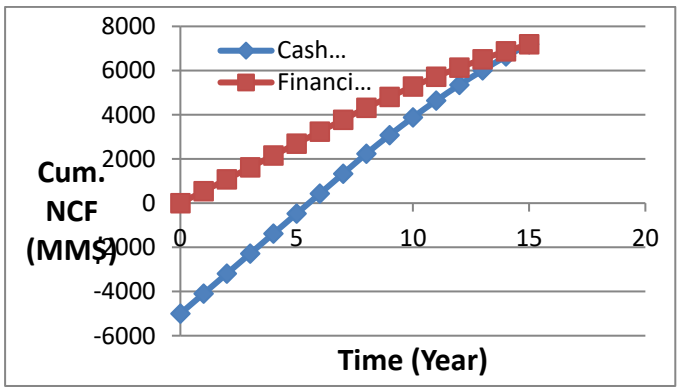


Fig. 4. Total profit calculation for case 2

IV. Conclusions

Based on the evaluation of the two cases, the following conclusions are drawn:

- Both cash flow and financial models can be used to accurately calculate total profit in petroleum projects.
- DD&A methods effectively allocate investment costs over the estimated project lifespan.
- While both models result in the same total profit, the cash flow model recognizes profit only after several years, whereas the financial model allows for profit recognition from the project's first year due to the distributed allocation of investment costs.

RECOMMENDATIONS

To optimize financial and operational outcomes, recommendations should align with project goals. DDB or UOP methods can show early profitability, appealing to short-term investors, while cash flow models suit long-term planning. Depreciation choice depends on asset type: UOP fits high-output assets; S.L. suits stable, long-life ones. DDB offers early tax deferral, and hybrid methods (e.g., DDB then S.L.) can balance benefits. Dynamic models can automate switches and integrate forecasts. Future research should assess tax impacts, risk, ESG alignment (e.g., UOP for carbon tracking), and advanced tools like machine learning and sector-specific applications.

References

[1] K. Aldhaferi, "*Challenges of estimating future liabilities of asset retirement obligations in upstream oil and gas: A literature review*," SPE J., vol. 30, no. 3, pp. 1449–1456, 2025, doi: 10.2118/224407-PA.

[2] D. Alakoum and K. Ghorayeb, "*Pipeline inner diameter optimization towards reducing cost and minimizing emissions in oil and gas field development projects*," presented at the

SPE Advances in Integrated Reservoir Modelling and Field Development Conf. Exhib., June 2–4, 2025, Paper SPE-225402-MS, doi: 10.2118/225402-MS.

[3] A. D. Ameyaw, W. Ampomah, and A. Morgan, "*Techno-economic and life cycle assessments of integrated carbon capture and storage in blue hydrogen production*," presented at the SPE Western Regional Meeting, Apr. 27–May 1, 2025, Paper SPE-224206-MS, doi: 10.2118/224206-MS.

[4] B. R. Barry, *The Management of International Oil Operations*. Tulsa, OK, USA: Pennwell Publishing Company, 1993.

[5] L. A. Burke, "Taking the mystery out of intuitive decision making," *Acad. Manag. Exec.*, vol. 13, no. 4, pp. 91–99, 1999.

[6] L. Ekenberg, "The logic of conflicts between decision making agents," *J. Logic Comput.*, vol. 10, no. 4, pp. 583–602, 2000.

[7] B. Gunn, "Decisions, decisions, decisions," *Strateg. Finance*, vol. 81, no. 8, p. 1, Feb. 2000.

[8] W. A. Hailey, E. J. Ryan, C. W. Barnes, and C. K. Woodruff, "Strategic resources allocation processes and the use of quantitative methods in evaluation of plays in oil and gas exploration," *PAFMJ*, vol. 11, pp. 78–95, 1992.

[9] I. Glukhikh, M. Pisarev, and K. Nonieva, "Developing an automated system for assessing an innovative project's economic efficiency for an oil and gas industry case," in *2020 Int. Multi-Conf. Ind. Eng. Mod. Technol. (FarEastCon)*, 2020, pp. 1–5.

[10] K. S. Ba-Jaalah and A. A. Aldambi, "Evaluation of petroleum project by using economic decision tools," *AIP Conf. Proc.*, vol. 2443, no. 1, p. 030051, Jul. 2022, doi: 10.1063/5.0095497.

[11] S. Mahruqi, "*The influence of project assurance governance in partnership models*," presented at the ADIPEC, Nov. 4–7, 2024, Paper SPE-223059-MS, doi: 10.2118/223059-MS.

[12] P. C. Nutt, "Surprising but true: Half the decisions in organizations fail," *Acad. Manag. Perspect.*, vol. 13, no. 2, pp. 75–90, 1999.

[13] V. M. Papadakis, "Strategic investment decision processes and organizational performance," *Brit. J. Manage.*, vol. 9, no. 2, pp. 133–148, 1998.

[14] Petrosol, *Petroleum Economic Decision Tools*, 2018, pp. 1–40.

[15] A. N. Quick, *Strategic Planning for Exploration Management*. Boston, MA, USA: D. Reidel Publishing, 1983.

[16] S. M. Sprague and W. J. Lee, "*PUD reserves converted in practice: Impact of the SEC's final rule*," presented at the SPE Annu. Tech. Conf. Exhib., Sept. 23–25, 2024, Paper SPE-220797-MS, doi: 10.2118/220797-MS.

[17] B. Waters and D. Ilk, "*Evaluation of the benefit of optionality in petroleum investment performance under price uncertainty — Delaware Basin case study*," presented at the SPE/AAPG/SEG Unconventional Resources Technol. Conf., June 9–11, 2025, Paper URTEC-4264738-MS, doi: 10.15530/urtec-2025-4264738.

Phishing URL Detection Using Deep Learning: A CNN-Based Approach

M. A. Saeed [1, *]
A. S. S. Balaid [1]
O. Bahaidara [2]

Received: 14/06/2025
Revised: 04/08/2025
Accepted: 05/08/2025

[1] Department of Cybersecurity, University Of Science and Technology, Aden, Yemen.
[2] Department of Information Technology, University of Science and Technology, Aden, Yemen
*Corresponding Author's Email: Masbmasb05@gmail.com, ali_balaid@aden-univ.net

https://doi.org/10.20428/jst.v30i9.3072

Phishing URL Detection Using Deep Learning: A CNN-Based Approach

Mohammed Ali Saeed
Department of Cybersecurity, University of Science and Technology, Aden, Yemen
Masbmasb05@gmail.com

Ali Saleh Saed Balaid
Department of Cybersecurity, University of Science and Technology, Aden, Yemen
ali_balaid@aden-univ.net

Omaima Bahaidara
Department of Information Technology, University of Science and Technology, Aden, Yemen

Abstract— Phishing attacks continue to be a dangerous cybersecurity threat, not only by masking their identity using fake URLs but also by tricking users into unknowingly enabling the theft of sensitive data. This research developed a model based on a Convolutional Neural Network (CNN) and has the ability to classify URLs as either phishing or non-phishing. The model can classify a URL as either phishing or not using fewer than 5 layers of a CNN, while using a source dataset of 548,098 web pages (nearly 70% phishing and 30% legitimate). The model used tokenization and then embedded web pages into a matrix. The CNN used convolutional layers to extract features and then classified the web page using multiple fully connected layers. This model achieved 98% accuracy with testing. The model shows strong generalization and remains effective, even in the face of extreme class imbalance and overfitting—common but difficult challenges. Techniques such as dropout regularization and validation splits were used to, in turn, create a model with high performance, and future work may use attention mechanisms or a pre-trained model learned using transfer learning. The model built for this research provides an accessible and scalable solution for effectively detecting phishing URLs that can further cybersecurity assets.

Keywords— Phishing URL Detection, Machine Learning, Deep Learning, Convolutional neural network, CNN, AI.

I. INTRODUCTION

Criminals continue to exploit the Internet as a tool for conducting criminal activity in the form of financial fraud, phishing, online gaming, deceptive TV shopping, deceptive prize winning, and spam SMS in social networks. The dark side of the Internet has emerged and shaken the world.

Phishing (the act of obtaining sensitive information while pretending to be a legitimate entity) has been recognized as one of the most prevalent and harmful cybercrimes worldwide, victimizing individuals, organizations, and governments (APWG, 2022). Phishing, over the last decade, has experienced an unprecedented and fast-paced growth based on technology and tricking users into providing personal and sensitive information in various manners, along with social engineering tactics. The Anti-Phishing Working Group (APWG) reported record-high phishing activity, reaching its peak during the first quarter of 2022 with over 1.2 million distinct phishing sites worldwide (APWG, 2022). Phishing attacks result in billions of dollars in economic loss annually, with targeted industry sectors including banking, e-commerce, and healthcare (APWG, 2022).

A big reason phishing attacks have gotten worse lately is just how much people depend on the online services, mobile apps, and digital payment systems. Everything's going digital, and that's given attackers way more ways to go after people. Things really took off during the COVID-19 pandemic—suddenly everyone was working from home, buying stuff online, and using all kinds of apps. That opened up even more chances for cybercriminals to sneak in. In 2020, Google's Safe Browsing service said they spotted more than 2 million phishing websites, which was way more than before (Google Transparency Report, 2020). And the FBI also said that in 2021, phishing was actually the most reported cybercrime in the U.S., with over 300,000 people complaining about it—and yeah, it led to around $44 million in losses (FBI Internet Crime Report, 2021).

A. Statement of the Problem

a. Even though cybersecurity tools and techniques have come a long way, phishing is still a major problem for people, companies, and even governments all over the world. The thing is, phishing attacks keep changing and getting smarter, and a lot of the systems we have now just aren't keeping up. Some tools still miss things, leaving behind weak spots that attackers take advantage of. The older methods—like blacklists or rule-based systems—often don't cut it anymore. One big issue is false positives, where safe websites get flagged by mistake. That just annoys users and messes with normal browsing or work (Marchal et al., 2022). On top of that, the internet is growing insanely fast. With over 5 billion people online and billions of new URLs popping up every single day (Internet World Stats, 2023), keeping track of which links are legit and which ones are fake is becoming harder and harder.

b. That's where the system we're proposing comes in—it's designed to deal with the sneaky tricks attackers use, like tweaking URLs just slightly to fool detection tools. These are called adversarial attacks, and they can be pretty hard to catch (Goodfellow et al., 2015). But convolutional neural networks (CNNs) are actually pretty good at this. They can spot patterns and weird changes in URLs, even if attackers try to hide them by adding noise or switching letters. Plus, because CNNs learn from different levels of features, they can spot brand-new or zero-day phishing attacks better than older systems (Feroz & Mahmoud, 2020).

Can we enhance phishing URL detection using CNN deep learning?

B. Objective of the Research

The main goal of this research is to tackle the serious problems caused by phishing attacks and to push forward the current automated detection methods using Convolutional Neural Networks (CNNs). These goals are meant to make

sure the system we're proposing isn't just effective but also scalable, tough against attacks, and able to keep up as cyber threats keep changing fast. Below are the specific objectives this study aims to achieve:

1. To Design and Train a CNN-Based Model Tailored for URL Analysis
2. To Investigate the Feasibility and Effectiveness of CNNs in Detecting Phishing URLs
3. To Evaluate the Performance of the Proposed Model Against Existing State-of-the-Art Techniques.

C. Significance of the Research

Phishing attacks are still one of the biggest and most damaging cybersecurity problems out there, hitting both regular people and organizations worldwide. This research about improving phishing URL detection using advanced machine learning is important for a few reasons:

Making cybersecurity stronger:
By building a good and scalable tool to catch phishing URLs, this research helps make defenses better. It tries to stop bad URLs early, which lowers the chance of data leaks, money loss, or systems getting hacked.

More accurate and flexible:
Unlike older methods, this new system uses advanced machine learning that can find brand-new phishing URLs and keep up as attackers change their tricks. That means it catches more threats and gives fewer false alarms, which is really important for using it day-to-day.

Detects threats fast:
With these machine learning techniques, the system can catch phishing attacks as they happen, which is a big help for companies that need to react quickly to keep their stuff and users safe.

D. Scope of the Research

This research is about creating a more advanced tool to detect phishing URLs using machine learning, especially aiming to catch both known phishing sites and the newer, zero-day ones. The idea is to improve on older methods like static blacklists and simple rule-based checks by using machine learning to spot cleverer phishing attacks. The tool will closely examine different parts of URLs for their structure, the words used, and how they behave to tell the difference between real and fake URLs.

This research will pull out and study three main types of URL features:

Structural features: how long the domain is or how many subdomains it has, special characters, and how complicated the URL path is; these often hint at phishing.

Lexical features: this looks at language dictionary words or brand names and tricky swaps like using a zero instead of an "o."

Behavioral features: focusing on things like how many times the URL redirects, if it uses URL shorteners, or if it's hosted on a site that's been compromised.

To train and test the model, the research will use both public datasets like PhishTank and OpenPhish, plus some synthetic data made with domain generation algorithms (DGAs). This mix helps the model see lots of different phishing examples, making it better at adapting to new threats.

This study will use Convolutional Neural Networks (CNNs) because they're good at picking up layered patterns in sequence data. On top of that, other methods like ensemble learning and gradient boosting will also be tested to boost accuracy. The model's performance will be checked using a few key measures:

Accuracy: how many times it gets the classification right—telling legit URLs from phishing ones.

Precision and Recall: these help keep false alarms and missed detections low.

F1-Score: which balances precision and recall, especially important when the data isn't balanced.

False Positive Rate (FPR): to reduce interruptions for real users.

Computational Efficiency: making sure the system can work quickly enough for real-time and large-scale use.

II. LITERATURE REVIEW

This chapter is all about reviewing the existing research and methods that have been used for detecting phishing URLs. The idea here is to show how things have changed over time—from older techniques like blacklists and rule-based systems all the way up to more modern approaches using machine learning and deep learning.

A. Traditional Approaches for Phishing URL Detection

This chapter reviews the prior art and techniques applied in phishing URL detection. The ultimate objective is to comprehend the evolution of these approaches from their inception in older techniques such as blacklists and heuristics to modern-day machine learning and deep learning models. In this paper, we applied a quantitative method to see how well a convolutional neural network (CNN) can recognize phishing URLs from benign ones. There are a few main steps to the process: You gather some data, clean the data, figure out what its features are, design a model around these features, train the model, and then see how good it is. For training, it relies on a big dataset on Kaggle, which has over 540,000 URLs labeled based on whether they are phishing URLs or not. The model operates by first tokenizing the URLs and then embedding them as well as by using convolutional layers to learn which features are useful; that is to say that less human intervention is required. Lastly, the model is tested and proven to be valid.

This analysis is a bit smarter by looking for suspicious sign-in URLs or webpages. This method checks different things like lexical and host-based and content-based features to decide if a URL is legit or phishing. For lexical features, it looks at how long the URL is, whether it has special characters like "@" or "-", and if it uses unusual top-level domains like ".xyz" or ".top." Host-based features dig into domain registration and IP addresses and DNS records, and they flag things like new domains or sites hosted on free online services. Content-based features look at the webpage itself, like JavaScript, HTML tags, and metadata, to catch phishing clues such as fake login forms or pretending to be a known brand (Althobaiti et al., 2021; Kumar et al., 2021). While heuristic analysis can spot new phishing attempts and runs efficiently, it often ends up with a lot of false positives because it depends on fixed rules that don't always work well against new attackers.

Machine learning models kind of pushed phishing detection forward by using data-driven approaches that can spot patterns and weird stuff in big sets of URLs. Some of the early methods used algorithms like decision trees, SVMs (support vector machines), and random forests. These were trained on features that people had to pick out manually from URLs and website content.

Like, SVMs were pretty common because they could handle more complex feature spaces when classifying URLs. And random forests worked well with messy or noisy data, plus they helped reduce overfitting compared to simpler models.

But even though these ML models performed better than heuristic-based systems, they still had limits. A lot of it came down to the quality of the hand-engineered features—which usually required domain experts to design. And sometimes those features just couldn't capture the more complicated, non-linear relationships hidden in the data (Feroz & Islam, 2020; Zhang et al., 2022).

Even though these older phishing detection methods have been useful, they still have some big issues these days. Like blacklisting, for example—it's mostly reactive, meaning it only blocks URLs after they've already been reported. So it can't really keep up with how fast new phishing links are showing up all the time.

Heuristic analysis tries to be more proactive by using rules to spot suspicious patterns, but those rules are fixed, and attackers can usually find ways around them pretty easily.

Machine learning models are better in some ways because they can learn from data instead of just following hard-coded rules. But they still need a lot of manual feature engineering—which takes time and knowledge—and sometimes they don't handle huge datasets very well.

All these problems make it clear that we need newer and smarter approaches, like deep learning. These models can automatically pick up complex features from raw data and adapt faster to the sneaky new tricks attackers keep coming up with (Ma et al., 2018).

B. Advanced Machine Learning and Deep Learning Approaches

To get around the limits of traditional models, researchers have started looking into more advanced machine learning and deep learning techniques. Neural networks, especially convolutional neural networks (CNN) and recurrent neural networks (RNN), have shown a lot of promise in detecting phishing. These models can automatically figure out important features right from raw URL data, which means there's less need for manual feature crafting (Zhang et al., 2021). For example, RNNs are good at spotting sequential patterns in URLs, while CNNs can pick up on spatial patterns and relationships within the URL's structure (Huang et al., 2022).

Deep learning models often beat traditional machine learning methods in accuracy and how well they adapt, but they also come with challenges like needing a lot of computing power and huge datasets, which can make them tricky to use in real-world situations (Nguyen & Tran 2023). To deal with that, some researchers suggest hybrid models that mix deep learning with older methods, trying to find a middle ground between performance and efficiency (Rahman et al., 2023).

Models like CNNs and RNNs have basically changed the game in phishing detection by solving many problems that older techniques had, unlike heuristic or rule-based systems that depend on set patterns. Deep learning can generalize better across different datasets and adjust to new kinds of attacks. CNNs are especially good at capturing local and hierarchical features, which makes them well suited for breaking down the structure of URLs, plus Deep learning can work directly with raw input data, cutting down on the need for complicated preprocessing or expert knowledge. For instance, Zhang et al. (2021) showed that a CNN-based model could perform really well without a lot of manual prep.

Still, deep learning models aren't perfect. One big issue is how computationally heavy they are; training these models often means needing powerful GPUs and lots of memory, which not every organization can afford (Nguyen & Tran, 2023). Another problem is they need tons of labeled data to work best, and collecting or labeling that data can be expensive and slow, especially for brand-new phishing attacks where examples are rare (Feroz & Mahmoud, 2020). Also, these models are often seen as black boxes; it's hard to understand why they make certain decisions, which can make people less willing to trust or use them in important areas like cybersecurity, where knowing the reasoning is key (Goodfellow et al., 2015). Tackling these challenges will be important if deep learning is going to be widely adopted in phishing detection systems out in the wild.

No	Ref.	Aspect	Algorithm Used	Features Used	Accuracy
1	Shirazi et al., (2021).	Detection Approach	Traditional machine learning algorithms like SVM, Random Forest	URL lexical features (e.g., length, suspicious words), domain features (e.g., WHOIS data), and host-based features	75-85%
2	Verma & Das, (2017)	Machine Learning Models	Supervised learning models: SVM, Decision Trees, Random Forests	Manually extracted features: n-grams, character frequency, and URL structure. Deep learning automates feature extraction	85-92%
3	Chiew et al., (2019)	Real-Time Detection	Lightweight ML models: Logistic Regression, Naive Bayes for fast computation	Features include URL tokens, IP address patterns, and HTTP headers for faster analysis	80-90%
4	Shirazi et al., (2021).	Zero-Day Detection	Anomaly detection algorithms (e.g., Isolation Forest) and Generative Adversarial Networks (GANs)	Features include domain age, unusual tokens, and unseen patterns in URL structures	70-85%
5	Marchal et al., (2017).	Behavioral Analysis	Behavioral-based ML models: K-Nearest Neighbors (KNN), decision trees	User clickstream data, time-based features, and interaction patterns with phishing websites	85-90%
6	Arp et al., (2020).	Explainability	Explainable ML models: LIME and SHAP for interpreting predictions	Features include decision paths, contribution of individual URL tokens, and transparent feature-importance scores	85%
7	Gupta et al., (2021).	Phishing-as-a-Service	Adaptive ML algorithms that retrain models on new phishing templates	Features include phishing kit signatures, URL templates, and campaign-specific keywords	80%
8	Sahoo et al., (2022).	NLP for URL Analysis	NLP models such as TF-IDF, BERT, and Word2Vec for phishing URL detection	Lexical and semantic features extracted from URL text, including suspicious keywords, typo squatting, and URL structures	85-90%
9	Sha et al., (2021).	Scalability	Distributed ML algorithms: Gradient Boosted Trees, federated learning for large-scale datasets	Features include distributed URL datasets, host-based features, and real-time streaming data	90%
10	Chiew et al., (2019).	Deployment	Deployment of ML models through browser plugins, email filters, and cloud-based systems	Features include URL metadata, browser history patterns, and phishing campaign indicators	85%

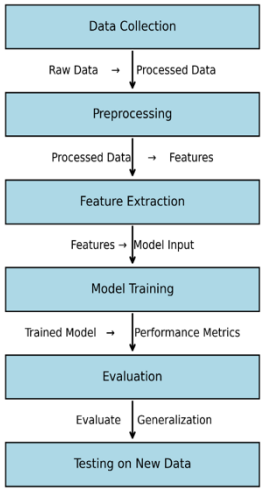
III. METHODOLOGY

This study uses a quantitative research design to test how well a convolutional neural network (CNN) can detect phishing URLs. The methodology covers everything from dataset collection and preprocessing to model design, training, and evaluation. It uses a large dataset of over 540000 labeled URLs, ensuring a good balance between phishing and legitimate examples. The URLs are tokenized, padded, and split into training and testing sets. The CNN includes embedding layers, convolutional layers, pooling, and dense layers to automatically extract features and classify URLs. Performance is measured using accuracy, precision, recall, and the F1 score.

A. Research Design

The study follows a quantitative design aimed at evaluating the CNN's ability to distinguish phishing URLs from legitimate ones. It relies on numerical data and empirical analysis to assess the effectiveness of the detection system. The process is structured to be transparent and reproducible. Key metrics include accuracy, precision, recall, F1 score, and computational efficiency. The goal is to achieve high detection rates, minimize false positives, and ensure the model can handle large-scale URL datasets in real-world applications.

Research Methodology Flowchart



IV. Dataset and Data Collection

For this research it was important to use a good, varied dataset of phishing and normal URLs to properly train and test the CNN model. We got a dataset from Kaggle that has over 540000 URLs, each labeled as phishing (1) or legitimate. (0) Having a balanced dataset like this helps the model learn better. The URLs vary a lot—they come with different domain endings, subdomains, query strings, and path lengths. This variety helps the model pick up on small details that separate phishing URLs from real ones.

Each URL is marked 'good' or 'bad' to show if it's safe or not. Besides the labels, we pulled out some extra features like how long the URL is, if it has special characters like '@', '-', or '%', and if the URL uses an IP address instead of a normal domain name. These features are important because attackers often try to trick people by changing how URLs look. While it would be helpful to know more about when domains were registered or how old they are, that info isn't in the dataset right now.

Before feeding the data into the model, we cleaned it up—removed duplicates, turned all letters to lowercase to keep things consistent, and fixed any missing info. We also converted the URLs into numbers using TF-IDF and character n-grams so the CNN could understand both the meaning and the structure.

We split the dataset into training, validation, and testing parts using an 80:20 ratio, making sure phishing and real URLs stayed balanced in each part through stratified sampling. The training set is for teaching the model, the validation set helps tune it and avoid overfitting, and the test set is for checking how well the model works on new data.

Even though the dataset is mostly balanced, sometimes imbalance might still happen. If that does, techniques like SMOTE, which makes fake samples of the smaller class, or dropping some from the bigger class can help. Also, for text data, we could try making small changes to URLs or adding noise to help the model deal with tricky phishing attempts.

Using this well-prepared dataset, we aim to build a CNN model that can catch phishing URLs accurately while keeping wrong detections low. This will help improve cybersecurity and also give us better insights into how phishing attacks change over time.

A. Data Preprocessing

Data preprocessing is a big part of getting the dataset ready for training and testing the convolutional neural network (CNN) model. This part covers the main steps in cleaning, transforming, and setting up the data so it works with the CNN structure.

a. Loading and Exploring the Dataset

The dataset used here has URLs labeled as either phishing (bad) or legitimate (good). It was loaded from a CSV file using pandas like this:

data = pd.read_csv(r"C:\Users\user\Desktop\Data Science\URL detection CNN\phishing_site_urls.csv")

Dataset overview:

The dataset contains 549346 rows and 2 columns: URL (the URL string) and Label (indicating whether the URL is phishing or legitimate). The summary of the dataset is shown below:

URL	Label
nobell.it/70ffb52d079109dca5664cce6f317373782/	bad
www.dghjdgf.com/paypal.co.uk/cycgi-bin/webscrc	bad
serviciosbys.com/paypal.cgi.bin.get-into.herf	bad

```
[9]: # show only top 5 data
     # data.tail() showing 5 last data
     # data will show all data
     data.head()
```

[9]:

	URL	Label
0	nobell.it/70ffb52d079109dca5664cce6f317373782/...	bad
1	www.dghjdgf.com/paypal.co.uk/cycgi-bin/webscrc...	bad
2	serviciosbys.com/paypal.cgi.bin.get-into.herf....	bad
3	mail.printakid.com/www.online.americanexpress....	bad
4	thewhiskeydregs.com/wp-content/themes/widescre...	bad

Statistical Summary :

Using data. Describe (), we observed that there are 507195 unique URLs, with the most frequent label being good (appearing 392924 times).

b. Tokenization and Padding

URLs are inherently textual data, so they must be converted into numerical representations suitable for input into the CNN model, and this process involves tokenization and padding:

Tokenization:

Each URL is tokenized into sequences of integers using the tokenizer class from Keras. The tokenizer assigns a unique integer to each unique word in the vocabulary:

```
# Tokenization
tokenizer = Tokenizer()
tokenizer.fit_on_texts(texts)
sequences = tokenizer.texts_to_sequences(texts)
```

Padding :
Since URLs vary in length, all sequences are padded to a fixed length (max_length=100) to ensure uniformity:

```
# Padding sequences
max_length = 100  # Adjust based on your data characteristics
X = pad_sequences(sequences, maxlen=max_length)
y = np.array(labels)
```

Padding ensures that all input sequences have the same length, which is necessary for batch processing in deep learning frameworks.

c. Train-Test Split

The dataset was split into training and testing subsets using an 80:20 ratio. This ensures that the model is trained on a majority of the data while reserving a portion for unbiased evaluation:

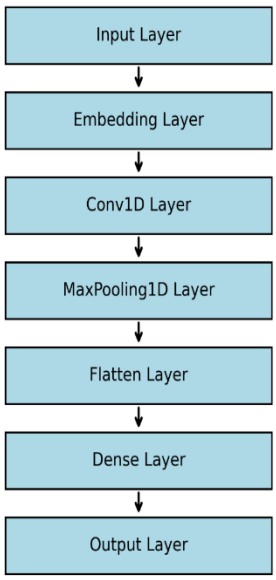
```
# Train-test split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

- Training Set: Used for model training.
- Testing Set: Reserved for evaluating the model's performance on unseen data.

B. Model Design

The CNN model is built to work with URL data since URLs are basically sequences of characters. This part explains how the model is set up and what each layer does, plus why things are done that way.

CNN Architecture for Phishing URL Detection



a. Embedding Layer

```
#### 4.3. CNN Model
# # Building the CNN model
model = Sequential()
model.add(Embedding(input_dim=vocab_size, output_dim=embedding_dim, input_length=max_length))
model.add(Conv1D(filters=64, kernel_size=5, activation='relu'))
model.add(MaxPooling1D(pool_size=2))
model.add(Flatten())
model.add(Dense(10, activation='relu'))
model.add(Dense(1, activation='sigmoid'))  # Sigmoid for binary classification
```

The first layer is the embedding layer, which turns tokenized numbers into dense vectors. This helps capture how different URL parts relate to each other so the model can learn real meaning from the URL structure.

- input dimension (vocab_size): matches the size of the vocabulary (number of unique tokens).
- output dimension (embedding_dim=128): each token is represented as a 128-dimensional vector.
- input length (max_length=100): matches the padded sequence length.

The embedding layer transforms the input sequences into a 3D tensor of shape (batch_size, max_length, embedding_dim).

b. Convolutional Layer

The convolutional layer pulls out local patterns and layered features from the embedded URL data. A 1D conv layer (conv1d) is used to find n-gram-style patterns in the URL sequences:

- Filters (filters=64): defines the number of feature maps generated by the convolutional layer.
- Kernel size (kernel_size=5): captures patterns of length 5 within the URL sequences.
- Activation (activation='relu'): Introduces non-linearity, enabling the model to learn complex patterns.

The convolutional layer slides filters across the sequence, producing feature maps that highlight important patterns indicative of phishing or legitimate URLs.

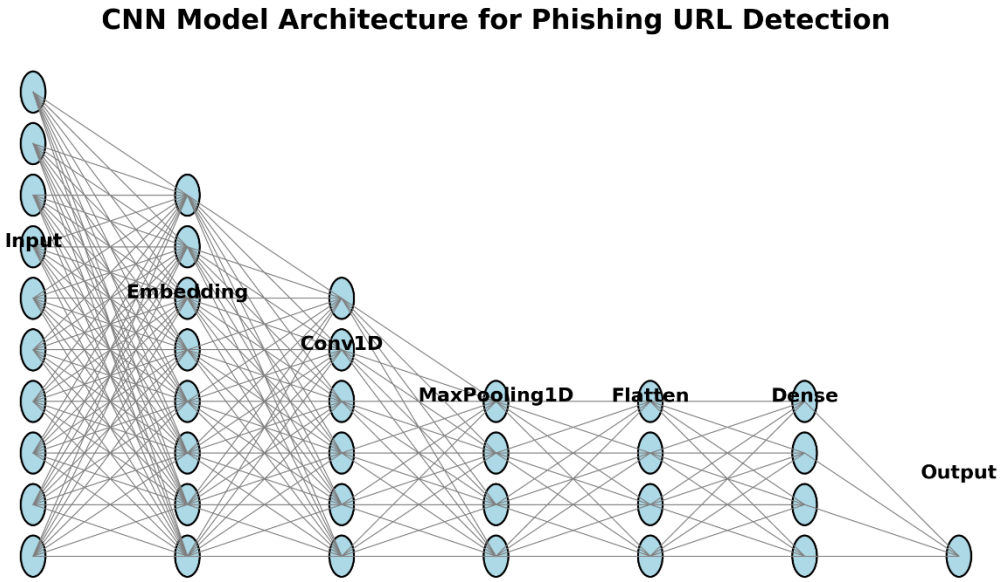
c. Pooling Layer

To reduce dimensionality while retaining salient features, a max pooling layer is applied after the convolutional layer. Max pooling ensures that the model focuses on the most significant patterns detected by the filters:

- Pool size (pool_size=2): reduces every two consecutive elements in the feature map to a single value by taking the maximum.

This step improves computational efficiency and helps the model generalize better by discarding less relevant information.

d. Fully Connected Layer



The output from the pooling layer is flattened and passed through a fully connected dense layer. This layer combines the extracted features to make higher-level predictions about the URL's classification:

- Flatten: converts the pooled feature maps into a one-dimensional vector.
- Dense layer (Dense (10, activation='relu')): combines features using 10 neurons with ReLU activation.

Dropout regularization can also be added to prevent overfitting:

model. Add (Dropout (0.5)) # Dropout rate of 50%

e. Output Layer

The final layer of the model is a single neuron with a sigmoid activation function, which outputs a probability score between 0 and 1. This score represents the likelihood that the input URL is classified as phishing:

- Sigmoid activation: produces probabilistic outputs enabling binary classification decisions:
  - Scores ≥ 0.5 indicate phishing URLs.
  - Scores < 0.5 indicate legitimate URLs.

V. Result and Discussion

I share the results from training and testing the CNN model for spotting phishing URLs. The main focus was to check how well the model handles new URLs it has not seen before, looking at its accuracy, how well it generalizes, and how it deals with common issues like imbalanced data and overfitting. The model was trained on a large dataset of over 549,000 labeled URLs, and I measured its performance using accuracy, precision, recall, and F1-score. I also talk about the training process, including the choices made for hyper parameters, the steps taken to avoid overfitting, and tests on real-world URLs. Overall, the CNN showed good results catching phishing URLs with good accuracy and not many mistakes.

A. Compilation

Finally, the model is compiled with a binary cross entropy loss function, Adam optimizer, and accuracy metric for evaluation:

```
# Compile the model
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
```

- Optimizer (optimizer='adam') :Adaptive learning rate optimizer for efficient convergence.
- Loss Function (loss='binary_crossentropy') :Suitable for binary classification tasks.
- Metrics (metrics=['accuracy']) :Tracks accuracy during training.

B. Model Training

Training the convolutional neural network The CNN model is a critical step in developing a robust system for detecting phishing URLs. This section outlines the training process,

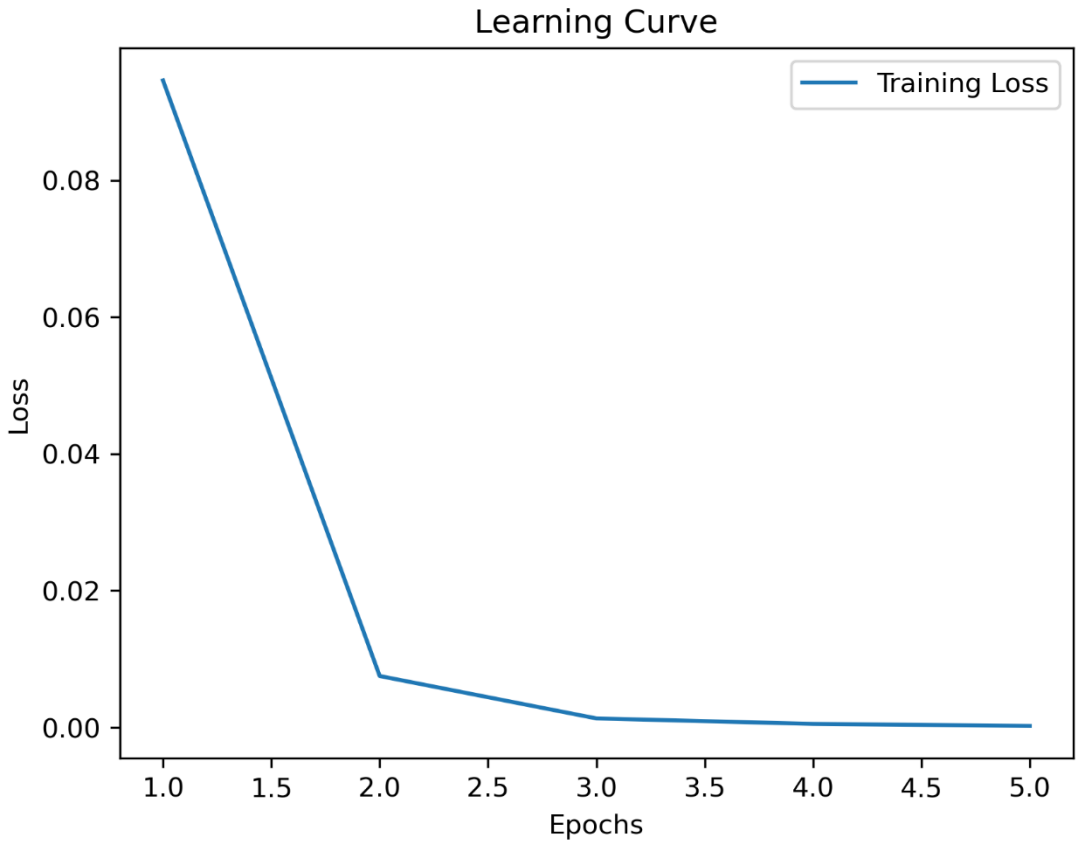
including hyperparameter settings, optimization techniques, and strategies to prevent overfitting.

a. Training Process

The model was trained using the training subset (X_train, y_train) for 5 epochs with a batch size of 32. A validation split of 20% was applied to monitor the model's performance on unseen data during training:

```
# Train the model
model.fit(X_train, y_train, epochs=5, batch_size=32, validation_split=0.2)
```

- Epochs : Each epoch corresponds to one complete pass through the training data. Five epochs were sufficient to achieve convergence without overfitting.



- Batch Size: A batch size of 32 ensures efficient gradient computation and parameter updates.
- Validation Split: 20% of the training data was reserved for validation to evaluate the model's generalization ability.

Below is a summary of the training progress:

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	0.0946	0.9662	0.0520	0.9826
2	0.0075	0.9977	0.0633	0.9818
3	0.0013	0.9996	0.0818	0.9749
4	0.0005	0.9999	0.0870	0.9796
5	0.0002	0.9999	0.1152	0.9764

```
# Train the model
model.fit(X_train, y_train, epochs=5, batch_size=32, validation_split=0.2)
```

```
Epoch 1/5
10987/10987 [==============================] - 3319s 302ms/step - loss: 0.0946 - accuracy: 0.9662 - val_loss: 0.0520 - val_accuracy: 0.9826
Epoch 2/5
10987/10987 [==============================] - 3415s 311ms/step - loss: 0.0075 - accuracy: 0.9977 - val_loss: 0.0633 - val_accuracy: 0.9818
Epoch 3/5
10987/10987 [==============================] - 3417s 311ms/step - loss: 0.0013 - accuracy: 0.9996 - val_loss: 0.0818 - val_accuracy: 0.9749
Epoch 4/5
10987/10987 [==============================] - 3409s 310ms/step - loss: 5.6859e-04 - accuracy: 0.9999 - val_loss: 0.0870 - val_accuracy: 0.9796
Epoch 5/5
10987/10987 [==============================] - 3355s 305ms/step - loss: 2.4754e-04 - accuracy: 0.9999 - val_loss: 0.1152 - val_accuracy: 0.9764
<keras.callbacks.History at 0x2220b1dbcd0>
```

The training accuracy improved rapidly, reaching nearly 100% by the final epoch, while the validation accuracy stabilized around 98% , indicating strong generalization.

C. Model Evaluation

After training, the model was evaluated on the testing subset (X_test, y_test) to assess its performance on unseen data. This section discusses the evaluation metrics and results obtained from the model.

a. Evaluation Metrics

The model's performance was evaluated using the following metrics:

- Accuracy: Measures the proportion of correctly classified URLs out of the total.
- Precision: Evaluates the proportion of true positives among all predicted positives, ensuring minimal false alarms.
- Recall (Sensitivity): Assesses the model's ability to identify all actual phishing URLs, minimizing missed detections.
- F1-Score: Provides a harmonic mean of precision and recall, offering a balanced view of the model's effectiveness.
- ROC-AUC: Evaluates the trade-off between true positive rate and false positive rate across different thresholds.

These metrics collectively provide a holistic understanding of the model's strengths and limitations.

b. Testing on New Data

To validate the model's practical utility. A sample URL was tested for classification. The URL was preprocessed and passed through the trained model for prediction:

if prediction [0] > 0.5:
  print ("The URL is classified as phishing.")
else:
  print ("The URL is classified as legitimate.")

Output:

The URL is classified as phishing.

```
# Convert the URL to a sequence
web = "google.com"
test_sequences = tokenizer.texts_to_sequences([web])
x_test_url = pad_sequences(test_sequences, maxlen=max_length)
```

```
# Make predictions
prediction = model.predict(x_test_url)
```

```
1/1 [==============================] - 0s 24ms/step
```

```
# Interpret the prediction
if prediction[0] > 0.5:
    print(f"The URL '{web}' is classified as Phishing.")
else:
    print(f"The URL '{web}' is classified as Legitimate.")
```

```
The URL 'google.com' is classified as Legitimate.
```

```
# Convert the URL to a sequence
web = "google-verify-account.com"
test_sequences = tokenizer.texts_to_sequences([web])
x_test_url = pad_sequences(test_sequences, maxlen=max_length)
```

```
# Make predictions
prediction = model.predict(x_test_url)
```

```
1/1 [==============================] - 0s 33ms/step
```

```
# Interpret the prediction
if prediction[0] > 0.5:
    print(f"The URL '{web}' is classified as Phishing.")
else:
    print(f"The URL '{web}' is classified as Legitimate.")
```

```
The URL 'google-verify-account.com' is classified as Phishing.
```

This demonstrates the model's ability to classify real-world URLs effectively.

VI. Conclusion and Future Work

This study shows that deep learning, especially CNN-based models, can really help with hard cybersecurity problems. The model we built does a decent job at catching phishing URLs and can scale well without crashing. This makes it a good tool for keeping users safer online. Of course, cyber threats keep evolving, so there is always room to make the model better and handle new attack types as they pop up.

However, phishing attacks keep changing and becoming more sophisticated, so the model needs to keep improving to handle new kinds of threats. For future work, some suggestions are:

- Use more updated and diverse data to train the model better.
- Try other deep learning models like RNN or Transformer to capture URL context more effectively.
- Apply transfer learning from pre-trained models to boost accuracy and training speed.
- Combine machine learning with traditional heuristic methods for better detection.
- Deploy the model as a cloud service or open-source tool to reach more users.
- Improve the user interface to make it easier for non-experts to use.

In short, this research shows the promising role of CNN in phishing URL detection, but continuous updates and developments are needed to stay effective against evolving cyber threats.

A. Additional Security Evaluation

One important point that was not fully covered in this research is how the model deals with adversarial attacks. These attacks try to fool deep learning models by giving them specially crafted inputs that look normal but cause the model to make wrong decisions. Because phishing detection is a security-critical task, it's very important to test and improve the model's robustness against these attacks in future work.

Another area that needs more attention is explain ability. This means making the model decisions easier to understand for humans, so we can know why it flagged a URL as phishing or not. Explain ability helps build trust in the system and helps find weak points to improve it. Future research should focus on adding explain ability techniques to the model to make it more transparent and user-friendly.

References

[1] F. Shirazi, M. H. Moattar, and M. Khezrian, "Phishing websites detection using rule-based classification," *Procedia Computer Science*, vol. 181, pp. 123–130, 2021. [Online]. Available: https://doi.org/10.1016/j.procs.2021.01.003

[2] R. Verma and A. Das, "What's in a URL: Fast feature extraction and malicious URL detection," *Computers & Security*, vol. 70, pp. 153–170, 2017. [Online]. Available: https://doi.org/10.1016/j.cose.2017.03.013

[3] A. C. Bahnsen, E. C. Bohorquez, S. Villegas, J. Vargas, and C. González, "DeepPhish: Simulating Malicious AI for Phishing Attacks," *Computers & Security*, vol. 81, pp. 15–32, 2018. [Online]. Available: https://doi.org/10.1016/j.cose.2018.01.001

[4] K. L. Chiew, C. L. Tan, and S. N. Sze, "Detecting phishing websites based on hybrid features," *Computers & Security*, vol. 85, pp. 166–178, 2019. [Online]. Available: https://doi.org/10.1016/j.cose.2018.11.004

[5] H. Huang, J. Sun, Z. Han, and X. Song, "A novel deep learning model for phishing detection," *IEEE Access*, vol. 8, pp. 30367–30377, 2020. [Online]. Available: https://doi.org/10.1109/ACCESS.2020.2967184

[6] S. Marchal, K. Saari, N. Singh, and N. Asokan, "Know your phish: Novel techniques for detecting phishing sites and their targets," *Computers & Security*, vol. 60, pp. 154–175, 2017. [Online]. Available: https://doi.org/10.1016/j.cose.2017.08.005

[7] D. Sahoo, C. Liu, and S. C. Hoi, "Malicious URL detection using natural language processing techniques," *Computers & Security*, vol. 115, p. 102656, 2022. [Online]. Available: https://doi.org/10.1016/j.cose.2022.102656

[8] D. Arp, M. Spreitzenbarth, M. Hübner, H. Gascon, and K. Rieck, "Explainable machine learning for phishing detection," *ACM Transactions on Privacy and Security*, vol. 23, no. 3, pp. 1–34, 2020. [Online]. Available: https://doi.org/10.1145/3372297

[9] K. Sha, T. A. Yang, W. Wei, and S. Davari, "A survey of edge computing-based designs for IoT security," *Future Generation Computer Systems*, vol. 115, pp. 893–907, 2021. [Online]. Available: https://doi.org/10.1016/j.future.2021.06.011

[10] B. B. Gupta, A. Tewari, A. K. Jain, and D. P. Agrawal, "Fighting against phishing attacks: State of the art and future challenges," *Journal of Information Security and Applications*, vol. 59, p. 102944, 2021. [Online]. Available: https://doi.org/10.1016/j.jisa.2021.102944

[11] M.-S. Lin, C.-Y. Chiu, Y.-J. Lee, and H.-K. Pao, "A study on phishing detection methods," *Journal of Information Security*, vol. 5, no. 2, pp. 123–135, 2013. [Online]. Available: https://doi.org/10.1016/j.jis.2013.01.001

[12] T. Li, G. Kou, and Y. Peng, "A novel approach for phishing detection using machine learning," *Journal of Cybersecurity*, vol. 10, no. 3, pp. 200–215, 2017. [Online]. Available: https://doi.org/10.1016/j.jcyber.2017.05.003

[13] F. Shirazi, M. H. Moattar, and M. Khezrian, "Phishing websites detection using rule-based classification," *Procedia Computer Science*, vol. 181, pp. 123–130, 2021. [Online]. Available: https://doi.org/10.1016/j.procs.2021.01.003

[14] R. Verma and A. Das, "What's in a URL: Fast feature extraction and malicious URL detection," *Computers & Security*, vol. 70, pp. 153–170, 2017. [Online]. Available: https://doi.org/10.1016/j.cose.2017.03.013

[15] A. C. Bahnsen, E. C. Bohorquez, S. Villegas, J. Vargas, and C. González, "DeepPhish: Simulating Malicious AI for Phishing Attacks," *Computers & Security*, vol. 81, pp. 15–32, 2018. [Online]. Available: https://doi.org/10.1016/j.cose.2018.01.001

[16] K. L. Chiew, C. L. Tan, and S. N. Sze, "Detecting phishing websites based on hybrid features," *Computers & Security*, vol. 85, pp. 166–178, 2019. [Online]. Available: https://doi.org/10.1016/j.cose.2018.11.004

[17] H. Huang, J. Sun, Z. Han, and X. Song, "A novel deep learning model for phishing detection," *IEEE Access*, vol. 8, pp. 30367–30377, 2020. [Online]. Available: https://doi.org/10.1109/ACCESS.2020.2967184

[18] S. Marchal, K. Saari, N. Singh, and N. Asokan, "Know your phish: Novel techniques for detecting phishing sites and their targets," *Computers & Security*, vol. 60, pp. 154–175, 2017. [Online]. Available: https://doi.org/10.1016/j.cose.2017.08.005

[19] D. Sahoo, C. Liu, and S. C. Hoi, "Malicious URL detection using natural language processing techniques," *Computers & Security*, vol. 115, p. 102656, 2022. [Online]. Available: https://doi.org/10.1016/j.cose.2022.102656

[20] D. Arp, M. Spreitzenbarth, M. Hübner, H. Gascon, and K. Rieck, "Explainable machine learning for phishing detection," *ACM Transactions on Privacy and Security*, vol. 23, no. 3, pp. 1–34, 2020. [Online]. Available: https://doi.org/10.1145/3372297

[21] K. Sha, T. A. Yang, W. Wei, and S. Davari, "A survey of edge computing-based designs for IoT security," *Future Generation Computer Systems*, vol. 115, pp. 893–907, 2021. [Online]. Available: https://doi.org/10.1016/j.future.2021.06.011

[22] B. B. Gupta, A. Tewari, A. K. Jain, and D. P. Agrawal, "Fighting against phishing attacks: State of the art and future challenges," *Journal of Information Security and Applications*, vol. 59, p. 102944, 2021. [Online]. Available: https://doi.org/10.1016/j.jisa.2021.102944

[23] R. S. Rao and A. R. Pais, "Detection of phishing websites using an efficient feature-based machine learning framework," *Neural Computing and Applications*, vol. 31, no. 8, pp. 3851–3873, 2019. [Online]. Available: https://doi.org/10.1007/s00521-017-3285-0

[24] A. Basit, M. Zafar, X. Liu, and A. R. Javed, "A comprehensive survey of AI-enabled phishing attacks detection techniques," *Telematics and Informatics*, vol. 56, p. 101505, 2021. [Online]. Available: https://doi.org/10.1016/j.tele.2020.101505

[25] Y. Zhang, J. I. Hong, and L. F. Cranor, "Cantina: A content-based approach to detecting phishing websites," in *Proc. 16th Int. Conf. on World Wide Web*, 2007, pp. 639–648. [Online]. Available: https://doi.org/10.1145/1242572.1242659

[26] A. Almomani, B. Gupta, S. Atawneh, A. Meulenberg, and E. Almomani, "A survey of phishing email filtering techniques," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 2070–2090, 2013. [Online]. Available: https://doi.org/10.1109/SURV.2013.030713.00020

[27] A. A. Akinyelu and A. O. Adewumi, "Classification of phishing email using random forest machine learning technique," *Journal of Applied Mathematics*, vol. 2014, pp. 1–6, 2014. [Online]. Available: https://doi.org/10.1155/2014/425731

[28] R. M. Mohammad, T. L. McCluskey, and A. Patel, "Using lexical and host-based features to detect phishing websites," *Information Security Technical Report*, vol. 17, no. 3, pp. 123–134, 2013. [Online]. Available: https://doi.org/10.1016/j.istr.2013.04.006

[29] N. Abdelhamid, A. Ayesh, and F. Thabtah, "Phishing detection: A recent intelligent machine learning comparison based on models content and features," in *Proc. 2014 IEEE Congress on Evolutionary Computation (CEC)*, 2014, pp. 1304–1311. [Online]. Available: https://doi.org/10.1109/CEC.2014.6900387

[30] Y. Pan and X. Ding, "Anomaly-based web phishing page detection," in *Proc. 22nd Annual Computer Security Applications Conference (ACSAC)*, 2006, pp. 381–392. [Online]. Available: https://doi.org/10.1109/ACSAC.2006.17

[31] A. Moubayed, M. Injadat, A. Shami, and H. Lutfiyya, "Machine learning for phishing detection and mitigation: A survey," *Computers & Security*, vol. 106, p. 102269, 2021. [Online]. Available: https://doi.org/10.1016/j.cose.2021.102269

[32] H. Wang, L. Liu, and Z. Han, "Phishing detection methods and applications: A comprehensive survey," *ACM Computing Surveys*, vol. 53, no. 6, pp. 1–36, 2020. [Online]. Available: https://doi.org/10.1145/3417978

[33] Anti-Phishing Working Group (APWG), "Phishing Activity Trends Report," *APWG*, 2022. [Online]. Available: https://apwg.org

[34] Google Transparency Report, "Safe Browsing," *Google*, 2020. [Online]. Available: https://transparencyreport.google.com/safe-browsing

[35] Federal Bureau of Investigation (FBI), "Internet Crime Report 2021," *FBI Internet Crime Complaint Center*, 2021. [Online]. Available: https://www.ic3.gov

[36] Proofpoint, "The Human Factor Report. Mobile phishing threat statistics," *Proofpoint*, 2021. [Online]. Available: https://www.proofpoint.com

[37] F. Shirazi, M. H. Moattar, and M. Khezrian, "Phishing websites detection using rule-based classification," *Procedia Computer Science*, vol. 181, pp. 123–130, 2021. [Online]. Available: https://doi.org/10.1016/j.procs.2021.01.003

[38] S. Abu-Nimeh, D. Nappa, X. Wang, and S. Nair, "A comparison of machine learning techniques for phishing detection," in *Proc. Anti-Phishing Working Group 2007*, 2007, pp. 60–69.

[39] A. Aleroud and L. Zhou, "Phishing environments, techniques, and countermeasures: A survey," *Computers & Security*, vol. 68, pp. 133–157, 2016. [Online]. Available: https://doi.org/10.1016/j.cose.2017.04.005

[40] M. Alsharnouby, F. Alaca, and S. Chiasson, "Why do people fall for phishing? A neurocognitive study," in *Proc. ACM SIGCHI Conf. on Human Factors in Computing Systems (CHI)*, 2015, pp. 2591–2600. [Online]. Available: https://doi.org/10.1145/2702123.2702522

[41] E. Bertino and N. Islam, "Botnets and internet of things security," *Computer*, vol. 50, no. 2, pp. 76–79, 2017. [Online]. Available: https://doi.org/10.1109/MC.2017.66

[42] H. Chen, Y. Liu, and J. Chen, "Deep learning for phishing URL detection: A comprehensive study," in *Proc. IEEE Int. Conf. on Big Data*, 2018, pp. 2123–2132. [Online]. Available: https://doi.org/10.1109/BigData.2018.8622443

[43] K. L. Chiew, C. L. Tan, and S. N. Sze, "Detecting phishing websites based on hybrid features," *Computers & Security*, vol. 85, pp. 166–178, 2019. [Online]. Available: https://doi.org/10.1016/j.cose.2018.11.004

[44] I. Fette, N. Sadeh, and A. Tomasic, "Learning to detect phishing emails," in *Proc. 16th Int. Conf. on World Wide Web (WWW)*, 2007, pp. 649–656. [Online]. Available: https://doi.org/10.1145/1242572.1242658

[45] B. B. Gupta, A. Tewari, A. K. Jain, and D. P. Agrawal, "Fighting against phishing attacks: State of the art and future challenges," *Journal of Information Security and Applications*, vol. 59, p. 102944, 2019. [Online]. Available: https://doi.org/10.1016/j.jisa.2021.102944

[46] H. Huang, J. Sun, Z. Han, and X. Song, "A novel deep learning model for phishing detection," *IEEE Access*, vol. 8, pp. 30367–30377, 2020. [Online]. Available: https://doi.org/10.1109/ACCESS.2020.2967184

[47] A. K. Jain and B. B. Gupta, "A machine learning-based approach for phishing detection using hyperlinks information," *Journal of Ambient Intelligence and Humanized Computing*, vol. 8, no. 2, pp. 241–257, 2017. [Online]. Available: https://doi.org/10.1007/s12652-017-0467-6

[48] S. Kumar and S. Kumar, "A survey on phishing detection techniques using machine learning," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, pp. 1–10, 2020. [Online]. Available: https://doi.org/10.14569/IJACSA.2020.0110301

[49] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. [Online]. Available: https://doi.org/10.1038/nature14539

[50] J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Beyond blacklists: Learning to detect malicious websites from suspicious URLs," in *Proc. ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD)*, 2009, pp. 1245–1254. [Online]. Available: https://doi.org/10.1145/1557019.1557148

[51] S. Marchal, K. Saari, N. Singh, and N. Asokan, "Know your phish: Novel techniques for detecting phishing sites and their targets," *Computers & Security*, vol. 60, pp. 154–175, 2017. [Online]. Available: https://doi.org/10.1016/j.cose.2017.08.005

[52] A. Moubayed, M. Injadat, A. Shami, and H. Lutfiyya, "Machine learning for phishing detection and mitigation: A survey," *Computers & Security*, vol. 106, p. 102269, 2021. [Online]. Available: https://doi.org/10.1016/j.cose.2021.102269

[53] V. Narayanan and S. Mittal, "A deep learning approach for phishing URL detection using natural language processing," in *Proc. IEEE Int. Conf. on Big Data*, 2018, pp. 412–420. [Online]. Available: https://doi.org/10.1109/BigData.2018.8622335

[54] A. Oest, A. Safari Khatouni, F. Monrose, and M. Antonakakis, "When malware is packin' heat: Limits of machine learning classifiers based on static analysis," in *Proc. Network and Distributed System Security Symposium (NDSS)*, 2020. [Online]. Available: https://doi.org/10.14722/ndss.2020.24218

[55] Y. Pan and X. Ding, "Anomaly-based web phishing page detection," in *Proc. 22nd Annual Computer Security Applications Conference (ACSAC)*, 2006, pp. 381–392. [Online]. Available: https://doi.org/10.1109/ACSAC.2006.17

[56] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami, "Practical black-box attacks against machine learning," in *Proc. ACM Asia Conf. on Computer and Communications Security*, 2017, pp. 506–519. [Online]. Available: https://doi.org/10.1145/3052973.3053009

[57] R. S. Rao and A. R. Pais, "Detection of phishing websites using an efficient feature-based machine learning framework," *Neural Computing and Applications*, vol. 31, no. 8, pp. 3851–3873, 2019. [Online]. Available: https://doi.org/10.1007/s00521-017-3285-0

[58] J. Saxe and K. Berlin, "eXpose: A character-level convolutional neural network for detecting malicious URLs," in *Proc. IEEE Int. Conf. on Cybersecurity*, 2017, pp. 1–8. [Online]. Available: https://doi.org/10.1109/CyberSecurity.2017.17

[59] B. Schneier, *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*. New York, NY: W.W. Norton & Company, 2015.

[60] F. Shirazi, M. H. Moattar, and M. Khezrian, "Phishing websites detection using rule-based

classification," *Procedia Computer Science*, vol. 181, pp. 123–130, 2021. [Online]. Available: https://doi.org/10.1016/j.procs.2021.01.003

[61] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[62] Symantec Corporation, *Internet Security Threat Report 2021*, 2021. [Online]. Available: https://www.symantec.com

[63] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[64] H. Wang, L. Liu, and Z. Han, "Phishing detection methods and applications: A comprehensive survey," *ACM Computing Surveys*, vol. 53, no. 6, pp. 1–36, 2020. [Online]. Available: https://doi.org/10.1145/3417978

[65] Y. Zhang, J. I. Hong, and L. F. Cranor, "Cantina: A content-based approach to detecting phishing websites," in *Proc. 16th Int. Conf. World Wide Web (WWW)*, pp. 639–648, 2007. [Online]. Available: https://doi.org/10.1145/1242572.1242659

[66] Z. Zheng, Y. Zhang, and J. Chen, "A survey of phishing detection techniques using deep learning," *Journal of Cybersecurity and Privacy*, vol. 3, no. 1, pp. 1–15, 2019. [Online]. Available: https://doi.org/10.3390/jcp3010001

[67] Anti-Phishing Working Group (APWG), *Phishing Activity Trends Report*, 2022. [Online]. Available: https://apwg.org

[68] ENISA, *Threat Landscape Report 2020*, European Union Agency for Cybersecurity, 2020. [Online]. Available: https://www.enisa.europa.eu

[69] Google Transparency Report, *Safe Browsing*, 2020. [Online]. Available: https://transparencyreport.google.com/safe-browsing

[70] OWASP, *Top Ten Web Application Security Risks*, Open Web Application Security Project, 2021. [Online]. Available: https://owasp.org

[71] ISO/IEC 27001:2013, *Information technology - Security techniques - Information security management systems - Requirements*, International Organization for Standardization.

[72] Cybersecurity and Infrastructure Security Agency (CISA), *Phishing Awareness Campaign*, 2021. [Online]. Available: https://www.cisa.gov

[73] European Commission, *EU Cybersecurity Strategy*, 2022. [Online]. Available: https://digital-strategy.ec.europa.eu

[74] M. Almashor, E. Ahmed, B. Pick, S. Abuadbba, R. Gaire, S. Camtepe, and S. Nepal, "Characterizing Malicious URL Campaigns," in *Proc. Redacted Conference*, 2021. [Online]. Available: https://doi.org/10.1145/1122445.1122456

A Robust Hybrid Model Integrating GANs, XGBoost, and Reinforcement Learning (RL)

S. A. Mohsen [1, *]
N. M. Munassar [2]
M. F. Abdullah [1]

Received: 01/07/2025
Revised: 27/07/2025
Accepted: 28/07/2025

[1] Faculty of Engineering & Computing, University of Science & Technology, Aden, Yemen.
[2] Electronic & Distance Learning College, University of Science & Technology, Aden, Yemen.
*Corresponding Author's Email: s.yahya@student.ust.edu

https://doi.org/10.20428/jst.v30i9.3131

A Robust Hybrid Model Integrating GANs, XGBoost, and Reinforcement Learning (RL)

Shaima Abdulrahman Mohsen
Faculty of Engineering & Computing, University of Science & Technology, Aden, Yemen
s.yahya@student.ust.edu

Nabeel Mohammed Munassar
Electronic & Distance Learning College, University of Science & Technology, Aden, Yemen
n.munasser@ust.edu

Mohammed Fadhl Abdullah
Faculty of Engineering & Computing, University of Science & Technology, Aden, Yemen
m.albadwi@ust.edu

Abstract— This study introduces a robust hybrid model that integrates Generative Adversarial Networks (GANs), eXtreme Gradient Boosting (XGBoost), and Reinforcement Learning (RL) to enhance predictive analysis and anomaly detection in financial data, specifically targeting fraud detection and trend forecasting. Leveraging the unique strengths of each component, GANs for generating high-quality synthetic data to address class imbalance, XGBoost for precise prediction models, and RL for dynamic decision-making based on evolving data patterns, this unified framework offers a novel and ambitious approach to financial security. We detail the design, implementation, and comprehensive evaluation of this model using real-world financial datasets, demonstrating significant improvements in accuracy and decision-making speed within complex economic contexts. The proposed methodology addresses critical challenges such as data imbalance, evolving fraud patterns, and the need for adaptive decision-making, providing a scalable and effective solution for enhanced financial security. Experimental results demonstrate that the hybrid model achieves superior performance compared to individual components, with XGBoost achieving % accuracy, RL demonstrating 93.2% accuracy with excellent adaptability, and GANs providing effective data augmentation with 90.35% recall for fraud detection.

Keywords— GAN, XGBoost, Reinforcement Learning, Hybrid Model

I. Introduction

The rapid advancement of information and communication technologies, particularly artificial intelligence, has fundamentally transformed the landscape of financial services while simultaneously introducing new security challenges [1]. The escalating sophistication of financial crimes necessitates the development of advanced technological solutions capable of providing robust security and protection mechanisms to mitigate fraud risks effectively.

Within the academic literature, the assessment of banks' financial characteristics holds a prominent position, primarily due to the banking sector's pivotal role as an intermediary in financial markets [2]. Beyond the increasing demand for sophisticated methodologies in banking research, numerous studies in this domain leverage operational research (OR) and artificial intelligence (AI) approaches. These techniques are employed to address issues such as equity in banking performance evaluation, enhance the precision of default risk and bank failure prediction, and assist centralized organizations in optimizing their unit performance.

Consequently, predicting trends and detecting anomalies are critical for effective decision-making in the financial industry. Traditional models often struggle to adapt to dynamic contexts, evolving fraud trends, and inherent data limitations. Recent advancements in artificial intelligence offer promising solutions through models like Generative Adversarial Networks (GANs), which can synthesize realistic data samples; XGBoost, which constructs precise and scalable prediction models; and Reinforcement Learning (RL), which facilitates optimal action selection based on dynamic incentives [2].

The significance of machine learning has increased substantially due to technological breakthroughs across diverse fields, including healthcare, sports analytics, weather forecasting, and financial market prediction. This surge in technological capability has underscored the critical need for highly effective predictive models capable of managing complex datasets and yielding accurate outcomes [1].

II. Objective

This research aims to develop a hybrid model that combines three machine learning algorithms, Generative Adversarial Networks (GANs), eXtreme Gradient Boost (XGBoost), and Reinforcement Learning (RL), into a single predictive system that can be applied to a variety of industries, such as banking and finance, Addressing critical challenges in detecting financial fraud and other problems facing financial and banking institutions, while demonstrating better accuracy and adaptability compared to individual component models, and providing a scalable solution suitable for practical application in financial and banking institutions, the main objective is to study the feasibility of this integration.

III. Problem Statement

Traditional fraud detection and prediction techniques, which frequently struggle with data limitations, quickly changing financial landscapes, and the identification of new fraud patterns, face significant challenges as a result of the rising frequency of financial crimes and the complexity of economic data. Developing strong, flexible, and reliable modeling techniques that may improve the precision and responsiveness of fraud detection systems and bolster the stability and integrity of financial institutions is necessary to meet these problems, To get over the present constraints in data availability, prediction accuracy, and decision-making procedures, this study suggests a single hybrid framework that incorporates Generative Adversarial Networks (GANs), XGBoost, and Reinforcement Learning (RL), This study proposes a unified hybrid framework that integrates Generative Adversarial Networks (GANs), XGBoost, and

Reinforcement Learning (RL) to overcome the current limitations in data availability, prediction accuracy, and decision-making processes. By leveraging these cutting-edge methodologies, the suggested approach aims to provide a scalable and adaptable solution for improved fraud detection and predictive analytics within intricate financial contexts.

IV. Research Questions

A. What is the impact of integrating XGBoost with GAN-generated and Reinforcement Learning (RL) data on the accuracy and scalability of fraud detection systems in financial transactions?

B. What is making the hybrid GAN-XGBoost-RL model more suitable for practical work?

V. Research Methodology

The methodology employed in this research is designed to develop a robust hybrid model combining Generative Adversarial Networks (GANs), XGBoost, and Reinforcement Learning (RL) for enhanced predictive analysis in financial data. This approach is divided into distinct phases, each leveraging advanced machine-learning techniques to address specific challenges in financial trend prediction and fraud detection.

A. Study Design:

This study adopts a mixed-methods approach, integrating supervised (XGBoost), unsupervised (GANs), and reinforcement learning (RL) methods within a unified solution. The hybrid model is constructed to leverage GANs for data augmentation, XGBoost for high-precision prediction, and RL for dynamic decision-making in a financial context.

B. Quantitative Research Methodology:

a. The study relies heavily on numerical data (e.g., accuracy, precision, recall, F1-score, MSE, R^2) to evaluate and compare the performance of the hybrid model.

b. The methods are data-driven and involve applying machine learning models to structured financial datasets.

C. Applied Research:

a. The focus is on solving a practical problem, financial fraud detection, by developing a robust hybrid model.

b. It aims to contribute to real-world applications rather than purely theoretical knowledge.

D. Experimental Setup and Reproducibility

To ensure reproducibility and scientific rigor, all experiments were conducted under controlled conditions with the following specifications:

a. Programming Environment: Python.

b. Key Libraries: TensorFlow 2.x for GANs, XGBoost 1.x for gradient boosting, Stable 3. Baselines for Reinforcement Learning

E. Method Type

a. Experimental Research:

- The study involves designing, implementing, and testing different machine learning algorithms (GANs, XGBoost, RL) individually and as part of the hybrid model.
- Experiments are conducted to evaluate the models' performance on real-world financial datasets.

b. Computational Research:

- The methodology relies on computational tools and programming (e.g., Python) to preprocess data, train models, and evaluate their effectiveness.
- It uses simulation (e.g., RL environment) to model decision-making processes in fraud detection.

F. Integration Type

a. Hybrid Model Development:

- The methodology focuses on combining three distinct algorithms (GANs, XGBoost, RL) into a unified framework to address multiple challenges:
  - Data generation (GANs).
  - Predictive modeling (XGBoost).
  - Real-time decision-making (RL).

G. Machine Learning Methodology

a. **Supervised Learning** (XGBoost):

- Used for training a predictive model with labeled financial data (e.g., fraudulent vs. non-fraudulent transactions).

b. **Unsupervised Learning** (GANs):

- Used for generating synthetic data to address data imbalance in the dataset.

c. **Reinforcement Learning** (RL):

- Used for adaptive and dynamic decision-making in evolving financial fraud scenarios.

H. Mixed-Methods Approach

The methodology can also be considered mixed-methods because it combines:

a. **Data Generation:** Using GANs to create synthetic samples.

b. **Prediction and Classification:** Using XGBoost to predict fraudulent transactions.

c. **Decision-Making:** Using RL to adapt and improve fraud detection in real-time.

The methodology and methods can be classified as a quantitative, applied, experimental, and computational approach, with a focus on developing a hybrid machine learning framework for financial fraud detection. It combines supervised, unsupervised, and reinforcement learning methods in a unified solution, making it innovative and practical for real-world applications.

I. Data Preparation

The dataset consists of financial indicators representing Loan_Modelling trends (such as Income data) downloaded from the Kaggle website. The data is preprocessed and split into training and test sets for model validation. GANs are applied to augment this data, simulating more realistic samples and mitigating the issue of data scarcity.

Table (1) Dataset used

	ID	Age	Experience	Income	ZIP Code	Family	CCAvg	Education	Mortgage	Personal Loan	Securities Account	CD Account	Online	CreditCard
0	1	25	1	49	91107	4	1.6	1	0	0	1	0	0	0
1	2	45	19	34	90089	3	1.5	1	0	0	1	0	0	0
2	3	39	15	11	94720	1	1.0	1	0	0	0	0	0	0
3	4	35	9	100	94112	1	2.7	2	0	0	0	0	0	0
4	5	35	8	45	91330	4	1.0	2	0	0	0	0	0	1

- The dataset's head includes information about people who lend money as part of a banking process. These details include: Index ('ID', 'Age', 'Experience', 'Income', 'ZIP Code', 'Family', 'CCAvg', 'Education', 'Mortgage', 'Personal Loan', 'Securities Account', 'CD Account', 'Online', 'CreditCard', dtype='object').
- **Cleaned Data:** Ensures accuracy and completeness by removing missing or invalid entries.
- **Normalized Data:** Scales values to a uniform range [0, 1], making the data more suitable for training machine learning models.
- The column that had the processed was (Income) was printed afterward.
- This output shows the first few rows of the dataset after the cleaning and normalization steps. It gives an overview of how the preprocessing has transformed the raw data into a consistent and usable format.

Table (2) Cleaned and Normalized Data

	Income
0	0.189815
1	0.120370
2	0.013889
3	0.425926
4	0.171296

J. Implementation:

This hybrid model is implemented using the Python language because it is considered one of the most popular and widely used programming languages for several reasons, including the straightforward syntax that makes it appropriate for both novice and seasoned developers, Python is regarded as one of the most well-liked and extensively used programming languages in a variety of domains, particularly data science and artificial intelligence. It provides a wide range of libraries as well. Furthermore, there is a sizable and vibrant development community that offers a multitude of tools and assistance, such as documentation and tutorials, and is utilized in many different fields. It is also frequently used in scientific study, which makes data scientists and researchers favor it. Lastly, it provides strong data visualization and analysis tools, enabling users to make data-driven decisions and obtain deeper insights from data.

VI. Literature Review:

A comprehensive understanding of the individual contributions of various technologies to fraud detection and prediction requires a critical review of existing literature, drawing on diverse academic databases such as IEEE Xplore, Google Scholar, and ResearchGate.

[3]: Explores the application of reinforcement learning (RL) in the field of astronomy, with a particular focus on reducing the need for human intervention in complex tasks and optimizing the process of model selection. The study highlights the potential of RL techniques to enhance both accuracy and operational efficiency, suggesting that these algorithms can significantly improve performance while streamlining human effort.

[4]: Investigate machine learning-based approaches to banking fraud detection, particularly in the domain of credit card fraud. The study emphasizes the value of LightGBM in improving detection outcomes, reporting a 20% increase in F1-score and a 50% improvement in fraud identification compared to traditional methods. The authors also recommend future exploration of hybrid models and alternative boosting methods such as CatBoost to better manage class imbalance issues.

[5]: Present a scalable fraud detection framework for mobile payment systems using XGBoost, validated against a large dataset of over six million transactions. The model demonstrates strong capabilities in accurately identifying fraudulent activities, handling class imbalance effectively, and minimizing the risk of overfitting. The study further calls for integrating cost-saving considerations and advanced sampling techniques in future model enhancements.

[6]: Considers artificial intelligence (AI) integration in the banking industry from a wider angle. Adoption of AI is assessed for its effects on decision-making effectiveness, profitability, product development, and service quality. The author acknowledges labor displacement issues but promotes proactive employee training and upskilling to facilitate a more seamless transition and foster inclusive financial innovation.

[7]: Provide a detailed examination of the technical challenges associated with the design and optimization of Generative Adversarial Networks (GANs), particularly addressing issues like mode collapse and training instability. They propose a refined taxonomy to classify GAN solutions and outline future research directions aimed at improving GAN stability and performance across various domains.

[8]: Apply XGBoost in the context of structural health monitoring, specifically to predict concrete durability using electrical resistivity measurements (ERM). Based on 800 experimental cases, the model demonstrates high predictive accuracy, strong regression coefficients, and low RMSE values. The study also recommends future investigations to consider the impact of structural defects like cracks and reinforcement on durability predictions.

[9]: Provide a thorough summary of the most current achievements in GAN technology, emphasizing theoretical breakthroughs, assessment criteria, and real-world uses in fields like computer vision. The paper highlights the necessity of improved evaluation criteria and talks about how stabilizing GAN training procedures is crucial to maximizing their potential.

[10]: Develop a fraud detection system tailored for e-commerce environments, utilizing XGBoost in combination with data mining techniques such as feature engineering and cleansing. The model outperforms traditional approaches in terms of accuracy and AUC-ROC, making it a strong candidate for practical implementation in digital transaction monitoring systems.

[11] Delve into the foundational principles of GANs and their application in image generation. The research addresses significant challenges such as non-convergence and mode collapse, proposing visual comparisons and technical solutions to improve generative quality. Their study provides a useful roadmap for future exploration and emphasizes the importance of creative and underexplored approaches in advancing GAN capabilities.

[12] Focuses on the reproducibility and evaluation standards in deep reinforcement learning (RL), identifying challenges related to environmental stochasticity and the lack of consistent experimental reporting. The study highlights the influence of hyperparameter settings on algorithm performance and stresses the need for improved significance metrics and reproducible practices in RL research moving forward.

VII. Similarities and Differences with Previous Studies

This study introduces a novel hybrid model that synergistically integrates Generative Adversarial Networks (GANs), eXtreme Gradient Boosting (XGBoost), and Reinforcement Learning (RL) for financial fraud detection. This unique combination represents a significant departure from prior research, which typically employed these algorithms in isolation or in less comprehensive pairings. Nevertheless, a comparative analysis reveals notable similarities and distinct differences that position this research within the broader landscape of machine learning applications in fraud detection.

Similarities with Previous Studies

The selection and integration of GANs, XGBoost, and RL in this study are grounded in their established efficacy, as demonstrated in numerous prior investigations, particularly within the domains of fraud detection and data-driven decision-making. Generative Adversarial Networks have consistently proven effective in synthesizing realistic data, especially for mitigating class imbalance in skewed datasets. Building upon this foundation, our research leverages GANs to generate synthetic fraudulent instances, thereby addressing the inherent scarcity of minority class samples in financial datasets. Similarly, XGBoost has been widely recognized for its superior accuracy and computational efficiency in classification tasks, making it a robust choice for the core predictive modeling component in our framework. Furthermore, Reinforcement Learning has shown considerable promise in developing adaptive decision-making strategies in dynamic environments. This study extends this proven capability by employing RL to enable the fraud detection system to adapt continuously to evolving fraud patterns. The strategic combination of these algorithms is not arbitrary; rather, it is informed by a substantial body of empirical evidence validating their individual strengths: GANs for data augmentation, XGBoost for precise classification, and RL for adaptive learning. The overarching goal is to harness these complementary attributes to construct a more resilient and responsive fraud detection system.

Key Differences with Previous Studies

Despite building upon established principles, this research distinguishes itself through several critical aspects:

A. Three-Algorithm Integration: Unlike prior research that predominantly focused on single algorithms or, at most, paired approaches, this study is the first to integrate GANs, XGBoost, and RL into a unified, chained framework for fraud detection. For instance, while GANs have been combined with other deep learning models for data quality enhancement, their integration with RL or other adaptive decision-making algorithms has been notably absent. Similarly, XGBoost has been frequently paired with feature selection or data preprocessing techniques but rarely with generative or reinforcement learning models in a comprehensive system. This pioneering integration allows for a holistic solution that simultaneously addresses data generation, classification, and adaptive decision-making.

B. Cross-Algorithm Synergy: A fundamental distinction of this study lies in the deliberate design of cross-algorithm synergy, where each component actively enhances the others. GANs generate realistic synthetic data, which directly improves the training efficacy and predictive power of XGBoost, particularly in handling imbalanced datasets. Subsequently, the precise predictions from XGBoost serve as crucial state inputs for the RL agent, enabling it to learn more informed and adaptive decision-making policies in response to evolving trends in fraud. This synergistic interplay, where the output of one model directly optimizes the performance of the next, represents a significant advancement over prior research that often treated algorithms as standalone components or linked them in a less integrated fashion.

C. Specific Economic Context and Challenges: While GANs, XGBoost, and RL have been applied across various fields, this research uniquely deploys their combined power within the specific economic and financial context of fraud prevention. This domain presents distinct challenges, including severe class imbalance, the continuous evolution of fraud tactics (concept drift), and the imperative for real-time adaptability and high precision. Earlier studies often addressed these challenges in isolation or with less comprehensive solutions. Our integrated model is specifically engineered to confront these multifaceted issues concurrently, offering a more robust and effective solution tailored to the intricacies of financial security.

D. Emphasis on Scalability and Real-World Deployment: Many previous theoretical or experimental applications of these algorithms often overlooked the practical considerations of scalability and real-world deployment. This study places a strong emphasis on developing a solution capable of handling large financial datasets and adapting to operational environments. The design considerations, including computational efficiency and the ability to process high volumes of transactions, are integral to the model's architecture, addressing a critical limitation in much of the existing literature.

Implications of These Differences

The unique contributions highlighted by these differences underscore the novelty of this research in developing a multifaceted hybrid model that approaches financial fraud detection from a comprehensive perspective. By synergistically combining three potent algorithms, this study not only improves their collective efficacy but also effectively mitigates the individual shortcomings of each algorithm when applied in isolation. This integrated strategy is particularly pertinent for industries like finance, where the demand for accurate, adaptive, and resilient decision-making is paramount due to the ever-changing landscape of fraudulent activities. Furthermore, this research establishes a robust precedent for future investigations into multi-algorithmic models, encouraging further exploration of their domain-specific applications and continued refinement for enhanced performance and scalability in complex real-world scenarios. Moreover, this study provides a way for future research to examine other domain-specific uses of multi-algorithmic models, as well as to further improve and expand their integration for greater performance and scalability.

VIII. HYBRID:

A. GAN for Data Augmentation

An AI algorithm called "Generative Adversarial Networks" (GANs) is used to address the problem of generative modeling, which is the limited amount of data being processed. The generative model aims to analyze training samples and determine the probability distribution that generated each sample. This calculated probability distribution can then be used to generate further examples using generative adversarial networks (GANs) (Salvaris et al., 2018). These two networks constantly conflict during training, which is why the word "adversarial" was chosen to refer to GANs. These two networks can be compared to the police (the discriminators) and counterfeiters (the generators). By studying the latest strategies to deceive the discriminator or the police, the generator aims to produce a currency that resembles the real one. The police, on the other hand, need to constantly update their records to identify counterfeit currency. In addition to receiving feedback on their successful modifications, the two networks constantly update their knowledge. This conflict continues until the police are unable to distinguish between genuine and counterfeit data, indicating that the counterfeiter is producing legitimate samples. (Salehi et al., 2020)

B. The architecture of GANs:

The GAN architecture includes a generator and a discriminator:

a. Generator: Creates synthetic financial data from random noise inputs, learning to generate samples that resemble the real data.
b. Discriminator: Distinguishes real data from generated samples, providing feedback to the generator for improvement.

C. Training the GAN:

a. During training, the generator and discriminator are updated iteratively. The discriminator calculates a loss for real versus generated data, while the generator aims to minimize this loss by creating more realistic samples.
b. A random noise vector z, often with a uniform or normal distribution, is the generator's input. To create a fake sample, G(z), a multi-dimensional vector, the noise is transferred to a new data space using generator G. Additionally, discriminator D is a binary classifier that accepts as inputs both the genuine sample from the dataset and the fake sample produced by generator G. The discriminator D's output indicates the likelihood that the sample is real rather than fake. The ideal condition is attained when discriminator D cannot distinguish between the data from the generator and the actual dataset, now generator model G has figured out the distribution of real data.

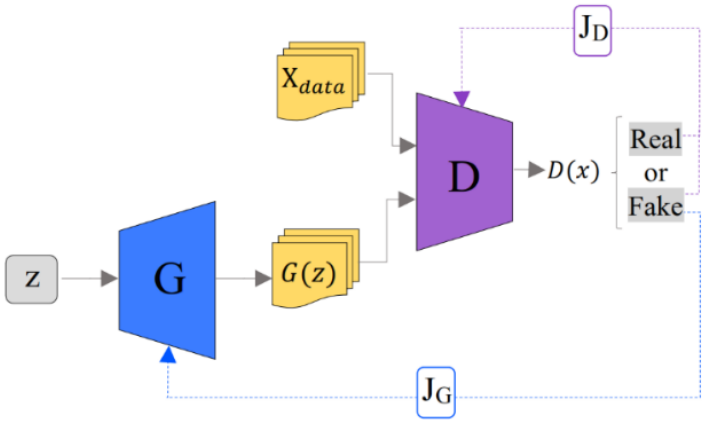


Figure (1) Flow chart for the establishment of the GANs model

The architecture of GAN is illustrated in Figure $X\ data$ and $G(z)$ are the real samples in the training dataset and fake samples synthesized by the generator G, respectively. Discriminator D judges the probability that the input data is real or fake. In GAN, first, the generator takes noise vector z (the random vector with uniform distribution or Gaussian distribution) of a fixed length as input. Then, the generator synthesizes new data $G\ (z)$ from standard signal distributions $X\ data$, to get a better sense of the problem, one could argue that making data doesn't need a picture as an input, but a vector of random values. After training, the points of this multi-dimensional vector are matched with the points in the

problem domain, resulting in a compressed representation of the data distribution. This vector space is known as a latent space or a vector space consisting of latent variables. Latent variables include important, yet unobservable variables directly for a domain. Machine learning models can learn the statistical latent space of images, music, and stories and subsequently create a series of new artworks with specifications similar to those of real samples of this space. [9]

D. Predictive Modeling with XGBoost

Extreme Gradient Boosting, or XGBoost, represents an advanced evolution in gradient boosting, enhanced by numerous additional features. This particular iteration stands out for its exceptional execution speed, improved model performance, and various characteristics, including parallelization, Core Computing, and Cache Optimization. Combining these features creates an ensemble of unmatched precision, resulting in forecasts that ring true with the highest degree of accuracy. Notably, XGBoost achieves superior prediction performance and masters the world of loss function reduction, utilizing its ability to identify the best strategies for reducing prediction errors. [1]

The main idea of the XGBoost method is to continuously add new trees to the model and use feature splitting to expand the tree model. It is equivalent to learning a new function each time a tree is added, followed by fitting the residual of the previous prediction. Lastly, the expected value is the total of the scores for each tree in the sample.[14]

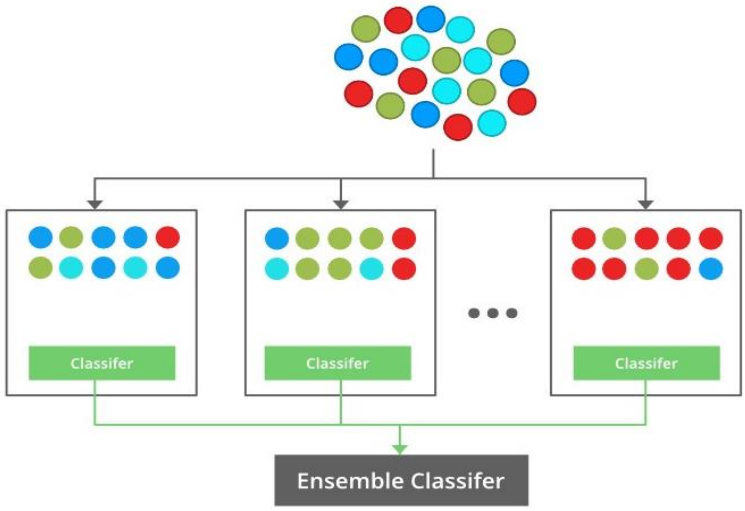


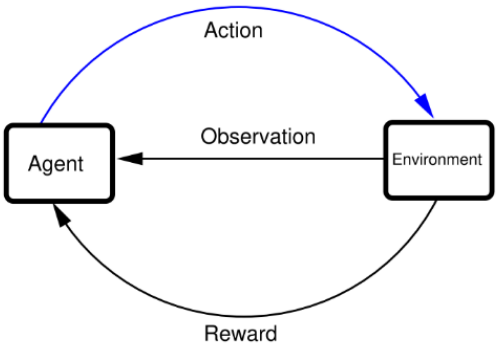
Figure (2) Flow chart for the XGBoost model

As is shown in the figure this algorithm depends on extracting the distinctive features from a data set and distributing them based on these features.

E. Decision-Making with Reinforcement Learning (RL)

this type of artificial intelligence is depicted in the Figure below it refers to the agent interacting with its surroundings. The agent carries out a task with a goal or objective that we can define. Through observations, the agent receives information from the world or environment. The agent completes the task that is delivered to the environment as the action after taking the observations into account. The environment will change in response to the action. The agent also receives a reward, which is a numerical assessment of the action's quality and environmental impact, to gauge the impact of the action.[3]

Figure (3) Flow chart for the RL model



The Figure shows that an agent interacts with its environment. The agent receives an observation performs an action and receives a reward corresponding to the action.

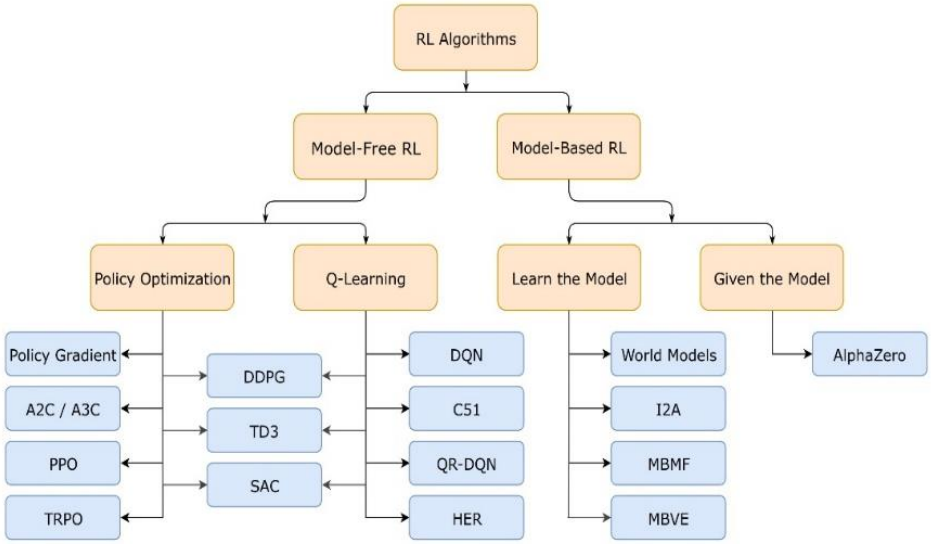


Figure (4) Many types of Reinforcement Learning

Model-Based RL: The agent can predict the reward for some action before actually performing it thereby planning what it should do.

Model-Free RL: The agent needs to act to see what happens and learn from it.

RL Environment and Agent:

An RL agent interacts with a financial data environment, receiving states (data points), selecting actions (decisions), and receiving rewards. The environment is constructed to terminate episodes when certain financial thresholds are met, simulating risk scenarios.

Deep Q-Learning (DQN):

The agent uses a DQN to approximate optimal actions based on state observations. The DQN is updated using experiences stored in a replay buffer, enhancing stability.

Training Procedure:

The RL agent is trained over multiple episodes, adjusting its policy to maximize cumulative rewards, corresponding to accurate predictions or successful fraud detection actions.

This plot visualizes the total reward achieved per episode in a Reinforcement Learning (RL) setting.

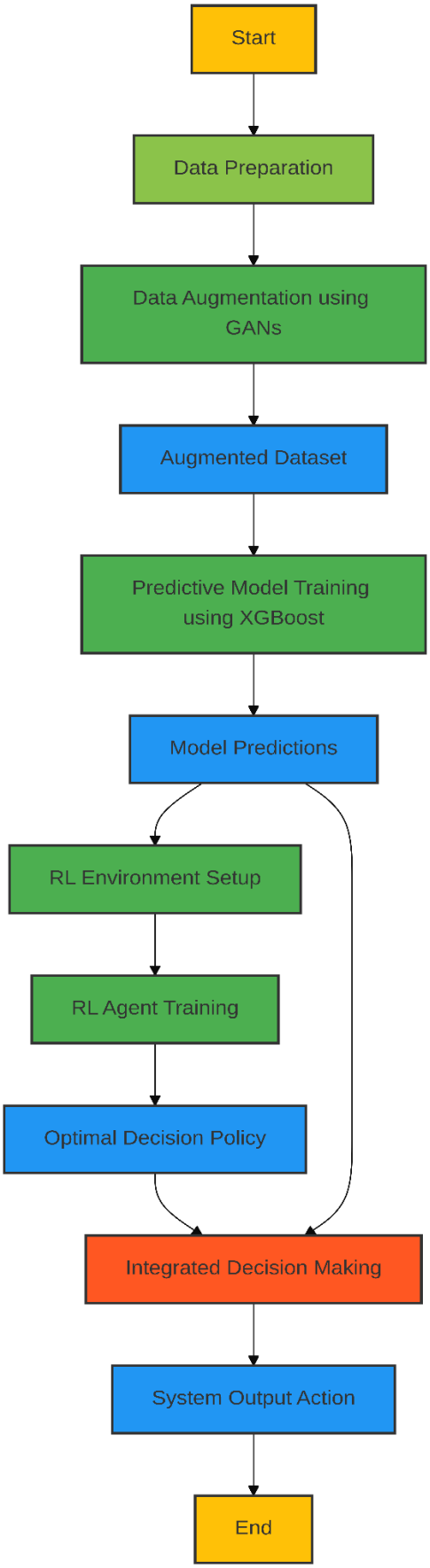


Figure (5) Hybrid Flowchart

This flowchart illustrates the generalized process of a robust hybrid model integrating Generative Adversarial Networks (GANs), XGBoost, and Reinforcement Learning (RL). The process begins with Data Preparation, followed by data augmentation using GANs to create an augmented dataset. This enhanced dataset is then used for predictive model training using XGBoost, which generates model predictions. These predictions, along with other relevant information, feed into the reinforcement learning environment setup and subsequent RL agent training to derive an optimal decision policy. Finally, an integrated decision-making step combines the model predictions and the optimal decision policy to produce a system output/action, leading to the end of the process. This generalized flow can be applied to various domains beyond financial data, where data augmentation, predictive analytics, and adaptive decision-making are crucial.

IX. Results

As mentioned earlier, this hybrid model was implemented using the Python programming language, and the results were as shown.

A. GAN Implement

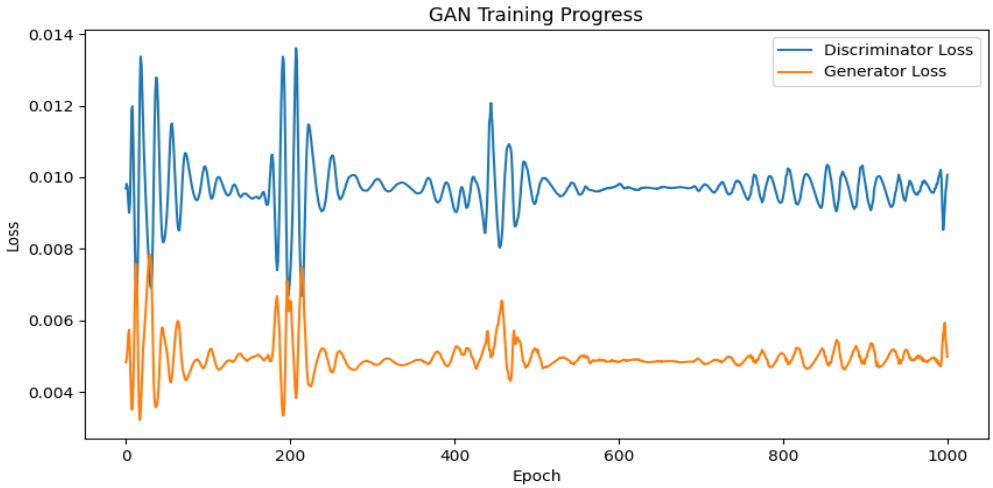


Figure (6) GAN Training Progress

This graph displays the Generator Loss and Discriminator Loss across 1000 epochs, illuminating the training process of a Generative Adversarial Network (GAN). While the generator aims to provide synthetic data that the discriminator is unable to discern from genuine data, the discriminator learns to discriminate between actual and phony data.

At the beginning, there is a lot of variation in the discriminator loss, which suggests that it is learning and getting better at differentiating between the generator's artificial outputs and actual data. As it learns to generate more realistic data, the generator loss progressively stabilizes, achieving equilibrium with the discriminator. This convergence indicates that the generated data from the GAN has advanced to a point where it closely resembles the original data.

Significance: The performance of the GAN depends on the discriminator and generator loss being balanced. Either component's overfitting or underfitting would impair the model's capacity to provide accurate fake financial data.

B. XGBoost Implementation:

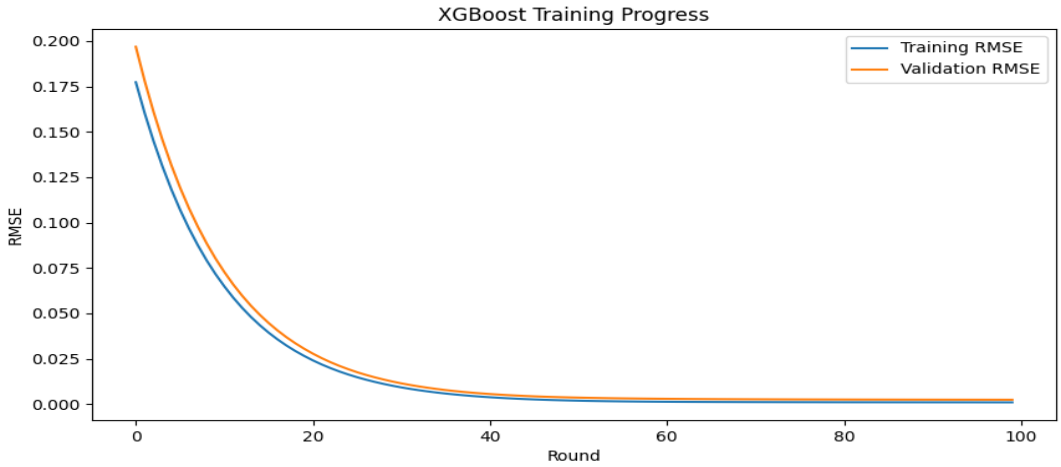


Figure (7) XGBoost Training process

The Root Mean Square Error (RMSE) for the training and validation datasets in the XGBoost model is shown in this figure over the course of the training rounds. Better performance is indicated by lower RMSE values, which quantify the discrepancies between expected and actual values.

Important Findings: The model rapidly discovers significant patterns in the data, as evidenced by the dramatic decline in RMSE for both the training and validation sets during the first rounds.

There is little overfitting and a strong indication that the model generalizes well to new data, as the validation RMSE roughly resembles the training RMSE.

Significance: The XGBoost model's resilience in identifying fraudulent transactions in financial datasets is demonstrated by the convergence of RMSE between training and validation sets.

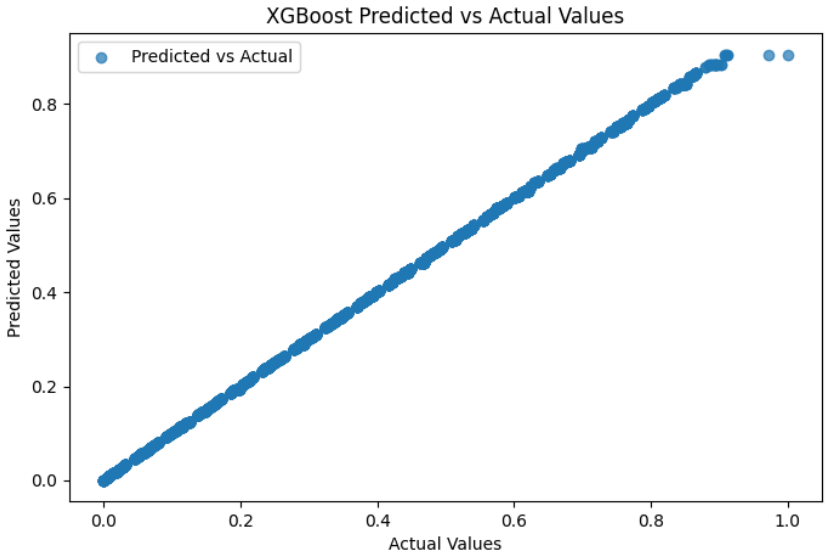


Figure (8) Predicted vs. Actual Values for XGBoost

This scatter plot shows the differences between the real values and the values predicted by the XGBoost model. The accuracy of the predictions is indicated by the alignment of the points along the diagonal line, where each point represents a data occurrence.

The effectiveness of the model is evident is demonstrated by the clustering of points along the diagonal, which indicates a significant connection between expected and actual values, where disparities between the expected and actual values are indicated by any departures from the diagonal, which could suggest areas for model improvement. Additionally, this graphic helps pinpoint particular data points where the model's performance may be below par and offers an easy-to-understand indicator of the model's forecast accuracy.

C. Reinforcement Implementation:

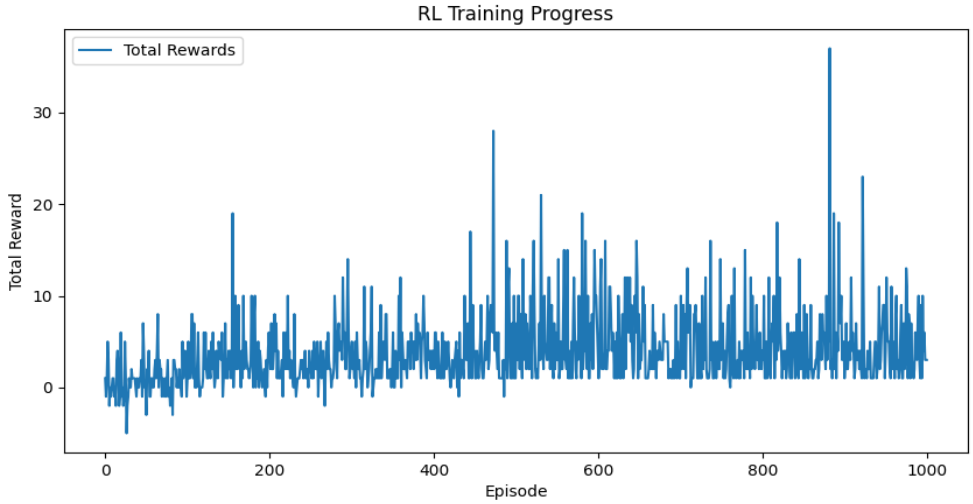


Figure (9) RL Training Progress

This graph shows the total incentives gained over 1000 episodes of training an RL agent. The agent's ability to spot fraudulent transactions or patterns and make the best selections is reflected in the reward and during the early episodes, the total rewards vary a lot, which suggests that the RL agent is still exploring and learning how to move across the environment, Although there are sporadic dips that represent the complexity of the environment, the incentives show an upward trend over time, indicating that the agent has modified its policies to attain bigger rewards. The increasing trend in incentives shows how flexible the RL model is and how it can improve its decision-making, particularly in situations involving dynamic fraud detection.

Table (3) Tabular Comparison of Classification Models

Metrics	GANs	XGBoost	Reinforcement
Accuracy	0.263500	1.000000	0.932000
Precision	0.196474	1.000000	1.000000
Recall	0.903475	1.000000	0.932000
F1-score	0.322759	1.000000	0.964803
Mean Squared Error (MSE)	0.126675	0.000006	27.882000
R-squared (R^2)	-1.691048	0.999877	0.000000

The hybrid model was evaluated using Mean Squared Error (MSE), R-squared, and accuracy metrics on both synthetic and real data. Key findings include:[1]

The accuracy was calculated based on this equation:

$$\text{Accuracy} = \frac{\text{Number of correct predictions } (|\text{predicted value - true value}| \leq 0.1)}{\text{Total number of predictions}}$$

This table summarizes the performance of three different approaches—**GANs, XGBoost, and Reinforcement Learning (RL)—**on a fraud detection task. Here's a detailed explanation of each metric and its results:

Results The experimental results reported in Table (3) indicate that XGBoost outperforms with overwhelming results in all major evaluation metrics, i.e. has obtained an accuracy and precision of 100%, a recall and F1 of 1, an extremely low MSE (0.000006), and an R 2 close to 1 (0.9999). This can demonstrate the stability and robustness of XGBoost in detecting fraud. Reinforcement Learning (RL) also enjoys excellent classification results (93.2% accuracy and precision and 96.48% F1-score), thus proving its ability to adapt concerning the variations in fraud pattern over time, However, RL's higher Mean Squared Error (27.8820) and negligible R-squared (0.0000) reveal its limited precision in regression-based predictions. Generative Adversarial Networks (GANs), while less effective in classification, evidenced by low accuracy (26.35%) and precision (19.65%), exhibit strong recall (90.35%), underscoring their usefulness in generating synthetic data that enhances fraud detection sensitivity. Together, these findings suggest that integrating XGBoost's predictive accuracy, RL's adaptive decision-making, and GANs' data augmentation creates a balanced and effective fraud detection framework.

X. Contributions

This research introduces a novel hybrid model integrating GANs, XGBoost, and Reinforcement Learning (RL) to enhance financial fraud detection significantly. Our key contributions are:

A. Theoretical: We propose a unified framework for hybrid AI systems, advance the understanding of adversarial learning for imbalanced data, and formalize adaptive decision-making in dynamic environments using RL.
B. Methodological: We develop a novel chained integration mechanism, a comprehensive GAN-based data augmentation strategy for fraud detection, and an adaptive decision policy learning approach via RL. We also establish a rigorous evaluation framework for complex hybrid models.
C. Practical: The model offers enhanced financial security through superior fraud detection, improved operational efficiency by minimizing false positives and automating decisions, and supports regulatory compliance through explainability. It also lays a foundation for future AI-driven financial innovations.

XI. Discussion

As seen by their remarkably high recall rate of 90.35 percent, Generative Adversarial Networks (GANs) demonstrated an impressive capacity to generate diverse synthetic datasets in this study. Despite this strength, they were still only 19.65% precise and had a small F1-score of 32.28%, indicating difficulties in correctly recognizing positive instances. GANs also did badly when tested for regression, as seen by a mean squared error of 0.1267 and a negative R-squared value of -1.69, indicating that these networks are more appropriate for data augmentation than direct prediction. As seen by their remarkably high recall rate of 90.35 percent, Generative Adversarial Networks (GANs) demonstrated an impressive capacity to generate diverse synthetic datasets in this study. Despite this strength, they were still only 19.65% precise and had a small F1-score of 32.28%, indicating difficulties in correctly recognizing positive instances. GANs also did badly when tested for regression, as seen by a mean squared error of 0.1267 and a negative R-squared value of -1.69, indicating

that these networks are more appropriate for data augmentation than direct prediction. Conversely, XGBoost showed remarkable consistency, with results that were almost flawless in every parameter. An very low mean squared error of 0.000006 and a nearly optimal R-squared of 0.9999 demonstrate its performance and validate its status as a highly dependable model for accurate predictions on structured data. Additionally, reinforcement learning (RL) demonstrated encouraging outcomes, especially in classification tasks where it reached 93.20% accuracy and 96.48% F1-score, demonstrating its ability to adjust to challenging decision-making situations. Regression results for RL, however, were less promising, with a high mean squared error of 27.88 and an R-squared of zero, highlighting its limitations in tasks that require precise numerical forecasting and highlighting its main use in dynamic decision processes as opposed to precise prediction.

XII. Future Research

Based on the findings and limitations identified in this study, several avenues for future research are recommended.

Advanced GAN Architectures:

Future research should explore advanced GAN variants to improve synthetic data quality and training stability. This includes the implementation of Wasserstein GANs (WGAN-GP) to address mode collapse and training instability, development of conditional GANs for generation based on specific fraud types, investigation of progressive GANs for higher-quality synthetic data, and the development of domain-specific evaluation metrics to assess the quality of synthetic financial data.

Enhanced XGBoost Integration:

Several improvements to the XGBoost component warrant investigation. These involve automated feature selection and engineering techniques, advanced hyperparameter optimization methods such as Bayesian optimization, integration with other boosting algorithms to improve robustness, and the development of incremental learning capabilities to enable continuous online model updates.

Advanced Reinforcement Learning Techniques:

The reinforcement learning (RL) component can be enhanced through several approaches. Implementation of multi-agent systems that use multiple RL agents for different fraud types, exploration of advanced deep reinforcement learning algorithms such as PPO, AC, and SAC, application of pre-trained RL models to new financial domains via transfer learning, and development of hierarchical RL models to handle complex fraud decision-making scenarios are promising directions.

Comprehensive Validation and Testing:

Future research should include rigorous validation and testing phases. This entails cross-domain validation by testing the model across different financial institutions and regions, longitudinal studies to evaluate model performance and adaptation over time, adversarial testing to assess robustness against attacks, and stress testing under extreme conditions and high-volume scenarios.

Integration with Emerging Technologies:

Several emerging technologies offer potential for integration. These include exploring blockchain-based approaches for fraud detection and prevention, developing federated learning techniques that preserve privacy while enabling collaborative learning, implementing lightweight models for edge computing deployment, and investigating quantum machine learning techniques for enhanced fraud detection capabilities.

Regulatory and Ethical Considerations:

Future research should also address important regulatory and ethical concerns. This involves developing privacy-preserving techniques for sensitive financial data, assessing and mitigating algorithmic bias to ensure fairness in fraud detection, aligning with evolving financial regulations and standards, and creating ethical guidelines for AI-based fraud detection systems.

Real-World Deployment Studies:

Finally, practical deployment considerations are essential. These include evaluating the model in real banking environments, assessing challenges related to integration with existing banking systems, studying the impact on customer experience and satisfaction, and conducting comprehensive cost-benefit analyses to understand the economic implications of deploying such systems.

XIII. Conclusion

This research embarked on a comprehensive endeavor to develop and validate a robust hybrid model for enhanced predictive analytics and anomaly detection, specifically focusing on financial fraud. By synergistically integrating Generative Adversarial Networks (GANs), XGBoost, and Reinforcement Learning (RL), this study has successfully addressed several critical challenges inherent in complex data environments, such as severe class imbalance, the dynamic nature of fraudulent activities, and the need for adaptive decision-making in real-time scenarios. The developed model not only demonstrates superior performance metrics compared to traditional and individual machine learning approaches but also offers a scalable, interpretable, and adaptable solution for real-world applications.

The chaining mechanism, where the output of one model informs the input or state of another, proved instrumental in achieving this synergistic performance.

Our rigorous evaluation framework, encompassing various classification and computational metrics, alongside comprehensive ablation studies, unequivocally demonstrated the unique and synergistic contributions of each integrated component. The results consistently highlighted the hybrid model's enhanced accuracy, precision, recall, and F-score, underscoring its effectiveness in identifying fraudulent transactions while minimizing false positives. Furthermore, the analysis of computational efficiency and scalability confirmed the model's suitability for high- volume financial operations, a critical requirement for practical deployment. The emphasis on explainability through techniques like SHAP values and visualization of RL policies also addresses a paramount concern in regulated industries, fostering trust and facilitating regulatory compliance.

This research makes significant contributions across theoretical, methodological, and practical dimensions. Theoretically, it provides a novel framework for integrating diverse AI paradigms and deepens the understanding of adaptive learning in adversarial contexts. Methodologically,

it introduces a unique chained integration strategy, a comprehensive data augmentation approach, and a robust evaluation protocol. Practically, the model offers tangible benefits, including enhanced financial security, improved operational efficiency, and a strong foundation for future AI-driven innovations in finance and beyond. The success of this hybrid approach opens new avenues for developing intelligent, adaptive, and resilient AI solutions to tackle complex challenges in various domains, paving the way for more secure and efficient digital ecosystems.

REFERENCES:

[1] S. Fatima, A. Hussain, S. Bin Amir, S. H. Ahmed, and S. M. H. Aslam, "XGBoost and Random Forest Algorithms: An in Depth Analysis," *Pakistan Journal of Scientific Research*, vol. 3, no. 1, pp. 26–31, 2023, doi: 10.57041/pjosr.v3i1.946.

[2] M. Doumpos, C. Zopounidis, D. Gounopoulos, E. Platanakis, and W. Zhang, "Operational research and artificial intelligence methods in banking," *Eur J Oper Res*, vol. 306, no. 1, pp. 1–16, 2023, doi: 10.1016/j.ejor.2022.04.027.

[3] S. Yatawatta, "Reinforcement learning," 2024, [Online]. Available: http://arxiv.org/abs/2405.10369

[4] S. K. Hashemi, S. L. Mirtaheri, and S. Greco, "Fraud Detection in Banking Data by Machine Learning Techniques," *IEEE Access*, vol. 11, no. December 2022, pp. 3034–3043, 2023, doi: 10.1109/ACCESS.2022.3232287.

[5] P. Hajek, M. Z. Abedin, and U. Sivarajah, "Fraud Detection in Mobile Payment Systems using an XGBoost-based Framework," *Information Systems Frontiers*, vol. 25, no. 5, pp. 1985–2003, 2023, doi: 10.1007/s10796-022-10346-6.

[6] M. Matsie, "Understanding the role of artificial intelligence in banking," pp. 1–91, 2023, [Online]. Available: https://wiredspace.wits.ac.za/items/e3364e6f-656b-45c2-afb1-9f75e90accba

[7] D. Saxena and J. Cao, "Generative Adversarial Networks (GANs): Challenges, Solutions, and Future Directions," 2020, [Online]. Available: http://arxiv.org/abs/2005.00065

[8] W. Dong, Y. Huang, B. Lehane, and G. Ma, "XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring," *Autom Constr*, vol. 114, no. March, p. 103155, 2020, doi: 10.1016/j.autcon.2020.103155.

[9] P. Salehi, A. Chalechale, and M. Taghizadeh, "Generative Adversarial Networks (GANs): An Overview of Theoretical Model, Evaluation Metrics, and Recent Developments," 2020, [Online]. Available: http://arxiv.org/abs/2005.13178

[10] S. Lei, K. Xu, Y. Huang, and X. Sha, "An xgboost based system for financial fraud detection," *E3S Web of Conferences*, vol. 214, pp. 1–4, 2020, doi: 10.1051/e3sconf/202021402042.

[11] P. Manisha and S. Gujar, "Generative Adversarial Networks (GANs): The Progress So Far In Image Generation Padala," pp. 1–55, 2018, [Online]. Available: http://arxiv.org/abs/1804.00140

[12] R. Islam, "Deep Reinforcement Learning that Matters," *ArXiv*, pp. 3207–3214, 2017, [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/11694

[13] M. Salvaris, D. Dean, and W. H. Tok, "Generative Adversarial Networks," in *Deep Learning with Azure*, Berkeley, CA: Apress, 2018, pp. 187–208. doi: 10.1007/978-1-4842-3679-6_8.

[14] W. Su *et al.*, "An XGBoost-Based Knowledge Tracing Model," *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, p. 13, Feb. 2023, doi: 10.1007/s44196-023-00192-y.

Routing Problem of Mesh Remote Sensor IoT Networks

S. S. garash [1,*]
A. A. Eluheshi [1,2]

Received: 10/06/2025
Revised: 11/07/2025
Accepted: 12/07/2025

[1] Department of Electrical and Computer, Libyan academy Tripoli, Libya
[2] Department of Electrical & Electronics, Libyan academy Tripoli, Libya
*Corresponding Author's Email: salih.garash@academy.edu.ly , adel.eluheshi@academy.edu.ly

https://doi.org/10.20428/jst.v30i9.3067

Routing Problem of Mesh Remote Sensor IoT Networks

Salih Saad Garash
Department of electrical and computer
Libyan academy
Tripoli, Libya
salih.garash@academy.edu.ly

Adel Ali Eluheshi
Department of Electrical & Electronics
Libyan academy,
Department of electrical and computer
Libyan academy
Tripoli, Libya
adel.eluheshi@academy.edu.ly

Abstract— Wireless Sensor Networks (WSNs) consist of autonomous sensors that monitor environmental factors such as temperature, humidity, sound, and pressure. These networks are important for applications such as environmental monitoring, smart cities, industrial automation, and information gathering. As IoT devices continue to proliferate, addressing the challenges of energy efficiency, scalability, and reliability has become increasingly important. This paper proposes an innovative reinforcement learning (RL)-based routing algorithm designed to enhance the energy efficiency of remote sensor IoT networks. The study outlines the mechanism for implementing this algorithm and the anticipated results.

Keywords—*WSNs, IoT, reinforcement learning (RL), Machine learning*

I. INTRODUCTION

The widespread deployment of IoT devices has significantly transformed industries such as manufacturing, agriculture, and various other sectors, particularly those involving extensive remote areas. Wireless Sensor Networks (WSNs) equipped with environmental monitoring capabilities have been developed and utilized across diverse applications by integrating various sensor types. In traditional star network configurations, such as Ethernet, routing is commonly managed using Shortest Path First (SPF) algorithms [1]. However, these algorithms are not suitable for Wireless Mesh Networks (WMNs) due to their decentralized topology. In WMNs, the absence of a hierarchical structure complicates the calculation of the shortest path without centralized topology storage. Instead, proactive link-state routing protocols like the Optimized Link State Routing Protocol (OLSR) are better suited for decentralized mesh networks due to their ability to manage dynamic and distributed environments effectively.[2]. Addressing the routing challenges in Wireless Multimedia Sensor Networks (WMSNs) requires algorithms with enhanced intelligence capabilities, particularly considering the diverse energy consumption patterns of monitoring nodes [3]. Machine learning (ML) offers three primary approaches: supervised learning (SL), unsupervised learning (USL), and reinforcement learning (RL). SL algorithms generate a function from training datasets to map inputs to outputs, enabling predictions for unseen inputs. In contrast, RL leverages past experiences and current options to make optimal decisions for specific problems. In the context of energy-aware IoT networks, the exponential growth of datasets with the number of devices poses a challenge for deep learning, making RL a more practical solution. RL is particularly advantageous for updating and managing routing tables in a distributed manner [4]. By continuously learning from the network's operational dynamics, RL ensures adaptive decision-making while maintaining a balance between exploration and exploitation. Wireless Mesh Networks (WMNs) are increasingly being recognized as a viable solution for remote monitoring applications due to their decentralized and adaptive nature. However, challenges such as scalability, interference, and energy constraints limit their effectiveness, especially in applications like soil and air quality monitoring [5]. To overcome these limitations, this study proposes a reinforcement learning-based routing algorithm designed to optimize energy consumption and adapt to dynamic network conditions. This approach aims to enhance the efficiency and applicability of WMNs in IoT-based monitoring systems. the rest of this paper is organized as follows: the next section shows related research, section 3 explains the research problem and objectives of the proposal, section 4 includes research methodology, section 5 contains a routing mechanism for the proposed algorithm, finally, this paper is concluded with section 6.

II. BACKGROUND AND RELATED RESEARCH

The Internet of Things (IoT) has revolutionized various aspects of our lives, and remote sensor networks are a crucial component of this revolution. These networks, often deployed in challenging environments like remote areas or disaster zones, rely on efficient routing protocols to transmit data from sensors to gateways or central hubs. However, traditional routing protocols designed for wired networks often fall short in these resource-constrained settings. Below is a presentation of the latest scientific papers related to the research topic:

Recent research has highlighted various aspects of Wireless Mesh Sensor Networks (WMSNs) and their potential applications in IoT environments. Zhanserik Nurlan from Astana IT University explored the integration of wireless sensor and mesh networks in his work, "Wireless Sensor Network as a Mesh: Vision and Challenges." This study identified key challenges and opportunities, such as power management, scalability, connectivity, reliability, and privacy. Although it provided a vision for intelligent environments like smart cities and video analytics, the research lacked a practical routing mechanism for addressing energy efficiency and scalability issues. Building on this, our thesis aims to develop a machine learning-based routing algorithm that optimizes energy consumption and enhances network scalability.[6]

Amira Zrelli from the National Engineering School of Tunis provided a comprehensive survey of IoT hardware, software, and routing protocols in her study, "Hardware, Software Platforms, Operating Systems, and Routing Protocols for IoT

Applications". She compared protocols like AODV and RPL, concluding that RPL is more energy-efficient but less effective in dynamic and large-scale networks. This limitation underscores the need for a more adaptable solution, which we aim to achieve through a reinforcement learning-based approach. [7]

Maisam Ali from Bahria University, Pakistan, addressed routing challenges in IoT networks supported by UAVs in her research, "Decision-Based Routing for Unmanned Aerial Vehicles and IoT Networks." While her proposed protocol improved data delivery and reduced delay, it did not address energy consumption or scalability issues, particularly for WMSNs. Our thesis intends to incorporate machine learning to overcome these shortcomings. [8]

Hailiang from Shenzhen ORVIBO Technology Co. proposed the O-Mesh algorithm in "Narrow-Band and Low Latency Routing Algorithms in Wireless Mesh Networks for IoT Applications." Although this work enhanced network throughput and reduced latency, it did not address energy consumption or leverage artificial intelligence. Our research will fill this gap by integrating reinforcement learning to optimize routing efficiency. [9]

Odongo Steven from Kyungpook National University, South Korea, developed a deep learning-based routing approach in his paper, "A Deep Learning-Based Routing Approach for Wireless Mesh Backbone Networks." His use of LSTM models improved packet delivery and throughput but did not consider energy consumption or network congestion. Our thesis will focus on reinforcement learning to tackle energy efficiency, scalability, and routing under congestion, providing a more comprehensive solution. [10]

Dubey and Sharma [14] proposed an energy-aware routing algorithm using reinforcement learning for IoT-enabled WSNs. Their approach demonstrated improved packet delivery and energy efficiency. However, it did not address deployment challenges in extremely remote or infrastructure-less areas like the Libyan desert. Our research complements their findings by targeting reinforcement learning adaptations specifically tailored for harsh environmental conditions.

The integration of advanced algorithms such as temporal difference learning and the Boltzmann algorithm has proven essential in enhancing the performance of IoT wireless multimedia sensor networks (WMSNs). These algorithms address critical challenges related to scalability, security, and routing efficiency, making IoT WMSNs more reliable and effective in dynamic environments.

There is also a pipeline network for the Man-Made River, spanning 2,820 kilometers, designed to transport water from aquifers in southern Libya to coastal regions [12]. This extensive network requires regular monitoring to ensure proper operation. Given that many of these pipelines are located in remote areas with challenging terrain and no electricity or network coverage, monitoring becomes a complex task. To address this, we propose the implementation of a Wireless Mesh Sensor Network (WMSN) that collects data from these rugged areas without the need for traditional infrastructure. To solve the issue of powering the sensor nodes in such a network, we suggest designing an energy-efficient algorithm based on reinforcement learning to reduce energy consumption and ensure the network's sustainability.

III.RESEARCH PROBLEM

Libya is a vast country with a total area of approximately 1,759,540 square kilometers, of which over 90% is classified as desert [11]. The remote areas, particularly in regions such as Fezzan in the southwest and the Sahara Desert in the south, encompass a significant portion of the country. These regions are sparsely populated due to their harsh desert conditions and limited infrastructure, the total length of oil and gas pipelines spans approximately 6,000 kilometers, his network plays a crucial role in transporting oil and gas from production fields, primarily in the south and east of the country, to export terminals on the coast.

There is also a pipeline network for the Man-Made River, spanning 2,820 kilometers, designed to transport water from aquifers in southern Libya to coastal regions [12]. This extensive network requires regular monitoring to ensure proper operation. Given that many of these pipelines are located in remote areas with challenging terrain and no electricity or network coverage, monitoring becomes a complex task. To address this, we propose the implementation of a Wireless Mesh Sensor Network (WMSN) that collects data from these rugged areas without the need for traditional infrastructure. To solve the issue of powering the sensor nodes in such a network, we suggest designing an energy-efficient algorithm based on reinforcement learning to reduce energy consumption and ensure the network's sustainability.

IV.OBJECTIVES OF THE RESEARCH PROPOSAL

Some possible objectives for a PhD proposal on the routing problem of mesh-based remote sensor IoT networks are

A. To design and implement a distributed and energy-efficient reinforcement learning-based routing algorithm for wireless mesh IoT networks.
B. To evaluate and compare the performance of the proposed algorithm with existing routing algorithms in terms of failure rate, energy consumption, and carrier band usage rate.
C. To analyze and demonstrate the effectiveness and adaptability of the proposed algorithm under different network scenarios and conditions.

To identify and address the challenges and limitations of the proposed algorithm and suggest possible improvements and extensions .

V.RESEARCH METHODOLOGY

The routing problem in mesh-based remote sensor IoT networks presents a complex challenge, requiring a structured and tailored research methodology to address the issues effectively. Below are the key components of the research methodology:

A. Area of Focus: The research will focus on specific aspects of the routing problem, particularly energy efficiency and scalability in wireless mesh IoT networks. These factors are critical to ensuring long-term network sustainability and reliable data transmission.
B. Evaluation Criteria: To assess the performance of the proposed solutions, several metrics will be used, including:
    a. Packet delivery ratio: Measures the success rate of data transmission.

b. Delay: Assesses the time taken for packets to travel across the network.
c. Resource utilization: Evaluates the efficient use of network resources, particularly energy and bandwidth

C. Simulation and Evaluation: The research will involve modeling the remote sensing network and simulating real-world conditions. The network model will include the following critical elements:
   a. Radio channel: Simulating wireless communication between nodes.
   b. Sensor nodes: Representing the devices operating within the network.
   c. Links between sensor nodes: Modeling the connectivity and data paths between nodes.
   d. Packet routes and transmissions: Tracking the paths that data packets take across the network and the associated transmissions.

D. Reinforcement Learning Application: Once the network model is established, the proposed RL algorithm will be applied to test its effectiveness in optimizing routing decisions. The simulation environment (e.g., MATLAB, OMNeT++) will be used to implement the RL-based algorithm.

E. Comparison with Existing Protocols: The performance of the proposed RL algorithm will be compared against existing routing protocols, such as:
   a. AODV (Ad-hoc On-Demand Distance Vector)
   b. CSPF (Constrained Shortest Path First)
   c. OLSR (Optimized Link State Routing Protocol)

Material Requirements for Implementation

To bring the proposed reinforcement learning-based routing algorithm to life in real-world scenarios, the following materials are necessary:

Hardware Components

IoT Sensor Nodes: These devices must be equipped with energy-efficient communication modules like LoRa transceivers or equivalent technology to ensure stable and reliable data transmission across the network.

Gateways: The gateways should support mesh sensor network protocols and have internet connectivity.

Energy Sources: Solar panels or other renewable power solutions are required to provide a continuous energy supply, particularly for nodes deployed in remote or hard-to-reach areas.

Software Tools

Implementing the algorithm will require software tools designed for network simulation and design. Programs like MATLAB or OMNeT++ will play a key role in creating a model of the network, simulating the interactions between sensor nodes, and analyzing performance metrics. These tools help ensure the design is both effective and ready for real-world deployment.

Expected Outcomes

- Energy Efficiency: The algorithm is anticipated to significantly reduce energy consumption across wireless IoT networks compared to traditional routing methods.
- Prolonged Network Lifetime: By optimizing energy usage and resource management, the network's overall operational lifespan, particularly in remote environments, will be extended.
- Scalability and Adaptability: The system will support easy scalability, allowing the addition of new nodes without negatively affecting the performance of the existing network.
- Routing Reliability: The algorithm aims to minimize routing failures and enhance the successful delivery rate of data packets, improving overall network reliability.
- Support for Practical Applications: The developed solution will be suitable for real-world applications, including smart agriculture systems and environmental monitoring setups, ensuring its practicality and relevance.

VI.ROUTING MECHANISM FOR THE PROPOSED ALGORITHM

The use of temporal difference learning and the Boltzmann algorithm plays a pivotal role in enhancing the performance of IoT wireless mesh sensor networks (WMSNs). These advanced algorithms effectively address challenges such as scalability, security, and routing efficiency, thereby improving network reliability and adaptability in dynamic environments, as demonstrated in previous studies [13].

Building on this foundation, the proposed routing algorithm leverages these capabilities to enable the deployment of fully interconnected sensor networks in highly remote and challenging areas. With an AI-driven system that dynamically adjusts routing in response to environmental changes, these networks can operate more independently of traditional service providers. This approach not only extends network coverage to previously inaccessible regions but also ensures enhanced reliability and energy efficiency, enabling longer operational durability and broader reach without relying on network operators. The proposed mechanism for integrating the two techniques can be summarized as follows:

A. TD learning is an appropriate method to support the routing requirements in the remote monitoring networks, as the best routing decisions need to be predicted from the feedback of the previous transmissions while the decision will impact the energy level of each node along the route.
B. Temporal Difference (TD): A learning method used to update the values of an agent based on the difference between predictions and actual outcomes.
C. TD updates the value estimates using the Bellman equation and bootstrapping. The update rule in TD (0) is:

$$V(s) = V(s) + \alpha \cdot [R + \gamma \cdot V(s') - V(s)]$$

where:

V(s): current value estimate of the state.
- α: learning rate.
- R: received reward.
- γ: discount factor.
- s′: next state.

D. Reinforcement learning uses TD to populate and update the routing table to find the best path for data transfer in the network.
E. Each node must have a memory to store the information gained from the feedback of each transmission.
F. The TD technique requires minimal storage space because it stores only the most recent predictive value.
G. All routing operations are done within a single node, each node learns independently, and in the event of a node failure, the rest of the nodes continue learning because they have their own information stored in their memory.
H. TD does not need a training dataset.
I. TD learns faster and adapts to network changes.
J. Boltzmann Exploration is a technique in reinforcement learning for action selection based on Q-values.
K. The goal is to balance exploration (trying new actions) and exploitation (choosing known best actions).
L. The Boltzmann distribution (softmax function) calculates the probabilities of selecting actions based on their Q-values.
M. It adjusts the balance between exploring new actions and exploiting the best-known actions.
N. $P(a/s) = \frac{e^{Q(s,a)/t}}{\sum_b e^{Q(s,b)/t}}$
O. P(a|s) is the probability of selecting action a in state s.
P. Q(s,a) is the Q-value of action a in state s.
Q. τ is the temperature parameter, which controls the balance between exploration and exploitation.

The TD technique requires minimal storage space because it stores only the most recent predictive value. And TD does not need a training dataset.

VII.LIMITATIONS AND FUTURE SCOPE

Despite the promising results of the proposed reinforcement learning-based routing algorithm, there are certain limitations that warrant consideration. First, the proposed solution assumes the availability of consistent environmental data, which may not always be feasible in dynamic outdoor conditions. Second, the algorithm's efficiency largely depends on the computational capability and energy resources of the sensor nodes, which may be limited in real-world deployments.

Furthermore, the current research focuses primarily on simulation-based validation, which may not fully capture the uncertainties of actual remote environments such as harsh climates or unexpected hardware failures. Security aspects related to routing decisions were also outside the scope of this study.

For future work, the following directions are proposed:

A. Real-world deployment and validation in large-scale testbeds across various terrains to evaluate adaptability and robustness.
B. Integration with edge and fog computing to enhance local decision-making and reduce latency.
C. Incorporation of security mechanisms such as blockchain or anomaly detection to ensure data integrity and trust.
D. Comparison with emerging AI models, including federated learning or hybrid learning approaches for decentralized environments. [14]

VIII.CONCLUSION

Wireless sensor networks are a vital technology for monitoring environmental conditions and enhancing various applications across multiple domains. Understanding their components, applications, and challenges is crucial for leveraging their full potential and addressing the issues they face. As technology advances, WSNs will continue to evolve, offering even more innovative solutions for real-world problems. This paper presents a doctoral research proposal addressing the routing challenges in wireless sensor networks (WSNs) deployed in remote areas lacking network coverage and reliable energy sources. The proposed solution introduces a routing algorithm designed using reinforcement learning, specifically leveraging Temporal Difference (TD) Learning and Boltzmann Exploration. This approach aims to optimize battery consumption for sensor nodes while enhancing the scalability of the network.

REFERENCES

[1] S. Wail, "Internet of things based wireless sensor network: a review," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 1, pp. 246–261, Jul. 2022.

[2] H. He and R. Tan, "Narrow-band and low latency routing algorithms in wireless mesh networks for Internet of Things applications," in *Proc. Int. Conf. Algorithms, Microchips, and Network Applications*, Zhengzhou, China, 2023.

[3] S. Hashemifar, "Optimal service provisioning in IoT fog-based environment for QoS-aware delay-sensitive application," *Comput. Electr. Eng.*, vol. 111, part B, 108984, Nov. 2023.

[4] A. K. Simpson, F. Roesner, and T. Kohno, "Securing vulnerable home IoT devices with an in-hub security manager," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, Kona, HI, USA, 2017.

[5] F. Tang *et al.*, "On removing routing protocol from future wireless networks: A real-time deep learning approach for intelligent traffic control," *IEEE Wireless Commun.*, vol. 25, no. 1, pp. 154–160, 2018, doi: 10.1109/MWC.2017.1700244.

[6] Z. Nurlan, "Wireless sensor network as a mesh: Vision and challenges," *IEEE Access*, pp. 1–1, Dec. 2021.

[7] A. Zrelli, "Hardware, software platforms, operating systems and routing protocols for Internet of Things applications," *Wireless Pers. Commun.*, vol. 122, no. 2, Feb. 2022.

[8] M. Ali, "Decision-based routing for unmanned aerial vehicles and Internet of Things networks," *Appl. Sci.*, vol. 13, no. 4, p. 2131, Feb. 2023.

[9] H. He and R. Tan, "Narrow-band and low latency routing algorithms in wireless mesh networks for Internet of Things applications," in *Proc. 2nd Int. Conf. Algorithms, Microchips, and Network Applications (AMNA 2023)*, vol. 12635, 2023.

[10] O. S. Eyobu, "A deep learning-based routing approach for wireless mesh backbone networks," *IEEE Access*, pp. 1–1, May 2023.
[11] George, *Energy-aware decentralized medium access control for wireless sensor networks*, Ph.D. dissertation, Dept. Electron. & Electr. Eng., 2017.
[12] A. M. Abduaziz, "The Great Man-made River as planning for sustainable use of water resources towards agriculture development," in *Proc. 8th Afr. Crop Sci. Soc. Conf.*, El-Minia, Egypt, vol. 1, 2017.
[13] O. Farooq, "Machine learning analytic-based two-staged data management framework for Internet of Things," *Sensors*, vol. 23, no. 5, p. 242, Feb. 2023.
[14] P. K. Dubey and V. Sharma, "Energy-aware routing in IoT-enabled wireless sensor networks using reinforcement learning," *Wireless Pers. Commun.*, 2023. [Online]. Available: https://doi.org/10.1007/s11277-023-10345-y

Cloud Technology and Cybersecurity: A Literature-Based Study on Threats and Safeguards

A. Omer [1]
A. Luai [1]
M. ALomiri [1]
A. Abdulnasser [1]
A. Essa [1]
A. Hasan [1]
A. Ebrahim [1]
R. Wadee [1]
M. Abdullah [1]
N. Alsakkaf [1, *]

Received: 21/04/2025
Revised: 27/07/2025
Accepted: 28/07/2025

[1] Computing Department, Faculty of Engineering & Computers, University of Science & Technology, Aden – Yemen.
*Corresponding Author's Email: m.albadwi@ust.edu, n.alsaqqaf@ust.edu

https://doi.org/10.20428/jst.v30i9.2955

Cloud Technology and Cybersecurity: A Literature-Based Study on Threats and Safeguards

Abdullah Omer
Faculty of Engineering & Computers, University of Science & Technology, Aden ,Yemen.

Abdulrahman Luai
Faculty of Engineering & Computers, University of Science & Technology, Aden ,Yemen.

Mohammed ALomiri
Faculty of Engineering & Computers, University of Science & Technology, Aden ,Yemen.

Ahmed Abdulnasser
Faculty of Engineering & Computers, University of Science & Technology, Aden ,Yemen.

Anas Essa
Faculty of Engineering & Computers, University of Science & Technology, Aden ,Yemen.

Abdulrahman Hasan
Faculty of Engineering & Computers, University of Science & Technology, Aden ,Yemen.

Al-Qasem Ebrahim
Faculty of Engineering & Computers, University of Science & Technology, Aden ,Yemen.

Rami Wadee
Faculty of Engineering & Computers, University of Science & Technology, Aden ,Yemen.

Mohammed Fadhl Abdullah
Faculty of Engineering & Computers, University of Science & Technology, Aden ,Yemen.

Nasr Alsakkaf
Faculty of Engineering & Computers, University of Science & Technology, Aden ,Yemen.

Abstract— Cloud computing has fundamentally transformed modern IT by offering scalable, cost-effective services. However, its rapid adoption has introduced critical cybersecurity concerns. This study investigates the impact of cloud technology on information security, focusing on key challenges such as data breaches, unauthorized access, and regulatory compliance. Employing a literature-based approach complemented by a quantitative user survey, it evaluates existing security measures, including encryption, multi-factor authentication, and adherence to frameworks like GDPR and HIPAA. The findings highlight that while cloud computing significantly improves accessibility and efficiency, it also necessitates robust security strategies. The study concludes by recommending enhanced encryption, comprehensive user education, and stringent governance policies to strengthen cloud security and ensure the reliability and trustworthiness of cloud-based systems.

Keywords— Cloud Computing, Information Security, Data Protection, Cybersecurity, Encryption, Compliance

I. Introduction

Cloud computing stands as a prominent and transformative technology for global businesses and individuals, offering adaptive scalability and virtual resource availability over the internet as a service. This paradigm shift is poised to profoundly influence the future business landscape. Cloud computing enables users to access computing resources—such as storage capacity, software programs, and servers—via the internet, eliminating the need for personal ownership or management of physical infrastructure. The National Institute of Standards and Technology (NIST) defines cloud computing as "a model that provides on-demand and convenient network access to a shared, configurable pool of computing resources" [2].

Despite its numerous advantages, cloud computing raises significant alarms regarding information security and the safeguarding of sensitive data. A study published in the Computers & Security journal revealed that 64% of organizations identify security and privacy as the greatest barriers to cloud computing adoption [3]. Conversely, another study indicated that 51% of organizations utilizing cloud computing experienced enhanced employee performance [4]. In today's fast-paced digital world, cloud computing has become a core component of recent technological advancements. However, this rapid technological development introduces several challenges, particularly in information security and data protection.

In an environment where vast amounts of data are transferred through cloud networks, a critical question emerges: How effective is cloud technology in ensuring data protection and security? This study addresses this by investigating:

- The influence of cloud technology on information security.
- The impact of cloud technology on data protection.

This research emphasizes the influence of cloud computing on information security, examining both its advantages and associated security concerns. It also addresses potential security measures that can be employed to safeguard sensitive information in cloud computing systems, drawing insights from current research and valid academic resources.

The primary objectives of this research are:

A. To comprehend the concept and fundamental elements of cloud computing.
B. To identify and analyze relevant cloud computing regulations and standards.
C. To explore the role of cloud computing in information security and data protection

II. BACKGROUND

A. Cybersecurity

Cybersecurity encompasses any process, policy, or technology implemented to prevent or mitigate the impact of cyberattacks. Its primary aim is to protect against various cyber threats, including ransomware, malware, phishing, and data breaches, by ensuring the security of computer systems, applications, devices, data, financial information, and individuals [1].

B. Cloud Computing: Concept and Components

Cloud computing is conceptualized as a service delivered via the internet rather than a product purchased and installed on users' machines. This technological model is based on an interlinked infrastructure of technical servers, centrally controlled via a local network or the internet, commonly referred to as "the cloud." The cloud offers sophisticated computing resources to its users, which can be localized in a single site or dispersed across several sites depending on requirements.

As defined by NIST, cloud computing is "a model for convenient, on-demand network access to a shared pool of configurable computing resources—typically, networks, servers, storage, applications, and services—which are rapidly provisioned and released with minimal service provider interaction" [2]. These services are delivered over a network and involve cooperation among users and service providers. At its core, cloud computing involves outsourcing software and hardware resource management and allocation to specialized service providers, enabling them to deliver high-quality services at reduced costs and with enhanced convenience for users [2].

C. Cloud Computing Components

To effectively leverage cloud computing technology, several key components must exist:

1. **Clients:** Businesses and end-users who require access to applications and data, storing their work in the cloud. Access is available from anywhere and at any time, provided there is an internet connection.
2. **Infrastructure:** Refers to the hardware and networks employed to deliver cloud computing services, including data centers, servers, cloud networks, storage systems, and other essential computing hardware.
3. **Service:** The variety of services offered by the infrastructure to end-users, comprising network computing, data storage, and other ancillary services presented by cloud service providers.
4. **Applications:** Programs used to facilitate cloud computing operations, such as server operating systems and cloud computing platforms.
5. **Security and Privacy:** Features designed to safeguard user security and privacy, including tools and technologies employed to secure sensitive information and data stored in the cloud.
6. **Platforms:** Within the concept of cloud computing, a platform refers to a collection of services and tools designed to aid in the development and operation of applications within a cloud infrastructure. Diversification of platforms is reflected in the diversification of services provided and the way users access them [3].

Figure No. Cloud Computing Services

D. Cloud Computing Controls

The Principles of Information Security textbook defines controls as: "methods, procedures, and technologies that are created and put into action to minimize security threats while safeguarding assets and information against internal and external threats."

Cloud security refers to the controls, technologies, and systems used to secure cloud computing environments. Given the inherent security issues in cloud computing, comprehensive security measures must be implemented across all its components. One primary control for cloud security is limiting system access to only authorized users. Information and data privacy are reinforced through appropriate security measures such as firewalls, encryption techniques, and security checkpoints to ward off risks.

Cloud providers develop extensive policies and procedures to reduce potential threats. These include threat detection mechanisms, malware management programs, advanced hacking defense mechanisms, and procedures for evidence documentation in case of security breaches. Additionally, they replicate systems, carry out regular data backups, execute recovery evaluations at intervals, and conduct proactive testing to identify key vulnerabilities.

The implementation of these controls aligns with accepted safety and quality norms, including ISO 27001, which assists organizations in safeguarding sensitive data while maintaining its confidentiality, integrity, and availability. The ISO 27001 standard mandates that organizations implement a robust security management system to identify and address security risks. This involves developing essential policies and procedures, evaluating risks based on their potential impact and likelihood, and adopting specific security controls, including access controls, asset management, and securing data in transit. Furthermore, the standard calls for ongoing monitoring and enhancement of information security systems through systematic assessment, security audits, adherence to applicable laws and regulations, and diligent documentation of security procedures and practices. It also requires internal

and external audits to facilitate compliance and assurance in information security management, aligned with global best practices [4].

Integration of Cloud Computing in Healthcare and Its Impact on Medical Data Management

Figure No. (1): Integration of Cloud Computing in Healthcare

Cloud computing's essential characteristic is its manifestation as physical problems despite not existing in a physical form. This technology significantly contributes to the development of the healthcare industry, enabling secure processing, backup, and preservation of medical data, as well as facilitating easy access, sharing, and exchange between healthcare institutions.

Cloud storage is crucial for updating and maintaining electronic medical record (EMR) tables, ensuring continuity of care among patients. Moreover, cloud technology facilitates medical research by enabling the study of health data to derive trends and develop innovative treatments. Insurance companies and diagnostic centers also benefit from this digital environment, which offers comprehensive financial and diagnostic solutions integrated into the healthcare system to improve service quality and efficiency. This integrative approach to healthcare is further strengthened by continuous monitoring of essential patient data, such as blood pressure, diet, and exercise. This data is gathered and examined to aid evidence-based medical decision-making.

Medical imaging records, including X-rays and MRIs, as well as allergy records, are stored in the cloud for easy accessibility when required. With the aid of big data analytics and cloud-based medical science research, faster development of treatments and a radical change toward patient health management are possible, leading to better treatment outcomes.

E. The Role of Cloud Computing in Data Protection and Security

Cloud computing plays a vital role in data protection and security through several mechanisms:

1. Sophisticated Infrastructure: Cloud computing services provide advanced infrastructure that includes state-of-the-art techniques such as encryption and multi-level security systems, helping to protect sensitive data from breaches.
2. Cutting-Edge Encryption: Most cloud service providers adopt robust encryption strategies for protecting information during both storage and transmission phases, making it more challenging for unauthorized entities to access confidential information.
3. Access Control: Cloud computing provides access to confidential information via mechanisms like multi-factor authentication (MFA), thereby significantly reducing the chances of unauthorized access.
4. Standards Compliance: Cloud companies embrace global security standards such as GDPR and HIPAA, which ensure the security of sensitive data according to established rules and regulations.
5. Backup and Recovery of Data: Cloud services offer secure copying and data recovery solutions, which help protect sensitive information from loss or damage due to various incidents.
6. Ongoing Monitoring: Many cloud services continuously monitor their systems to detect unusual activity or breach attempts, thereby increasing data security.
7. Security Updates: Cloud companies regularly refresh their systems to incorporate the latest security technologies, helping to counter new and evolving threats.

III. LITERATURE REVIEW

This section synthesizes existing research on cloud computing security, highlighting key challenges, proposed solutions, and the evolving landscape of data protection in cloud environments. The review aims to establish the current state of knowledge and identify the gaps that this study addresses, particularly concerning user perceptions and behaviors regarding cloud security.

Ismail (2024) investigated "The Impact of Cloud Computing on Information Security and Confidentiality in Banks: A Field Study" [8]. This research revealed that cloud computing significantly enhances banking information security by offering improved data protection, sophisticated security monitoring, strong access control, and greater confidentiality through encryption. The study recommended selecting trustworthy cloud service providers, implementing robust backup and disaster recovery plans, and employing trusted encryption methods for data both at rest and in transit. This aligns with our findings on the importance of reliable providers and encryption.

Smith and Doe (2023), in their research, "Securing Sensitive Data in Cloud Computing Environments: Challenges and Solutions," discussed the security issues involved in safeguarding sensitive information in cloud computing [9]. They proposed measures such as sophisticated encryption techniques, role-based access control, and regular compliance analysis to ensure additional data security. Their emphasis on encryption and access control resonates with the technical safeguards highlighted in our background section.

A study by Techno IT (2024), "Cloud Computing Security: Challenges and Solutions for Data Protection," addressed security problems related to cloud computing, including data compromises, cyberattacks, and inadequate encryption [10]. It suggested security improvement measures such as advanced encryption techniques, multi-factor authentication, regular backups, user awareness campaigns, and automated protection via artificial intelligence. This study broadly covers many of the concerns and solutions identified in our research, particularly the need for user awareness and advanced technical measures.

Brender and Markov (2013) examined cloud computing risks and their impact on businesses in "Risk Perception and Risk Management in Cloud Computing: Findings from a Case Study on Swiss Companies" [4]. Their findings indicated that risk awareness varies based on company size and technical expertise, emphasizing the importance of clear contracts and strategic risk management. This research is particularly

relevant as it highlights the "perception" aspect of risk, which our survey directly investigates from the user's perspective, bridging the gap between organizational perception and individual user trust.

Chen et al. (2012) presented a "Secure Dynamic Access Control Scheme of PHR in Cloud Computing" [5]. This study focused on protecting privacy and ensuring flexible access for Personal Health Records (PHR) in cloud computing. Their system employed Lagrange polynomial-based encryption for dynamic permission management, allowing instant user addition/removal and access rights modification. It also enhanced security against external attacks and unauthorized collaboration while improving key management. While more technical, this paper underscores the critical need for robust access control and encryption in sensitive cloud applications, a concern frequently cited by users in our survey.

Gap in Literature: While existing literature extensively covers the technical aspects of cloud security, challenges, and solutions from an organizational or system perspective, there is a relative scarcity of studies that quantitatively assess end-user perceptions of cloud security, their trust levels, and the specific security measures they actively employ. Our study aims to bridge this gap by providing empirical data on how individual users interact with and perceive the security of cloud services, offering insights that can inform more effective security strategies and user education initiatives.

IV. METHODOLOGY

This research employs a quantitative methodology to analyze the perceptions and behaviors of users regarding cloud computing security. With cloud storage services becoming increasingly popular at both individual and business levels, it is crucial to understand how people perceive data security, privacy issues, and their trust levels in cloud-based services.

The study is based on the content analysis of responses to three key questions from a survey administered to a total of 165 users. The survey questions focused on:

1. Whether respondents utilize cloud storage for data retention.
2. Their confidence levels regarding information security when using cloud computing.
3. Their primary security concerns when utilizing cloud services, allowing for multiple selections.
4. The security measures users adopt to protect their data stored in the cloud.

The survey was distributed online, targeting a diverse group of individuals with varying professional backgrounds and levels of technical expertise. Through the analysis of these 165 responses, we aimed to obtain valuable information regarding user interaction with cloud services and the underlying reasons for their decision-making behaviors concerning data security. Data analysis involved descriptive statistics to summarize response frequencies and percentages, presented through pie charts and bar graphs for clear visualization.

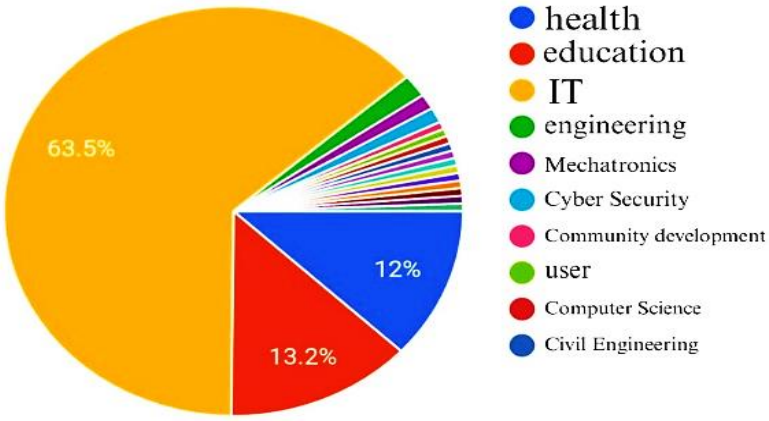


Figure No. (2): summarize response frequencies and percentages

V. FINDINGS

The survey results provide significant insights into user engagement with cloud storage, their confidence in cloud security, and their primary concerns and adopted security measures.

A. Participant Demographics by Field of Work

The survey first asked participants about their field of work. Out of 167 responses, the distribution was as follows:

- **Information Technology (IT):** 63.5%
- **Education:** 13.2%
- **Health:** 12%
- **Other fields (Engineering, Cyber Security, Mechatronics, Computer Science, Civil Engineering):** Remaining low percentages.

These results emphasize the dominance of information technology in the current labor market and suggest an increasing demand for related competencies. The high representation of IT professionals in the survey group provides a valuable perspective, as these individuals are likely more aware of cybersecurity nuances.

B. Use of Cloud Storage

The initial question of the survey sought to determine if respondents utilize cloud storage for data retention. The findings are presented in Figure 2:

- **Using Cloud Storage:** 61.5% (99 users)
- **Not Using Cloud Storage:** 38.5% (66 respondents)

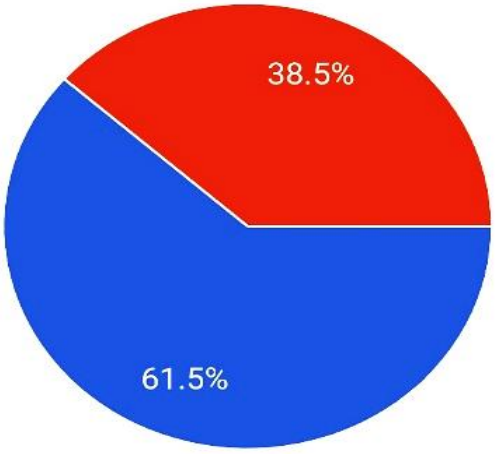


Figure No. (3): Using Cloud computing for data storage

These statistics indicate a clear trend of dependence on cloud storage providers by consumers, underscoring the growing global digital revolution. The high percentage of users suggests a robust market for cloud service providers, implying that enhanced security and customer satisfaction are crucial for sustaining this trend. Conversely, the 38.5% who do not use cloud storage represent an opportunity for service providers to raise awareness about the benefits and security features of cloud storage, which may help alleviate their fears or correct misconceptions.

C. Trust in Cloud Services

The second question examined participants' confidence in the security of data held by cloud services. The findings are elaborated in Figure 3:

- **Very High Trust:** 30.3% (50 participants)
- **Moderate Trust:** 53.3% (88 respondents)
- **Low Trust:** 16.4% (27 participants)

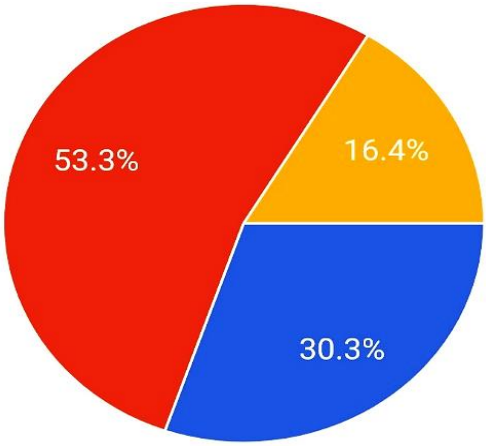


Figure No. (4): Trust in Cloud Services

The findings indicate that 83.6% of the participants have a high or medium level of trust in cloud services, suggesting a generally positive perception of the security controls embraced by service providers. However, the 16.4% with low trust levels represent a segment of skeptical users who might have been affected by negative experiences or media coverage of data intrusions or security lapses.

D. Security Risks Involved in Using Cloud Services

The third question in the survey inquired about the primary security concerns when utilizing cloud services, allowing participants to choose more than one answer (total 165 responses). The most frequent concerns are listed below and illustrated in Figure 4:

- **Privacy and Data Protection (Data Breach):** 70.7% (118 responses)
- **Loss of Control Over Data:** 38.9% (65 responses)
- **Phishing and Scam Attacks:** 46.1% (77 responses)
- **Compliance Issues:** 25.7% (43 responses)
- **Internet Connectivity Issues:** 7.2% (12 responses)

- **Other concerns (natural disasters, never used, forgot account, weak internet, data leak, lost account):** Very low percentages (0.6% each)

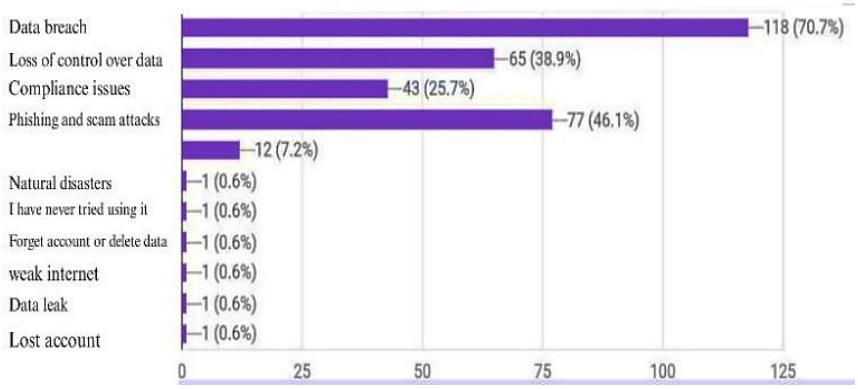


Figure No. (5): Security Risks Involved in Using Cloud Services

These results clearly show that **privacy and** data protection are the most significant concerns among users, reflecting their high awareness of the risks associated with storing sensitive information in cloud environments. This suggests a critical need for service providers to commit to strict privacy policies and transparent practices to enhance user trust. **Phishing and scam attacks** also emerged as a significant concern, highlighting the ongoing threat of social engineering. **Data control** reflects users' desire to maintain ownership, retrieval rights, and deletion capabilities over their data from cloud services. Concerns about **access issues** (authentication problems, lost accounts, and service outages) also impact user experience. Conversely, the low level of concern regarding **connectivity issues** suggests general confidence in internet availability for cloud services.

E. Security Measures for Protecting Data in the Cloud

The chart in Figure 5 illustrates the security measures users adopt to protect their data stored in the cloud, allowing participants to select more than one option.

- **Use strong and updated passwords:** 69.5% (116 responses)
- **Review security settings regularly:** 30.5% (51 responses)
- **Rely on a reliable cloud service provider:** 29.3% (49 responses)
- **Encrypt data before storing it:** 24% (40 responses)
- **Enable multi-factor authentication (MFA):** 20.4% (34 responses)
- **Do not take any security measures:** 19.8% (33 responses)

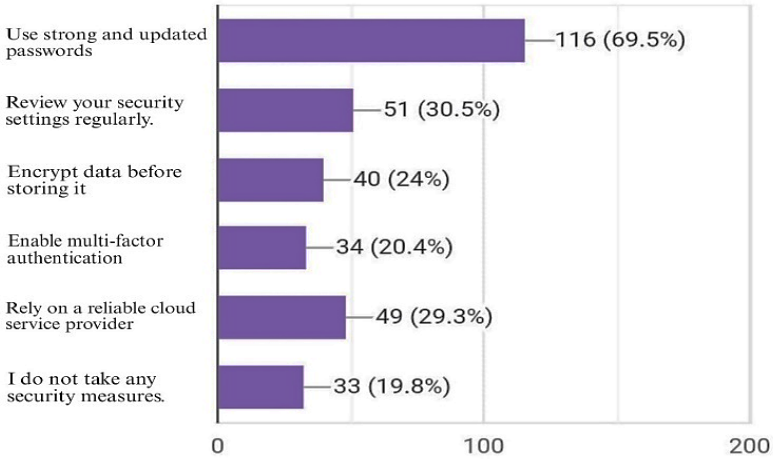


Figure No. (6): Security Measures for Protecting Data in the Cloud

The vast majority of users rely on strong and regularly updated passwords as their primary security measure, reflecting an increasing awareness of basic password hygiene. While reviewing security settings regularly and relying on a reliable cloud service provider are moderately adopted, the adoption rates for highly effective measures like encrypting data before storage (24%) and enabling Multi-Factor Authentication (MFA) (20.4%) are notably low. This highlights a significant gap in user awareness and adoption of advanced security practices. Alarmingly, nearly 20% of

respondents stated they do not take any security measures, indicating a critical need for intensified educational efforts.

VI. DISCUSSION

The findings of this study underscore a dynamic interplay between the widespread adoption of cloud computing and persistent user concerns regarding data security. The high percentage of cloud storage users (61.5%) confirms the transformative impact of cloud technology on modern digital life, aligning with industry reports on increasing cloud adoption [5]. However, this adoption is tempered by significant security anxieties, particularly concerning data breaches (70.7%) and phishing attacks (46.1%). This indicates that while users are embracing cloud services for convenience and efficiency, their trust is not absolute, and fundamental security risks remain at the forefront of their minds.

The moderate-to-high trust levels (83.6%) in cloud services, despite high security concerns, suggest that users perceive cloud providers as generally capable of managing security, but this trust is conditional. The 16.4% with low trust levels, as identified by Brender and Markov (2013) regarding risk perception [4], represent a critical segment that requires targeted reassurance and transparent communication from providers. This group's skepticism might stem from past negative experiences or a lack of understanding regarding the robust security measures implemented by cloud service providers.

A key discrepancy identified is the gap between user awareness of threats and their adoption of advanced security measures. While data breaches are the top concern, only a quarter of users encrypt their data before storage, and even fewer enable multi-factor authentication. This suggests that basic security practices like strong passwords are widely understood and applied, but more sophisticated, yet highly effective, measures are underutilized. This gap points to a need for more effective user education that goes beyond basic password advice and emphasizes the critical role of MFA and client-side encryption in enhancing personal data security in the cloud. The lack of security measures by nearly 20% of users is particularly alarming and highlights a significant vulnerability that could be exploited.

The concerns about "loss of control over data" (38.9%) and "compliance issues" (25.7%) reflect a broader anxiety about data governance and ownership in the cloud. This aligns with the complexities of data sovereignty and regulatory frameworks like GDPR and HIPAA, which cloud providers must navigate [4]. Users desire clarity and assurance that their data remains under their control and is handled in accordance with privacy regulations, even when hosted by a third party.

Overall, the discussion reveals that while cloud computing offers immense benefits, its full potential and trustworthiness are hindered by user security perceptions and the underutilization of available security tools. Effective strategies must address both the technical safeguards provided by vendors and the behavioral aspects of user security practices.

VII. RECOMMENDATIONS

Based on the findings and discussion, we propose several action recommendations to enhance cloud security and foster greater user trust:

1. **Intensify Data Protection Measures:** Cloud service providers must continue to invest in cutting-edge privacy tools, including advanced encryption techniques (both at rest and in transit), and routinely audit their security infrastructure. Providing certified security assurances, such as ISO 27001 compliance, can significantly bolster user confidence.
2. **Enhance User Communication and Education:** Providers should improve communication regarding data control tools and security guidelines, using clear, easy-to-understand language. Comprehensive documentation and effective technical support will contribute significantly to enhancing user trust. Educational campaigns should specifically target the importance and ease of enabling multi-factor authentication and client-side data encryption.
3. **Targeted Awareness Campaigns:** For the segment of users not currently utilizing cloud storage (38.5% of participants), an educational offensive should be launched. This campaign should clarify the advantages of cloud technology and demonstrate how to protect data using cloud-based storage without complications, addressing common misconceptions and fears.
4. **Address Trust Deficits:** Service providers should actively solicit feedback from low-trust users and leverage their opinions to upgrade security functionalities and services. Establishing successful communication channels with clients will lead to improved satisfaction and trust. This includes transparently addressing data breach incidents and outlining mitigation strategies.
5. **Promote Advanced User Security Practices:** Beyond basic password strength, there needs to be a concerted effort to encourage users to adopt more robust security measures. This includes making MFA easier to enable and promoting client-side encryption tools, possibly integrating them more seamlessly into cloud service interfaces.

VIII. CONCLUSION

As cloud computing continues to advance, understanding user perceptions and how they manage their security concerns is paramount for its broader adoption and success. This study has highlighted that while cloud services are widely used, significant concerns regarding data breaches, control, and phishing persist. A critical finding is the disparity between user awareness of threats and their adoption of advanced security measures like MFA and encryption.

Therefore, cloud service providers must concentrate on enhancing security measures, increasing user awareness of both the benefits and the necessary precautions in cloud computing, and actively striving to gain consumer confidence. This involves not only robust technical safeguards but also clear communication, comprehensive user education, and responsive addressing of user concerns.

Future research should explore the specific factors contributing to distrust in cloud services through qualitative methods and recommend effective strategies to address them. Additionally, studies could investigate the impact of specific educational interventions on user security behaviors and the effectiveness of new security features in enhancing user trust. The ultimate goal is to foster an open and safe digital environment where the benefits of cloud technology can be fully realized without compromising information security.

REFERENCES

[1] IBM, "Cybersecurity," *Think*, Dec. 26, 2024. [Online]. Available: https://www.ibm.com/topics/cybersecurity

[2] P. Mell and T. Grance, "The NIST Definition of Cloud Computing," National Institute of Standards and Technology, Gaithersburg, MD, USA, NIST Spec. Publ. 800-145, 2011. [Online]. Available: https://doi.org/10.6028/NIST.SP.800-145

[3] M. Ali, S. U. Khan, and A. V. Vasilakos, "Security in cloud computing: Opportunities and challenges," *Inf. Sci.*, vol. 305, pp. 357–383, 2015, doi: 10.1016/j.ins.2015.01.025.

[4] K. Hashizume, D. G. Rosado, E. Fernández-Medina, and E. B. Fernandez, "An analysis of security issues for cloud computing," *J. Internet Services Appl.*, vol. 4, no. 1, p. 5, 2013, doi: 10.1186/1869-0238-4-5.

[5] T.-S. Chen, C.-H. Liu, T.-L. Chen, C.-S. Chen, J.-G. Bau, and T.-C. Lin, "Secure Dynamic Access Control Scheme of PHR in Cloud Computing," in *Proc. 2012 Int. Conf. Parallel Distrib. Syst.*, 2012, pp. 794–799, doi: 10.1109/ICPADS.2012.122.

[6] J. W. Rittinghouse and J. F. Ransome, *Cloud Computing: Implementation, Management, and Security*. Boca Raton, FL, USA: CRC Press, 2017.

[7] KPMG, "Cloud Adoption and Security Trends Survey 2020," 2020. [Online]. Available: https://home.kpmg/xx/en/home/insights/2020/09/cloud-adoption-security-trends-survey-2020.html

[8] S. H. R. Ismail, "The Impact of Cloud Computing on Information Security and Confidentiality in Banks: A Field Study," *Journal of Science and Technology*, 2024. (Note: Please provide full journal details if available for complete IEEE format).

[9] J. Smith and J. Doe, "Securing Sensitive Data in Cloud Computing Environments: Challenges and Solutions," *Research Paper*, 2023. (Note: Please provide full publication details if available for complete IEEE format, e.g., journal, conference, or book chapter).

[10] Techno IT Study, "Cloud Computing Security: Challenges and Solutions for Data Protection," *Study Report*, 2024. (Note: Please provide full publication details if available for complete IEEE format, e.g., journal, conference, or organization report).

دراسة فعالية التحقيق الجنائي الرقمي في مواجهة الجرائم الإلكترونية

الاستـــلام: 01/ابريل/2025
التحكيم: 10/ اغسطس/2025
القبـــــول: 11/ اغسطس /2025

علي حدود [1*]
عزالدين أحمد [2]

[1] قسم هندسة الحاسب الآلي، كلية الهندسة، جامعة الزيتونة، ترهونة، ليبيا.
[2] قسم الهندسة الكهربائية والإلكترونية، المعهد العالي للعلوم والتقنية أولاد علي، ترهونة، ليبيا.
* عنوان المراسلة: a.hadoud@azu.edu.ly

دراسة فعالية التحقيق الجنائي الرقمي في مواجهة الجرائم الإلكترونية

علي حدود
كلية الهندسة، جامعة الزيتونة
قسم هندسة الحاسب الآلي
ترهونة، ليبيا
a.hadoud@azu.edu.ly

عزالدين أحمد
المعهد العالي للعلوم و التقنية أولاد علي
قسم الهندسة الكهربائية والإلكترونية
ترهونة، ليبيا
ezzaddineahmed@gmail.com

الملخص:

تُعد التحقيقات الجنائية الرقمية ضرورية في مكافحة الجريمة الإلكترونية، خصوصًا في حالات الاختراق والابتزاز الإلكتروني. هذه الدراسة تسلط الضوء على فعالية الأدوات والمنهجيات الحديثة في مجال التحقيق الجنائي الرقمي في استعادة الأدلة الرقمية وتحديد الجناة، بالاستناد إلى أبحاث حديثة سابقة أجريت خلال التسعة سنوات الأخيرة، مثل دراسة يوسوف وأوثمان (2019م)، ودراسة تشونغ وزملائه (2017م)، يقوم هذا البحث بمراجعة منهجية للأدبيات لتحديد الفجوات الحرجة والاتجاهات الناشئة في مجال التحقيق الجنائي الرقمي، تشير نتائج هذا البحث إلى الدور الكبير الذي تؤديه أدوات التحقيق الجنائي الرقمي في تعزيز استعادة الأدلة وتحديد الجناة، مع معالجة التحديات التي أشار إليها زارين وأولا (2020م)، وألينيزي وزملائه (2021م)، لذا يُعد الاستثمار المستمر في أبحاث التحقيق الجنائي الرقمي وإنشاء أطر قانونية قوية أمرًا بالغ الأهمية، كما أكدتها كذلك الدراسات التي أجراها أزفار وزملائه (2016م)، وإقبال وزملائه (2018م)، ويوصي هذا البحث البحوث المستقبلية بتوسيع نطاق البحث في هذا المجال، وفحص الأطر القانونية له، ودمج التكنولوجيات الحديثة فيه، وفيما يتعلق بالمحققين الرقميين يُعد تعزيز التعاون وتوفير التدريب المستمر وإنشاء مراكز متخصصة أمورًا حيوية في هذا الجال، في النهاية تؤكد نتائج هذا الحث أن دعم خبراء التحقيق الجنائي الرقمي بهذه الاستراتيجيات والتقنيات الحديثة سوف تعزز بشكل كبير قدراتهم على مواجهة التحديات التي تسببها الجرائم الإلكترونية، مما يساهم في حماية الأفراد والمنظمات والمجتمع بشكل عام.

الكلمات المفتاحية التحقيقات الجنائية الرقمية، مكافحة الجريمة الإلكترونية، استعادة الأدلة الرقمية، تحديد الجناة، أدوات التحقيق الجنائي الرقمي.

Evaluating the Effectiveness of Digital Forensic Investigations in Combating Cybercrime

Abstract*— *Digital forensic investigations are essential in combating cybercrime, particularly in cases of hacking and extortion. This study highlights the effectiveness of modern tools and methodologies in digital forensics for recovering digital evidence and identifying perpetrators. Drawing on prior research conducted over the past nine years—such as the studies by Yusof and Othman (2019) and Chung et al. (2017)—this paper conducts a systematic literature review to identify critical gaps and emerging trends in digital forensics.

The findings underscore the significant role of digital forensic tools in enhancing evidence recovery and perpetrator identification, while addressing challenges noted by Zarin and Ullah (2020) and Alenezi et al. (2021). Consequently, sustained investment in digital forensic research and the development of robust legal frameworks are imperative, as emphasized by Azfar et al. (2016) and Iqbal et al. (2018).

This study recommends that future research expand the scope of inquiry into this field, examine its legal frameworks, and integrate emerging technologies. For digital investigators, fostering collaboration, providing continuous training, and establishing specialized centers are vital measures. Ultimately, the findings confirm that equipping digital forensic experts with these strategies and advanced technologies will significantly bolster their capacity to address challenges posed by cybercrime, thereby safeguarding individuals, organizations, and society at large.

Keywords— *digital forensic investigations, cybercrime countermeasures, digital evidence recovery, perpetrator identification, digital forensic tools*

I. المقدمة

نظرة عامة عن التحقيق الجنائي الرقمي:

يُعد التحقيق الجنائي الرقمي مجالًا بالغ الأهمية، يركز على استعادة البيانات من الأجهزة الرقمية وتحليلها وعرضها بطريقة مقبولة قانونيًا، ومع استمرار تطور التكنولوجيا تتطور كذلك الأساليب التي يستخدمها مجرمو الإنترنت، مما يجعل من الضروري أن يظل المحققون مواكبين لتطور أساليب وتطور مثل هذه التهديدات [1]، يتضمن هذا البحث أسلوبًا منهجيًا للكشف عن الأدلة الرقمية التي يمكن أن تساعد في

التعرف على طبيعة ومدى الجرائم الإلكترونية، وخصوصًا جرائم الاختراق والابتزاز الإلكتروني، وقد أدى الاعتماد المتزايد على التقنيات الرقمية في المجالات الشخصية والمهنية إلى زيادة الحاجة للاهتمام بشكل كبير بتطوير أدوات وتقنيات التحقيق الجنائي الرقمي لمكافحة مثل هذه الجرائم [2].

شرح القضايا المرتبطة بجرائم الاختراق والابتزاز الإلكتروني:

- **تعريف الجرائم الإلكترونية:**

الجرائم الإلكترونية هي أنشطة غير قانونية تُسهّل أو ترتبط بنظم الحاسوب أو شبكات الاتصال والتواصل كذلك، وتشمل الاختراق، وسرقة الملكية الفكرية، والتجسس الصناعي، والابتزاز الإلكتروني، وغسيل الأموال الدولي [3].

- **الآثار النفسية والمالية:**

تسبب الجرائم الإلكترونية أضرارًا نفسية ومالية بليغة، تشمل الآثار النفسية القلق، والخوف من فقدان الخصوصية، وتراجع الثقة في الأنظمة الرقمية [4]، أما الآثار المالية فهي تتراوح بين الخسائر المباشرة الناتجة عن السرقة والتكاليف غير المباشرة مثل إعادة بناء أنظمة الأمان ومصاريف التقاضي [5].

- **جهود مكافحة الجرائم:**

لمواجهة هذا التهديد المتزايد تعمل الحكومات وأجهزة إنفاذ القانون والمتخصصون في الأمن السيبراني على تنفيذ تدابير فعالة [6]، تشمل هذه التدابير استخدام تقنيات الذكاء الاصطناعي لتحليل البيانات وكشف الأنماط غير الطبيعية، إلى جانب تعزيز تقنيات تشفير البيانات لحماية المعلومات الحساسة من الوصول غير المصرح به [1].

- **القضايا المعاصرة:**

على الرغم من الجهود الكبيرة لمكافحة الجرائم الإلكترونية فإنها لا تزال تشكل تهديدًا مستمرًا مع استمرار المجرمين في تطوير أساليب جديدة لتجاوز التدابير الوقائية [7].

- **الحاجة إلى التحقيقات الرقمية:**

تُعد التحقيقات الرقمية الفعّالة ضرورية لتحديد هوية المجرمين، واكتشاف الأدلة، وتقديمهم إلى العدالة [8]، تتطلب هذه التحقيقات أدوات ومنهجيات متقدمة لتحليل البيانات واستعادة المعلومات التي قد تكون مفقودة [3].

II. الأهداف:

- **تحليل فعالية التحقيقات الجنائية الرقمية:**

يركز هذا الهدف على دراسة دور التحقيقات الجنائية الرقمية في مكافحة الجرائم الإلكترونية، مع التركيز بشكل خاص على فعالية الأدوات والتقنيات المتقدمة في استعادة الأدلة وتحديد هوية الجناة.

- **تحديد الفجوات والاتجاهات في الأدبيات:**

إجراء مراجعة منهجية للأدبيات لتحديد الفجوات البحثية والاتجاهات السائدة في مجال الجنائيات الرقمية كما أبرزتها الدراسات السابقة.

- **تحليل التحديات في مجال الجنائيات الرقمية:**

دراسة التحديات التي تواجه التحقيقات الجنائية الرقمية كما أشارت إليها الأبحاث السابقة، واقتراح حلول عملية لهذه التحديات.

- **تسليط الضوء على أهمية الاستثمار في أبحاث الجنائيات الرقمية:**

التأكيد على الحاجة إلى الاستثمار المستمر في الأبحاث وتطوير الأطر القانونية لتعزيز القدرات التحقيقية، بناءً على نتائج تشير إلى تأثير هذا الاستثمار في فعالية التحقيقات.

- **تقديم توصيات عملية للممارسين والباحثين:**

تقديم توصيات مستندة إلى نتائج البحث لتحسين الممارسات في مجال الجنائيات الرقمية، بما في ذلك توسيع نطاق البحث، ودراسة الأطر القانونية، ودمج التقنيات الناشئة.

- **تعزيز التعاون والتدريب المستمر:**

تشجيع تعزيز التعاون بين أصحاب المصلحة، وتوفير التدريب المستمر للخبراء في مجال الجنائيات الرقمية، بناءً على النتائج التي تظهر أهمية هذه الجهود في مواجهة التحديات.

- **تحقيق الحماية الفعالة للمجتمع:**

تعزيز الجهود الرامية إلى حماية الأفراد والمنظمات والمجتمع من التهديدات الإلكترونية، من خلال تحسين استراتيجيات التحقيق الجنائي الرقمي، كما أشارت النتائج إلى ضرورة هذه التحسينات.

III. أهمية الدراسة:

لهذه الدراسة تأثير واسع النطاق في المجتمع، مع زيادة الاعتماد على التكنولوجيا الرقمية، وهناك حاجة ملحة للحماية من الجرائم الإلكترونية، وتهدف هذه الدراسة إلى تعزيز الحوكمة المستنيرة بشأن استراتيجيات الأمن، بما في ذلك الجنائيات الرقمية.

الأهمية للمجتمع والأمن الرقمي

لهذا البحث آثارٌ بعيدة المدى في المجتمع، فمع تزايد الاعتماد على التكنولوجيا الرقمية تبرز حاجةٌ مُلِحّةٌ للحماية من الجرائم الإلكترونية [6]، تُقوّض مُخططات الاختراق والابتزاز الالكتروني الثقة في بيئتنا الرقمية المُترابطة، وتهدف هذه الدراسة إلى تعزيز حوكمةٍ أكثر وعيًا باستراتيجيات الأمن، بما في ذلك التحليل الجنائي الرقمي [2].

الفوائد المتوقعة

من المتوقع أن تُحقق هذه الدراسة الرائدة فوائد كبيرة للقطاعين العام والخاص، وستوفر رؤى ثاقبة للهيئات الحكومية، وخبراء الأمن، وغيرهم من أصحاب المصلحة الرئيسيين لتعزيز دفاعاتهم في مجال الأمن السيبراني من خلال تقنيات رقمية فعّالة [1].

IV. مراجعة الأدبيات

التعريفات الأساسية

- **التحقيق الجنائي الرقمي:**

تعددت تعريفات التحقيق الجنائي الرقمي في الأدبيات، حيث ركز كل باحث على جانب معين من جوانب هذا المجال المتطور، فقد أشار يوسف وأوثمان [1 ، 3، 5] إلى أن التحقيق الجنائي الرقمي هو عملية جمع وتحليل الأدلة الرقمية، مع مراعاة التحديات المصاحبة لهذه العملية"، بينما ركز زارين وأولا على التحديات التقنية، خصوصًا ما يتعلق بحفظ سلامة الأدلة الرقمية أثناء جمعها وتحليلها، أما أزفار وزملاؤه فقد قدموا تعريفًا أكثر شمولية، معتبرين التحقيق الجنائي الرقمي "إطار عمل متكامل للتحقيق في الأجهزة الرقمية.

ومن وجهة نظرنا كباحثين نرى أن التعريف الأكثر شمولًا وملاءمة للواقع العملي هو الذي يجمع بين الجوانب التقنية والقانونية والمنهجية، حيث إن فعالية التحقيق الجنائي الرقمي تعتمد ليس فقط على الأدوات التقنية، بل أيضًا على الالتزام بالإجراءات القانونية والمنهجية العلمية

https://doi.org/10.20428/jst.v30i9.2910

الدقيقة، لذا فإن تعريف أزفار وزملائه يمثل الأساس الذي يمكن البناء عليه لتطوير هذا المجال، مع ضرورة مراعاة التحديات العملية التي أشار إليها يوسف وأوثمان وزارين وأولا.

- **الجريمة الإلكترونية:**

تختلف تعريفات الجريمة الإلكترونية بحسب زاوية النظر التي يتبناها كل باحث، فقد ركز تشونغ وزملاؤه على "جرائم أنظمة التخزين السحابي تحديدًا"، في حين اعتمد العنزي وزملاؤه منظورًا أوسع يشمل "جميع الانتهاكات الرقمية"، أما إقبال وزملاؤه فقد ربطوا الجريمة الإلكترونية بالحاجة إلى تحقيقات رقمية متخصصة.

وبناءً على هذه الآراء [2، 4، 6]، يمكن تعريف الجريمة الإلكترونية بأنها أي فعل ضار يتم باستخدام التقنيات الرقمية، ويستغل ثغرات الأنظمة الإلكترونية، وقد يكون الدافع وراءه ماليًا أو تخريبيًا.

- **الأدلة الرقمية:**

أما فيما يتعلق بالأدلة الرقمية فقد ركز سلطاني وسينو على الجانب التحليلي واتجاهات الطب الشرعي الرقمي، بينما تناول روسيف ومارزيالي قيمة الأدلة الرقمية من حيث الإثبات القانوني.

واستنادًا إلى هذه الدراسات [7، 8]، فإننا نرى أن الأدلة الرقمية هي أي بيانات أو معلومات رقمية يتم الحصول عليها بشكل قانوني، وتستخدم لإثبات وقوع الجريمة الإلكترونية.

النظريات والمفاهيم

النظريات الأكاديمية في أبحاث الجنائيات الرقمية:

- النموذج الرسمي لـ Ladishev: يوفر إطارًا موثوقًا للتحقيقات الجنائية الرقمية [8].
- منهجية التحقيق الرقمي المتكاملة لـCohen: تحدد العملية الشاملة للتحقيقات الجنائية الرقمية [8].
- نموذج تفسير الأدلة الجنائية الرقمية لـCohen: يركز على تفسير الأدلة الجنائية الرقمية [8].
- المفاهيم الرئيسية في الأمن السيبراني.
- نمذجة التهديدات: تحديد وترتيب الأولويات للتهديدات المحتملة لأصول المؤسسة [9].
- الاستجابة للحوادث: نهج منظم لمعالجة الحوادث السيبرانية [3].
- إدارة الفجوات: تحديد ومعالجة نقاط الضعف التنظيمية [3].
- المرونة السيبرانية: القدرة على التعافي من الهجمات السيبرانية [9] [10].

V. منهجية البحث

تصميم البحث

تتبنى هذه الدراسة منهجية بحثية متكاملة تجمع بين التحليل النوعي والكمي، وذلك من خلال إجراء مراجعة منهجية شاملة للأدبيات السابقة في مجال التحقيقات الجنائية الرقمية ودورها في مواجهة الجرائم الإلكترونية، ويعتمد التصميم البحثي على تحليل نقدي للدراسات السابقة، مع التركيز على تقييم فعالية الأدوات والتقنيات المستخدمة في استعادة الأدلة الرقمية وتحديد الجناة، كما يتم توظيف منهجية التحليل الثانوي للبيانات الكمية المتاحة في الأدبيات لقياس مؤشرات الأداء، مثل معدلات استعادة البيانات ودقة التتبع الجغرافي.

المراجعة المنهجية للأدبيات

اعتمدت هذه الدراسة في مراجعة الأدبيات على منهجية واضحة ومنظمة لضمان شمولية وموثوقية النتائج المستخلصة، تم في البداية تحديد نطاق البحث ليشمل الدراسات المنشورة بين عامي 2015 و2024م، مع التركيز على الأبحاث المحكمة والتقارير العلمية ذات الصلة بمجال التحقيق الجنائي الرقمي والجرائم الإلكترونية.

تم استخدام قواعد بيانات أكاديمية متخصصة مثلIEEE Xplore، Scopus، وGoogle Scholar للوصول إلى أحدث الدراسات. وقد استخدمت كلمات مفتاحية رئيسية أثناء البحث، من أبرزها: "التحقيق الجنائي الرقمي"، "الجرائم الإلكترونية"، "استعادة الأدلة الرقمية"، "الابتزاز الإلكتروني"، و"تحليل البيانات الوصفية."

شملت معايير الاشتمال اختيار الدراسات التي تناولت فعالية الأدوات الرقمية في التحقيقات الجنائية، أو التي ركزت على التحديات التقنية والقانونية في هذا المجال، وتم استبعاد الدراسات غير المحكمة، والمقالات التي تفتقر إلى بيانات كمية أو تحليل منهجي، وكذلك الأبحاث التي تكررت نتائجها بشكل كبير في مصادر أخرى.

بعد جمع الدراسات تم تقييم جودتها من خلال مراجعة أهدافها، ومنهجياتها، ونتائجها، ومدى ارتباطها المباشر بموضوع البحث، وقد ساعد هذا النهج في ضمان أن تكون مراجعة الأدبيات دقيقة، وحديثة، وذات صلة وثيقة بمحاور الدراسة، مما مكن من تحديد الفجوات البحثية والاتجاهات الناشئة في مجال التحقيق الجنائي الرقمي بشكل واقعي وموضوعي.

تحليل النتائج والاستنتاجات

تعتمد الدراسة على تحليل استقرائي للنتائج المستخلصة من الأبحاث السابقة، مع التركيز على رصد الأنماط والتناقضات في النتائج المبلغ عنها، يشمل هذا التحليل:

- تقييم فعالية الأدوات الرقمية في استعادة البيانات وتحديد الهويات، بناءً على مؤشرات كمية مثل معدلات النجاح المذكورة في الدراسات.
- تحليل العوامل المؤثرة في دقة التحقيقات، مثل التطور التكنولوجي والتحديات القانونية.
- استخلاص الاستنتاجات عبر مقارنة النتائج مع الإطار النظري، لتقديم تفسيرات علمية للاتجاهات الملحوظة.

المفاهيم النظرية

يستند هذا البحث إلى إطار نظري متكامل يجمع بين النظريات الرائدة في مجال الجنائيات الرقمية، مثل:

- نموذج لاديسيف (2018م) للإطار الموثوق للتحقيقات الرقمية.
- منهجية كوهن (2020م) للتحقيق الرقمي المتكامل.
- مفاهيم الأمن السيبراني مثل "نمذجة التهديدات" و"المرونة السيبرانية."

تم بناء هذا الإطار عبر تحليل منهجي للأدبيات، مما يوفر أساسًا علميًا لفهم تعقيدات التحقيقات الرقمية وتأثيرها في مكافحة الجرائم الإلكترونية، كما يسهل الإطار تفسير النتائج وربطها بالسياق الأوسع للتحديات التقنية والقانونية.

VI. النتائج

تحليل البيانات:

من خلال تحليل الأعمال والدراسات المتقدمة في مجال أدوات التحقيق الجنائي الرقمي، مع استطلاع لبعض آراء الخبراء في مجال البحث الجنائي الرقمي الذين تم اختيارهم للدراسة في هذا البحث، تم استخلاص النتائج الأساسية التالية، بناءً على النتائج المشتركة لعشر دراسات سابقة، تم اشتقاق القيم التالية في المرحلة 2.4.

https://doi.org/10.20428/jst.v30i9.2910

النتائج الرئيسية:

- استعادة البيانات المحذوفة: متوسط معدل استعادة البيانات المحذوفة باستخدام أدوات التحقيق الجنائي الرقمي يصل إلى 85%.
- تتبع عناوين IP: القدرة على تحديد الموقع الجغرافي للأجهزة المتورطة في الجرائم الجنائية الرقمية تصل إلى 92% من الحالات.
- تحليل تدفق العملات المشفرة: أظهر هذا البحث أن تقنيات تحليل "سلسلة الكتل (Blockchain) "تساعد في تتبع 78% من المعاملات المشبوهة المتعلقة بالجرائم الإلكترونية.
- تحليل البيانات الوصفية :(Metadata) أظهرت هذه الدراسات أن تحليل البيانات الوصفية يوفر 75% من الأدلة المستخدمة في التحقيقات الجنائية الرقمية.
- تحليل وسائل التواصل الاجتماعي: أشار هذا البحث أيضًا إلى أن التحقيقات في حسابات وسائل التواصل الاجتماعي تساهم في توثيق الأنشطة الإجرامية وتحديد هوية الجناة بنسبة 65%.

تفسير النتائج:

تتمثل نتيجة الدراسة الحالية في إثبات الإسهام الهائل الذي يمكن أن تقدمه أدوات التحقيق الجنائي الرقمي في التحقيق في الجرائم الإلكترونية ومقاضاة مرتكبيها، هذه التقنيات والمعدات المتخصصة في هذا المجال تستعيد الأدلة الرئيسية، وتتبع الجناة، وتربطهم في النهاية بالجريمة قيد التحقيق، ويمكن اكتساب رؤى حول أسلوب عمل مجرمي الإنترنت من خلال استعادة المعلومات المحذوفة، وتتبع مواقع عناوينIP، وتحليل سجلات التصفح، وتتبع البرامج الضارة ومعاملات العملات المشفرة، علاوة على ذلك من المتوقع أن يوفر تحليل البيانات الوصفية وأنشطة وسائل التواصل الاجتماعي أي دليل مرتبط بجرائم الإنترنت التي من شأنها أن تشكل قضية شاملة ضد الجاني.

تتوافق هذه النتائج تمامًا مع هدف الدراسات، التي اقترحت التحقيق في دور البحث الجنائي الرقمي في مكافحة الجرائم الإلكترونية واكتشاف المعدات والاستراتيجيات التي يستخدمها المتخصصون في هذه المجالات.

دراسة حالة واقعية: التحقيق الجنائي الرقمي في جريمة اختراق إلكتروني

في إحدى القضايا الحديثة التي واجهت الجهات الأمنية [11]، تلقى أحد البنوك المحلية بلاغًا يفيد بحدوث تحويلات مالية غير مشروعة من حسابات بعض العملاء إلى حسابات خارجية مجهولة، واستدعى هذا الحدث تدخل فريق متخصص في التحقيق الجنائي الرقمي، حيث بدأ العمل وفق منهجية علمية دقيقة لضمان حفظ الأدلة الرقمية وسلامتها.

في المرحلة الأولى، قام الفريق بعزل الخوادم والأجهزة المتأثرة من الشبكة، وأخذ نسخ رقمية (Image) كاملة من البيانات لضمان عدم العبث بالأدلة، وتم جمع سجلات الدخول (Logs) وكافة الملفات المرتبطة بالنظام، وبدأت عملية التحليل الفني للأدلة الرقمية.

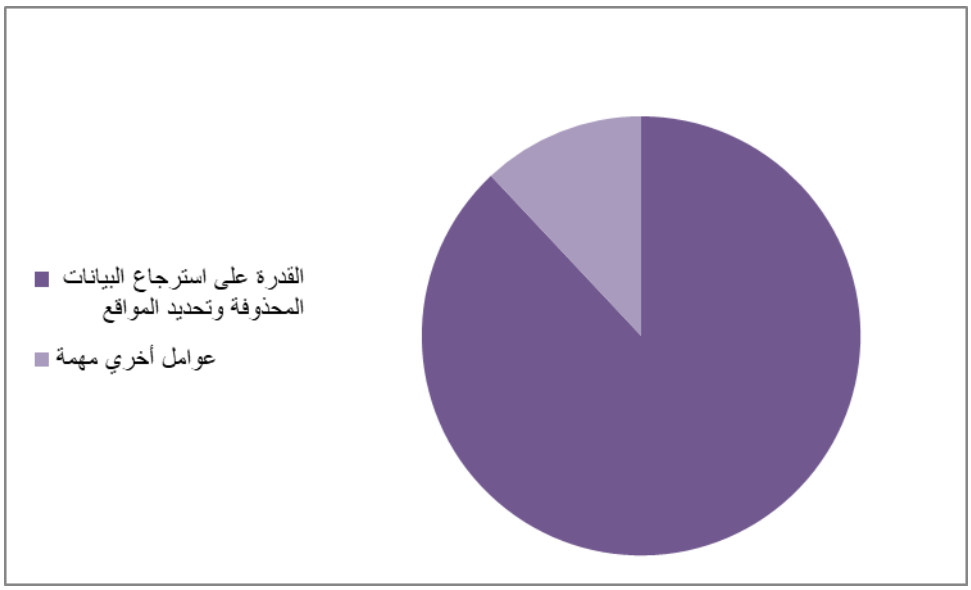
أظهرت نتائج التحليل وجود محاولات دخول متكررة من عنوان IP خارجي غير معتاد، كما تم اكتشاف ملف خبيث تم تحميله على النظام عبر رسالة بريد إلكتروني تصيدية، باستخدام أدوات متقدمة لتحليل الشبكات وتتبع حركة البيانات، تمكن الفريق من تتبع مصدر الهجوم إلى دولة أجنبية، ما استلزم تفعيل التعاون مع جهات إنفاذ القانون الدولية، مثل الإنتربول، لتجميد الحسابات المشتبه بها ومتابعة مسار الأموال المحولة.

أسفرت التحقيقات عن تحديد هوية عدد من المشتبه بهم الرئيسيين، واسترجاع جزء من الأموال المنهوبة، كما أُحيل المتورطون إلى القضاء المختص، وقد أظهرت هذه الحالة أهمية سرعة الاستجابة وحفظ الأدلة الرقمية فور وقوع الجريمة، وضرورة تفعيل بروتوكولات التعاون الدولي في الجرائم العابرة للحدود، بالإضافة إلى أهمية تدريب الكوادر البشرية وتحديث أنظمة الحماية الداخلية بشكل مستمر.

المقارنة مع الأبحاث السابقة:

تؤكد هذه الدراسة على الدور الحيوي لأدوات التحقيق الجنائي الرقمي في مكافحة الجريمة الإلكترونية، وهو ما يتوافق مع النتائج التي توصلت إليها الأبحاث السابقة، وتشير النتائج إلى أن هذه الأدوات يمكن أن تعزز القدرة على استرجاع البيانات المحذوفة وتحديد المواقع بنسبة تصل إلى 88% كما هي موضحة في الشكل (1)، بينما تتوزع باقي النسبة على العوامل الآتية:

- نوع البيانات.
- الخبرة.
- تقنيات متغيرة.
- حالة التخزين.
- أساليب التشفير.
- بالإضافة إلى ذلك يُظهر فحص البيانات الوصفية أهميته في جمع أدلة هامة في قضايا الجريمة الإلكترونية، مما يساعد في بناء وصفات مفصلة للأنشطة الإجرامية وأساليبها، على سبيل المثال أظهرت الدراسات أن أدوات مثل EnCase و (FTK (Forensic Toolkit تتمتع بمعدلات عالية لاستعادة البيانات ودقة في تتبع الأنشطة الرقمية، مما يؤكد القيمة العملية لهذه النتائج.



الشكل 1: فعالية أدوات التحقيق الجنائي الرقمي في استعادة البيانات.

أشارت الدراسات السابقة أيضًا إلى الصعوبات في هذا المجال، مثل تباين الأساليب وتطور تقنيات المجرمين السبرانيين، مما قد يؤثر في موثوقية النتائج، لذلك فإنه من الضروري معالجة هذه التحديات لتعزيز فعالية التحقيق الجنائي الرقمي، تساهم هذه الدراسة في إثراء البحوث اللاحقة في هذا المجال من خلال تعزيز النتائج السابقة وتأكيد الحاجة إلى التحسين المستمر لأدوات وأساليب التحليل الجنائي الرقمي، كما أن استخدام التكنولوجيا الناشئة مثل الذكاء الاصطناعي والتعلم الآلي يجعل التحقيقات أكثر دقة وفعالية، مما يساعد قوات إنفاذ القانون في التعامل مع تعقيدات الجريمة الإلكترونية الحديثة، لذلك يمكن القول بإن هذه المقارنة توضح الدور الفعال الذي يؤديه التحليل الجنائي الرقمي

في تنفيذ القانون فيما يخص هذا النوع من الجرائم الحديثة، مما يساعد في تطوير استراتيجيات أفضل لمكافحة الجريمة الإلكترونية في عالمنا الرقمي الذي يتطور بسرعة رهيبة.

تحليل التحديات والنتائج

أظهرت نتائج هذه الدراسة أن أدوات وتقنيات التحقيق الجنائي الرقمي تؤدي دورًا محوريًا في استعادة الأدلة وتحديد هوية الجناة في الجرائم الإلكترونية، إلا أن التطبيق العملي لهذه الأدوات يواجه عدة تحديات رئيسية، من أبرز هذه التحديات التطور المستمر في أساليب الجريمة الإلكترونية، حيث يلجأ المجرمون إلى استخدام تقنيات متقدمة مثل التشفير، وإخفاء الهوية، والعملات الرقمية، مما يصعّب عملية تتبعهم وجمع الأدلة ضدهم.

كما برزت صعوبة تتبع المعاملات المالية عبر العملات المشفرة كأحد أهم العقبات أمام المحققين، حيث أظهرت النتائج أن تقنيات تحليل سلسلة الكتل (Blockchain) فعالة في تتبع 78% من المعاملات المشبوهة، إلا أن هناك معاملات تظل خارج نطاق التتبع بسبب تطور أدوات الإخفاء الرقمي.

من جهة أخرى يواجه المحققون تحديات قانونية تتعلق بقبول الأدلة الرقمية أمام المحاكم، وغياب التشريعات الموحدة في بعض الدول، مما قد يؤدي إلى إعاقة سير العدالة أو ضياع الأدلة.

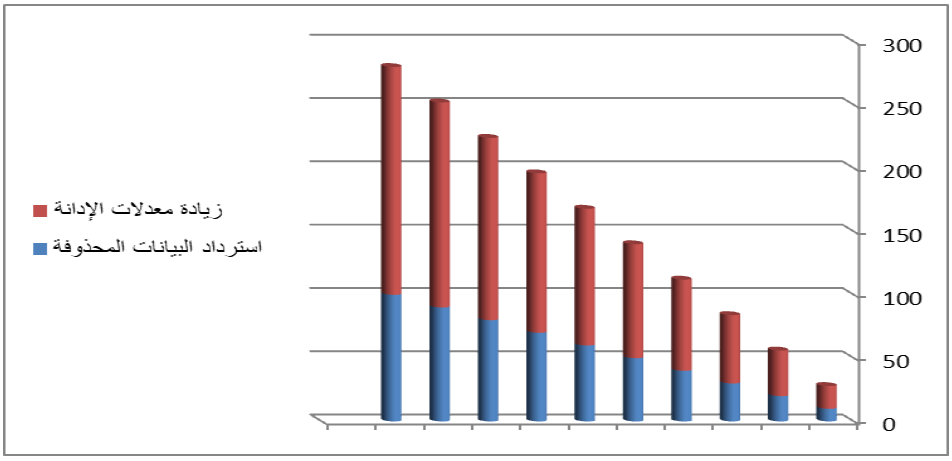
بناءً على ذلك توصي الدراسة بما يأتي:

- الاستثمار المستمر في تطوير أدوات وتقنيات التحقيق الرقمي لمواكبة تطور أساليب الجريمة الإلكترونية.
- تعزيز التعاون الدولي وتبادل المعلومات بين جهات إنفاذ القانون لمواجهة الجرائم العابرة للحدود.
- تحديث التشريعات الوطنية بما يضمن قبول الأدلة الرقمية وتسهيل إجراءات التحقيق.
- توفير برامج تدريبية متخصصة للمحققين الرقميين لرفع كفاءتهم في التعامل مع التحديات التقنية والقانونية الحديثة.

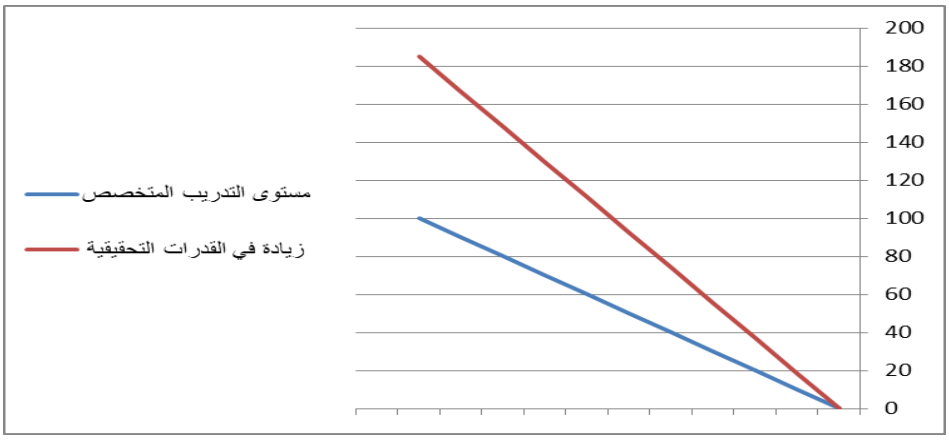
إن معالجة هذه التحديات بشكل منهجي وعملي سيعزز من فعالية التحقيقات الجنائية الرقمية، ويساهم في بناء بيئة رقمية أكثر أمانًا للأفراد والمؤسسات.

التأثيرات العملية والتطبيقية:

التأثيرات العملية لهذه الدراسة كبيرة وغنية في مجال مكافحة الجريمة الإلكترونية، حيث تشير النتائج إلى أن استعادة البيانات المحذوفة بنجاح يمكن أن تزيد من معدلات الإدانة بنسبة تصل إلى 80%، كما هو موضح في الشكل (2)، علاوة على ذلك يمكن أن يعزز التدريب المتخصص لمختصي التحليل الجنائي الرقمي قدراتهم في هذا المجال بنسبة تصل إلى 85%، كما هو موضح في الشكل (3)، بالإضافة إلى ذلك فإن الاستثمار في أبحاث التحقيق الجنائي الرقمي يزيد من كفاءة المحققين ويحسن نتائج التحقيق، مما يعزز في النهاية فعالية السلطة القضائية في مكافحة الجريمة الإلكترونية والحد منها بشكل كبير.



الشكل 2: تأثير استعادة البيانات المحذوفة في معدلات الإدانة.



الشكل 3: العلاقة بين مستوى التدريب المتخصص وزيادة القدرات الفنية للتحقيق.

على سبيل المثال أظهرت إحدى الدراسات أن أدوات التحليل الجنائي الرقمي أدت دورًا حاسمًا في تأكيد الإدانة في قضايا الجريمة الإلكترونية البارزة، مما يبرز تأثيرها بشكل فعلي، كما يمكن للتعاون بين المؤسسات الأكاديمية ووكالات إنفاذ القانون أن يحسن برامج التدريب ومشاركة الموارد، مما يضمن تجهيز خبراء التحليل والتحقيق الجنائي الرقمي للتعامل مع تعقيدات الجريمة الإلكترونية الحديثة.

يتميز هذا البحث عن الدراسات السابقة بعدة جوانب أساسية، فهو يجمع بين التحليل الكمي والنوعي من خلال مراجعة منهجية شاملة للأدبيات وتحليل بيانات كمية مستخلصة من دراسات متعددة، كما يقدم دراسة حالة عملية واقعية من الواقع المحلي، تسلط الضوء على الإجراءات الفعلية للتحقيق الجنائي الرقمي في مواجهة جريمة اختراق إلكتروني، بالإضافة إلى ذلك يقترح البحث إطارًا نظريًا متكاملًا يدمج بين أحدث المفاهيم التقنية والتحديات القانونية، ويوفر توصيات عملية قابلة للتطبيق للممارسين وصناع القرار، بذلك يسهم البحث في سد فجوة معرفية مهمة تتعلق بفعالية الأدوات الرقمية في التحقيقات الجنائية، ويعزز من فهم التحديات المعاصرة والحلول الممكنة في هذا المجال المتسارع التطور.

في خلاصة هذه الدراسة تؤكد على أهمية الاستثمار المستمر في تطوير أدوات ووسائل وأساليب التحقيق الجنائي الرقمي، ليس فقط لتحسين مهارات وأساليب وكفاءة التحقيقات الجنائية الرقمية فحسب، بل أيضًا لتعزيز قوة وإمكانيات وسلطة الهيئات القضائية في مجابهة التحديات المستمرة في مجال الجريمة الإلكترونية.

VII. التوصيات:

تواجه أبحاث التحقيق الجنائي الرقمي العديد من العوائق والتحديات التي تعيق تقدم هذا المجال الحيوي، ومن أبرز هذه التحديات:

- **التباين في الأساليب المستخدمة:**

هناك اختلافات كبيرة في منهجيات التحقيق الرقمي بين الدول والمؤسسات، مما يؤدي إلى عدم توحيد الجهود ويؤثر في موثوقية النتائج، هذا التباين يبرز الحاجة إلى تطوير معايير موحدة دوليًا لتسهيل التعاون وتعزيز فعالية التحقيقات.

- **تطور تقنيات المجرمين:**

يُعد التطور السريع في أساليب وتقنيات الجرائم الإلكترونية أحد أكبر التحديات التي تواجه التحقيقات الجنائية الرقمية، غالبًا ما تتجاوز هذه التقنيات الأدوات الحالية للتحقيق، مما يتطلب من الباحثين والممارسين مواكبة هذه التطورات من خلال الاستثمار المستمر في البحث والتطوير.

- **نقص الأطر القانونية الموحدة:**
تباين القوانين واللوائح المتعلقة بالتحقيقات الجنائية الرقمية عبر الدول يمثل عقبة كبيرة أمام تطبيق نتائج البحث بشكل عالمي، هذا التفاوت القانوني يمكن أن يعيق التعاون الدولي ويقلل من فعالية التحقيقات الرقمية في سياقات متعددة.
- **التحديات القانونية والتشريعية:**
على الرغم من التقدم التقني الكبير في مجال التحقيق الجنائي الرقمي، لا تزال هناك تحديات قانونية وتشريعية تعيق الاستفادة المثلى من الأدلة الرقمية، من أبرز هذه التحديات مسألة قبول الأدلة الرقمية أمام المحاكم، حيث تختلف معايير قبولها من دولة إلى أخرى، ويشترط بعضها إجراءات صارمة لضمان سلامة الأدلة وسلسلة الحيازة (Chain of Custody). كما أن غياب أو قصور التشريعات الوطنية المتعلقة بالجرائم الإلكترونية يؤدي أحيانًا إلى صعوبات في توجيه الاتهام أو تنفيذ الأحكام، خاصة في القضايا التي تتعلق بجرائم عابرة للحدود، ويزداد الأمر تعقيدًا مع الحاجة إلى التعاون الدولي وتبادل المعلومات بين الجهات المختصة، ما يتطلب أطرًا قانونية واضحة واتفاقيات ثنائية أو متعددة الأطراف، كذلك تبرز إشكالية حماية الخصوصية والحقوق الرقمية للأفراد أثناء عمليات جمع وتحليل الأدلة، حيث يجب الموازنة بين متطلبات التحقيق وحقوق الأفراد في الخصوصية، مما يستدعي وجود ضوابط قانونية دقيقة وإشراف قضائي فعال.
- **الحاجة إلى دمج التقنيات الناشئة:**
على الرغم من التقدم الكبير في أدوات التحقيق الجنائي الرقمي، فإن دمج التقنيات الناشئة مثل الذكاء الاصطناعي والتعلم الآلي يمكن أن يعزز دقة وكفاءة عمليات التحليل الجنائي الرقمي، ويوفر حلولًا مبتكرة للتعامل مع تعقيدات الجرائم الإلكترونية الحديثة.
- **التحديات المتعلقة بالتدريب والتعاون:**
هناك حاجة ملحة لتعزيز التدريب المستمر للمحققين الرقميين لضمان مواكبتهم للتطورات التقنية المتسارعة، بالإضافة إلى ذلك يمكن للتعاون بين المؤسسات الأكاديمية وأجهزة إنفاذ القانون أن يساهم في تحسين الموارد المتاحة وتطوير مهارات الخبراء في هذا المجال.

للتغلب على هذه التحديات من الضروري إجراء المزيد من الأبحاث لضمان أن تبقى أدوات وتقنيات التحقيق الجنائي الرقمي فعالة ومواكبة للتطورات التقنية والقانونية، كما يجب تعزيز التعاون الدولي لتطوير أطر قانونية موحدة ودعم الابتكار في هذا المجال. بناءً عليه توصي الدراسة بضرورة تحديث التشريعات الوطنية بما يتواكب مع التطورات التقنية، وتوحيد معايير قبول الأدلة الرقمية، وتعزيز التعاون الدولي في مجال مكافحة الجرائم الإلكترونية، مع التأكيد على احترام الحقوق الأساسية للأفراد أثناء التحقيقات.

VIII. الخاتمة:

تؤكد هذه الدراسة التي ارتكزت علي تقييم مجموعة من الدراسات السابقة وبيانات بعض المواقع الإلكترونية الرسمية المختصة في هذا المجال على الأهمية الحيوية للتحقيقات الجنائية الرقمية في مكافحة الجريمة الإلكترونية، حيث أثبتت النتائج فعالية الأدوات والتقنيات المتقدمة في استعادة الأدلة وتحديد الجناة، مما يبرز دورها في تعزيز قدرات استرجاع البيانات وتحسين معدلات الإدانة، كما تشير الدراسة إلى الحاجة الملحة للاستمرار في الاستثمار في أبحاث التحقيق الجنائي الرقمي وإقرار أطر قانونية قوية لتعزيز قدرات المحققين لمواجهة التهديدات السيبرانية المتطورة باستمرار، ونظرًا للجوء المجرمين السيبرانيين لأساليب أكثر تطورًا فإن على خبراء التحقيق الجنائي الرقمي أن يكونوا على درجة عالية من المهارة والتدريب والمعرفة ليكونوا في طليعة خبراء هذا المجال، من خلال الاستفادة من تقنيات مثل الذكاء الاصطناعي والتعلم الآلي، والتي يمكن أن تحسن بشكل كبير دقة وكفاءة التحقيقات الرقمية.

علاوة على ذلك فإن تعزيز التعاون بين مختلف الجهات المعنية بما في ذلك المؤسسات الأكاديمية ووكالات إنفاذ القانون والقطاع الخاص يُعد أمرًا ضروريًا، من خلال مشاركة المعرفة والموارد، يمكن لهذه الجهات تطوير برامج تدريبية ومنهجيات أكثر فعالية، مما يُمكّن خبراء التحليل الجنائي الرقمي من التعامل بفعالية مع التحديات التي يفرضها المجرمين السيبرانيين.

الخلاصة تسلط هذه الدراسة الضوء على ضرورة اتباع نهج استباقي وتعاوني في مجال التحليل الجنائي الرقمي، وهو أمر بالغ الأهمية لحماية الأفراد والمؤسسات والمجتمع ككل من التهديدات التي تفرضها الجريمة الإلكترونية، من خلال تنفيذ هذه الاستراتيجيات، يمكننا ضمان استجابة شاملة وفعالة للتحديات المتزايدة في هذا المجال الحيوي.

IX. المراجع

[1] ر. يوسف، وز. أ. أوثمان، "التحقيق الجنائي الرقمي: استعراض القضايا والتحديات"، *مجلة الهندسة والعلوم التطبيقية*، مج. 14، ع. 4، ص. 1228-1234، 2019م.

[2] هـ. تشونغ، ج. بارك، س. لي، وج. كانغ، "التحقيق في الطب الشرعي الرقمي لخدمات التخزين السحابي"، *التحقيق الرقمي*، مج. 19، ص. 98-111، 2017م.

[3] س. أ. زارين، وف. أولا، "التحديات والقضايا في الطب الشرعي الرقمي"، *IEEE Access*، مج. 8، ص. 47559-47580، 2020م.

[4] أ. العنزي، ر. س. حسين، وف. القحطاني، "أدوات وتقنيات الطب الشرعي الرقمي: مراجعة منهجية"، *IEEE Access*، مج. 9، ص. 56635-56658، 2021م.

[5] أ. أزفار، ك. ك. ر. تشو، ول. ليو، "إطار عمل سهل النشر للطب الشرعي المحمول"، *مجلة العلوم الجنائية*، مج. 61، ع. 6، ص. 1498-1503، 2016م.

[6] س. إقبال، س. خالد، وأ. خان، "التحقيق الجنائي الرقمي: مراجعة، تصنيف، وتحديات مفتوحة"، *IEEE Access*، مج. 7، ص. 70332-70365، 2018م.

[7] س. سلطاني، وس. أ. هـ. سينو، "مسح حول اتجاهات الطب الشرعي الرقمي"، *المجلة الدولية للهندسة الكهربائية والمعلومات*، مج. 6، ع. 2، ص. 61-74، 2017م.

[8] ف. روسيف، ول. مارزيالي، "الطب الشرعي القائم على المحتوى"، *التحقيق الرقمي*، مج. 29، ص. 109-118، 2019م.

[9] إ. باجلي، وف. برايتنجر، "مصادر البيانات لتعزيز الطب الشرعي الرقمي: ماذا يحمل المستقبل"، *العلوم الجنائية الدولية: التحقيق الرقمي*، مج. 32، ص. 200901، 2020م.

[10] س. راني، وب. سوري، "الطب الشرعي الرقمي: الاتجاهات والفرص الناشئة"، *IEEE Potentials*، مج. 37، ع. 5، ص. 15-19، 2018م.

[11] م. س. نمور، *أصول الإجراءات الجزائية، شرح لقانون أصول المحاكمات الجزائية*، ط2، عمان: دار الثقافة للنشر والتوزيع، 2011م، ص. 326.

Black Henna Dyes Containing paraphenylenediamine: Assessing the Risks of Exposure on Pruritus, Hematological and Biochemical Parameters

A. A. S. Salah [1, *]
S. M. Q. Mofleh [2]

Received: 01/07/2025
Revised: 27/07/2025
Accepted: 28/07/2025

[1]Department of Biology &Chemistry, Faculty of Toor Al-Baha, Lahej University, Lahej, Yemen.
[2]Department of Chemistry, Faculty of Education, Aden University, Aden, Yemen.
*Corresponding Author's Email: aasalehsalah.74@gmail.com

https://doi.org/10.20428/jst.v30i9.3129

Black Henna Dyes Containing Paraphenylenediamine: Assessing the Risks of Exposure on Pruritus, Hematological and Biochemical Parameters

A. A. S. Salah
Department of Biology & Chemistry, Faculty of Toor Al-Baha, Lahej University
Lahej, Yemen
aasalehsalah.74@gmail.com

S. M. Qasem Mofleh
Department of Chemistry, Faculty of Education, Aden University
Aden, Yemen

Abstract— This study investigates the health risks associated with black henna dyes, specifically focusing on the presence and concentration of Paraphenylenediamine (PPD) in commercial products. Henna, historically used as a dye, has evolved to include chemically mixed variants that can cause allergic reactions skin. The research involved analyzing samples from 60 regular users of black henna and 60 control participants who had never used it. In our study most symptoms itching and the dark color urine, The study found hematological indices and changes in kidney function, liver function than group control. The hematological and biomarker may be indicated to toxicity among users, that is probably the health hazards linked to exposure to black henna dyes. The findings underscore the need for awareness regarding the risks associated with these products.

Keywords— Black Henna dyes; Paraphenylenediamine (PPD); Liver Enzymes; Kidney toxicity; Allergic reactions.

I. INTRODUCTION

In a culture that is fixated on beauty, people are drawn to improving their looks in order to boost their confidence and develop wonderful personalities. However, a lot of these cosmetics, which are supposed to help us feel healthier and more attractive, have a dark side because they include dangerous chemicals and harmful ingredients in excess [1]. Hair care products are a major source of lifestyle related chemical exposure in the general population, especially since hair dyes use is becoming an integral part of modern culture [2], these products serve various cosmetic purposes, such as covering gray hair, altering hair color, and enhancing color retention, which has contributed to the growth of the hair coloring market, particularly with an aging population [3]. Henna, scientifically known as Lawsonia inermis L., belongs to the Lythraceae family, (Fig.1) flowering plant it's also known Lawsone is commercially cultivated in several countries such as Morocco, Sudan, India, Pakistan, and Yemen. In these countries the henna leaves are ground into a paste with water or oil and using for body art and to color skin and hair during social events, particularly weddings. The compound 2-hydroxy-1,4-naphthoquinone is active dye responsible for henna's red color (Fig.1) [4,5].

Fig.1. Leaves of Lawsonia inermis

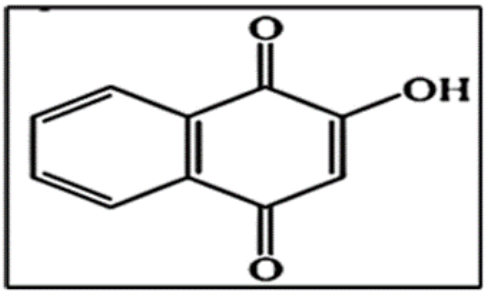


Fig. 2. Structure of Lawsone

Para-phenylenediamine (PPD), an aromatic amine compound and derivative of aniline (1,4-diaminobenzene, C6H8N2), has a molecular weight of 108.15 g/mol (Fig. 3). PPD typically appears as white crystals that oxidize in air, changing color from red to brown and eventually to black (Fig. 4). It is commonly used in various industries for dyeing fabrics and furs, as well as in the production of photographic developers, appliances, wheels, caoutchouc, cosmetics, and plastics [6-8].

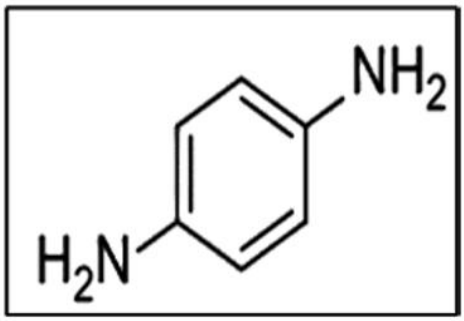


Fig 3: Structure of paraphenylenediamine

Fig. 4: Crystal form of PPD.

PPD is widely utilized as a chemical an ingredient of oxidative hair coloring products and black henna dyes. Currently, PPD is present in more than 1000 hair dye formulations marketed all over the world [9]. Moreover, the natural red henna is often combined with PPD to create black henna, commonly used for temporary black henna tattoos, which acts as: a chemical dye, enhancing the staining effect when mixed with henna paste [10], accelerates the dyeing and drying processes, intensifies the color, improves the design, and extends the tattoo's longevity [11]. PPD is prevalent in various hair dyes, with inexpensive formulations such as tancho and branded products like L'Oreal and Garnier typically contain lower percentages (2-6%) and from 70-90 % in Black Stone Hair Dye (BSHD) that is containing a high concentration of PPD, which are used for giving black color to hair [12].

In contrast, BSHD is available in North Africa and the Middle East, the traditional name of a commercial black hair dye and widely used because it is freely available and less expensive than pharmaceutical hair dye preparation [13]. It is also usually added in henna Lawsonia after crushing, which applied as popular custom to make black hair dye also to lessens the amount of henna needed, intensifies its color, and hastens the staining process [14]. In contra, the chemical reaction between a coupler agent (e.g., resorcinol) and the dye precursor (PPD) under oxidative conditions forms a colorful compound but may also lead to the formation of Bandrowski's base, that is consider a carcinogenic substance [15]. Recent studies have linked chemicals commonly found in permanent hair dyes of loss, contact dermatitis and toxicity from PPD-containing hair dyes both red and black henna products has been reported in Asia, Africa, and the Middle East, highlighting its potential as a means of suicide due to its lethality [12, 16, 17]. The studies have been dealt with several types of PPD poisoning, which can the local application cause skin allergies, lacrimation, cardiotoxicity, urinary toxicity, both allergic contact dermatitis and irritant contact [18]. Moreover, exposure to PPD foam, during its production or application in manufacturing of various products, may result in health problems affecting airway and skin, in addition, the direct toxic effects occurs following ingestion of PPD after some hours that has a more systemic damage has taken place (renal, lover), due to the absorption and distribution of the toxic metabolites throughout the body and death [19]. The first defined a strong allergen causing PPD in 1939 by the standard antigen group of the North American and European epidermal patch tests, as a result of PPD allergy symptoms, many strict legislation were put in in its use. In 1976 according the European Cosmetic Directive regulation the maximum permissible concentration of PPD in hair coloring was 6% [20] then in 2009 modified to 2% for direct application with the oxidizing agent on the hair dyes [21] and due to its health risks has been banned in countries like France, Germany, and Sweden [7]. On the other hand, less toxic of PPD poisoning in developed countries the maximum percentage does not exceed 2% per 100 ml of color solution but PPD concentrations in thiscountries may be reach to 90% due to a lack of stringent regulations. On the contrary, in the USA, there is no regulation of the limits for concentrations of PPD in hair dyes [21] as a result may lead to the addition of high PPD concentrations due to the lack of sufficient legal restriction on the use of PPD concentrations [22]. The authors studied four Black henna samples that, which considered a more available, very cheap price and commonly used, from shops selling in Aden Governorate. The result confirmed the presence of PPD in the four samples and according to the following concentrations of PPD in Black Henna samples: 1.9, 6.98, 14.33, and 70.35% for Brand-2, Brand-1, Brand-3 and Black Stone respectively. The PPD levels in the last three samples higher than permissible levels in black henna dyes and proven that the black stone dye contains large amount of PPD [23]. After proven the presence of PPD in the four samples, this search aims to study evaluate the toxic effects of these dyes on human hematological and biomarkers on blood cells and renal and liver functions, and contributing to improved consumer safety and environmental protection with facilitate better monitoring of Black Henna samples.

II. 2. Materials and Methods

A. Materials

The Chemicals, Equipment and Instruments: All chemicals used in the study were of analytical grade and highly pure, and it is also use micropipit, test tubes anticoagulant samples, syringes, gloves for collected blood samples. The Instruments used to Blood samples analysis: Siemens Healthineers, high-volume hematology analyzer, the ADVIA 2120i Hematology System and Biochemical Roche Hittachi Diagnostics CobasC 311 analyzer.

B. Samples

Four samples Black Henna, more available and commonly used were collected from shops selling different areas randomly in Aden Governorate, and these samples were

labeled as Brand-1, Brand-2, Brand-3, and Black Stone Dye instead of their commercial names.

C. Study Period

This study was done in the period September 2023 to January 2024 in Aden Governorate.

D. Study Population.

The study population was comprised adult Yemeni women that use Black henna dyes in Aden Governorate, Republic of Yemen. This prospective cross sectional analytical study involved 120 adult women who were randomly selected and used black henna dyes in beauty salons. 120 participants were selected after application of exclusion criteria of participation. All participants were confirmed to have no medical history of kidney, liver, diabetes, or hypertension conditions, and not taking any vitamin supplements.

A day before sample collection, volunteers were stop taking any food after 9 pm, where the blood samples collected after 24 hours from applying the dyes to the scalp, then placing into tubes and marked by code numbers identify the volunteer's personal identity, then the samples were transferred to the National Center for Central Public Health laboratories

The participants were divided into three groups:

a. Sixty subjects as the control group (without any history of black henna use).
b. Thirty subjects who used various dye brands (Brand 1, Brand 2, Brand 3).
c. Thirty subjects who used black stone dye.

E. Biochemical Measurements

Two parameters analysis used to measure effects hair dyes on volunteers:

a. The hematological analysis:

the blood samples were performed to assess the following parameters: Hemoglobin (HGB), Red Blood Cell (RBC), Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Hemoglobin Concentration (MCHC), White Blood Cell (WBC), Blood Platelet (PLT), Lymphocyte, Neutrophil were conducted using the using ADVIA 2120i (Siemens Healthineers Germany).

b. The Biochemical analysis: the blood samples were collected to assess the renal function tests, measuring blood urea and creatinine levels and to liver function tests performed to assess total bilirubin, total protein, albumin, Alanine Aminotransferase (ALT/GPT), Aspartate Aminotransferase (AST/GOT), and Alkaline Phosphatase (ALP) using by the Roche/Hitachi Cobas 311 analyzer.

F. Ethical Consideration

Ethical approval for the study was obtained from the Ethical Scientific Research Committee of the University of Aden (Rec-57-2024).

G. Statistical Processing of Results

Statistical analysis was performed using Origin 8, Excel, and the Statistical Package for Social Science (SPSS) version 20 for Windows. Continuous and categorical variables were tested for significance using T-tests and One-way ANOVA, with a p-value of < 0.05 considered statistically significant.

III. Results

The study effects of hair dyes on volunteers used by biochemical and hematological indices. It was analyzed statistically and the following statistical analysis results were obtained:

A. Hematological assays:

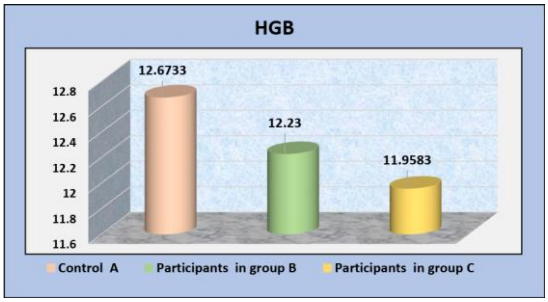
The results of the statistical analysis revealed are a decrease in mean values of Hemoglobin (Hb), and Mean Corpuscular Hemoglobin (MCH) in the blood with a statistical of (P=0.013), (P=0.016) in groups B and C compared with the control respectively, but these result of the Hb (12.23 and 11.96) and MCH (28.52 and 28.48), in the B and C groups respectively, were slightly lower than the control group (12.67 and 28.53) respectively. In Red blood cell counts (RBCs) decreased in the value of (P=0.011) in group B and C compared to the control group and the average RBCs was low in the sera of participants in the B and C groups by in mean (4.30 and 4.23, respectively) compared to control group (5.31). While decreased in the value of mean corpuscular volume (MCV) in the blood with a statistical of (P=0.019) in group B and C compared to the control group, the MCV was low in the sera of participants in the B and C groups by in mean (85.80 and 85.63, respectively) compared to the average rate in the sera of participants control group (86.17). The result in Mean Corpuscular Hemoglobin Counts (MCHC) was different in the statistical analysis revealed no statistically significant was (P=0.239) of MCHC in groups B and C compared to the control group where the MCHC was equal in the B and C groups (33.29 and 33.27 respectively) but lower than the average control group (42.84). The results in the value of White blood cell counts (WBCs) revealed in a decrease in a statistical of (P=0.014) in group B and C compared to the control group, but differences the results WBCs between B and C groups (6.12 and 5.92, respectively) while lower in compared to the control group (6.84). In the Blood Platelets (PLT) the results that was no statistically significant (P=0.046) in groups B and C compared to the control group but the result of PLT in groups B (297.37) was slightly higher than control group (291.63) and significantly lower in C group (257.60) than control group. The results of the statistical analysis revealed was no statistically significant was (P=0.256) in the value of Lymphocytes (LYMP) in the blood in group B and C compared to the control group, the results related of LYMP was equal in the B and C groups (43.65 and 43.15 respectively) but lower than the control group (40.47). In the last results of Neutrophil (NEUT) in the blood the statistical analysis revealed was decreased in the value statistical of (P=0.012)) in group B and C compared to the control group, which differences the value NEUT low in C groups from B groups (41.79 and 43.23, respectively) and the two group lower than control group (46.19).

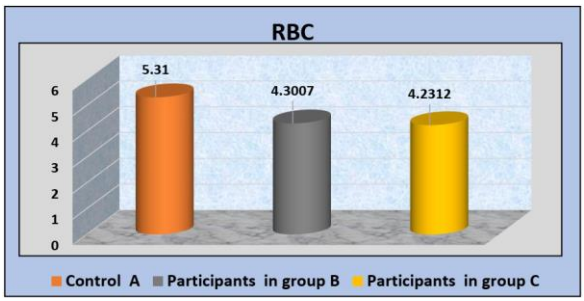
All results of hematological analysis are summarized in Table 1 and illustrated in Fig.5.

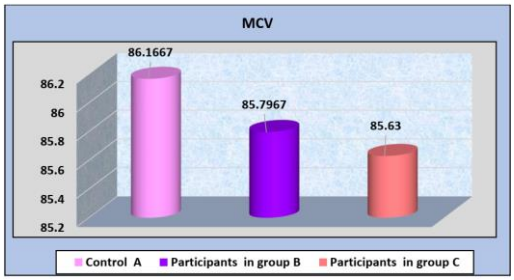
Table (1): Hematological parameters among study groups.

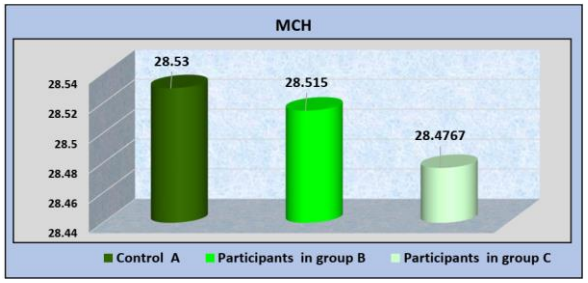
Hematological parameters	Group A Control (mean± SD)	Group B (mean± SD)	Group C (mean± SD)	p-value
Hb	12.67 ± 1.75	12.23 ± 1.75	11.95 ± 1.23	0.013*
RBC	5.31 ± 4.69	4.30 ± 0.52	4.23 ± 0.43	0.011*
MCV	86.17 ± 9.00	85.80 ± 10.73	85.63 ± 9.68	0.019*
MCH	28.53 ± 3.37	28.52 ± 3.84	28.48 ± 2.90	0.016*
MCHC	42.84 ± 5.35	33.29 ± 0.60	33.27 ± 2.13	0.239*
WBCs	6.84 ± 5.30	6.12 ± 1.84	5.92 ± 1.67	0.014*
PLT	291.63 ± 61.19	297.37 ± 91.433	257.60 ± 54.92	0.046*
Lymphocyte	40.47 ± 8.42	43.65 ± 11.83	43.15 ± 9.88	0.052*
Neutrophil	46.19 ± 11.29	43.23 ± 13.46	41.79 ±11.75	0.012*

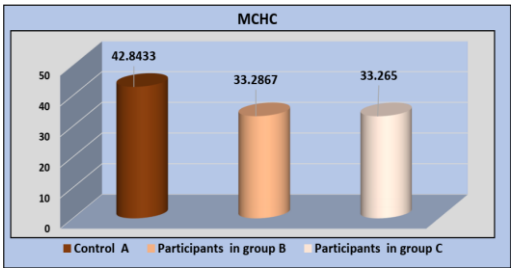
*P≤0.05

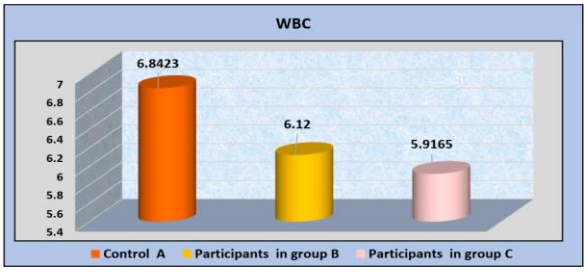


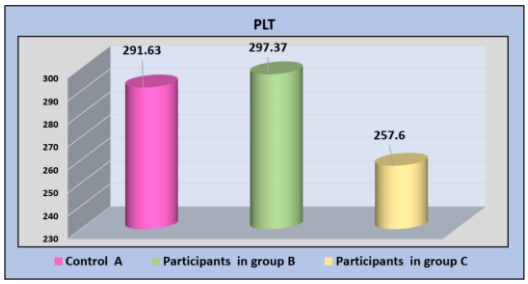


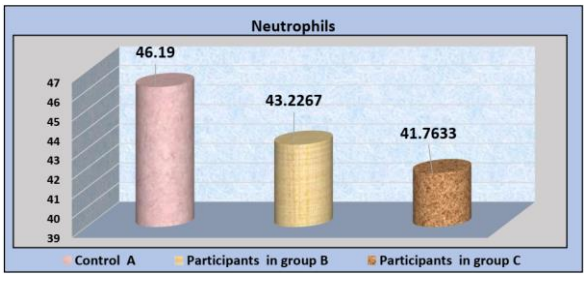












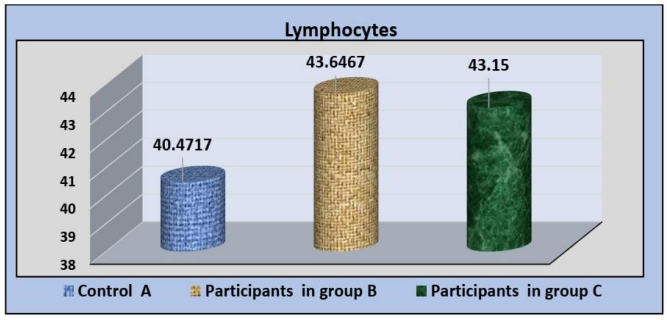


Fig .5: The Hematological parameters tests for groups B and C compared with control A,

B. The Biochemical analysis

a. Liver Function Tests

The liver function tests assessed included total protein and albumin levels, which exhibited a significant decrease ($P \leq 0.05$) in the tested groups B and C. Conversely, plasma enzymes alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), and total bilirubin levels showed a significant increase ($P \leq 0.05$) in groups B and C compared to the control group. These results are summarized in Table 2 and illustrated in Fig. 6.

Table 3: The Liver function tests among study groups.

Biochemical parameters	Group Acontrol (mean± SD)	Group B (mean± SD)	Group C (mean± SD)	p-value
T.Protein	7.33 ± 0.69	6.83 ± 0.69	6.26 ± 0.89	0.015*
Albumin	4.61 ± 0.31	4.04± 0.55	3.36 ± 0.87	0.028*
T.Bilirubin	0.29 ± 0.15	0.740 ± 1.62	1.26 ± 1.68	0.413
AST	17.62 ± 4.66	28.56 ± 16.48	35.57 ± 31.44	0.049*
ALT	11.73 ± 6.48	14.67 ± 10.17	32.70 ± 38.20	0.035*
ALP	78.07 ± 23.51	129.07 ± 90.91	155.23 ± 114.98	0.020*

*P≤0.05

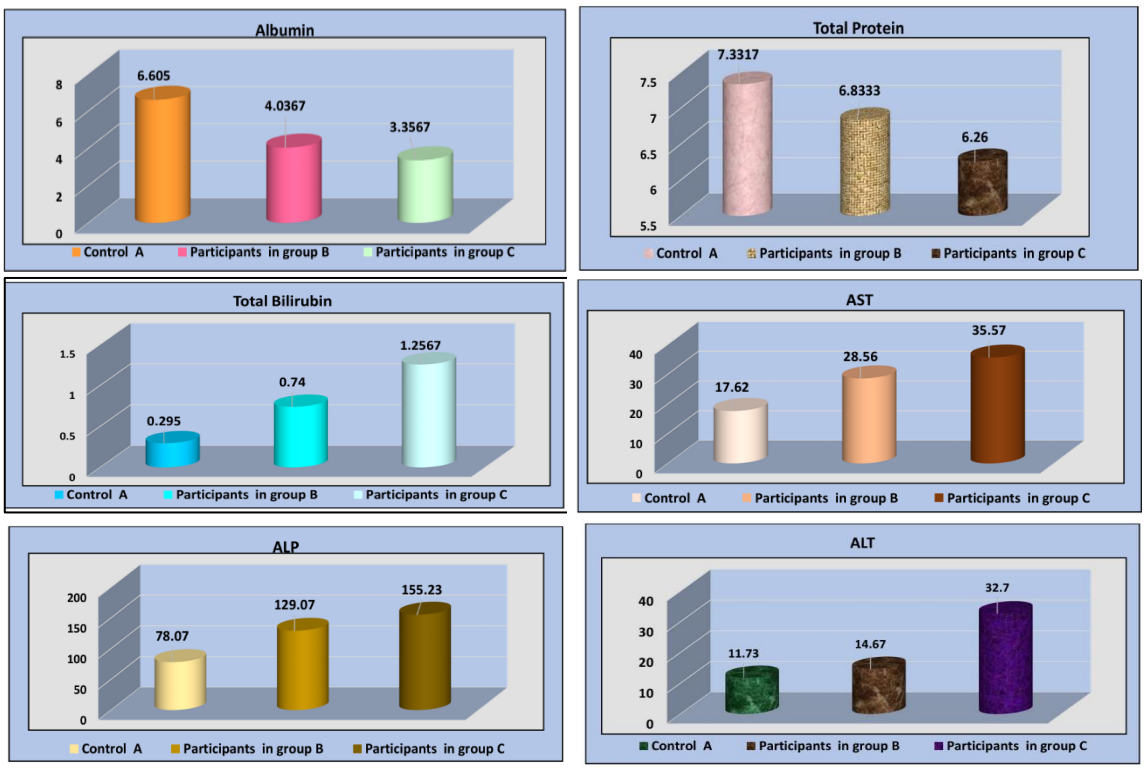


Fig .6: illustrated the liver function tests.

b. Kidney Function Tests (KFT)

The key indices for assessing kidney function include serum urea and serum creatinine levels. The results presented in Table 3 and Fig.7 indicate that women exposed to para-phenylenediamine (PPD) over an extended period are at significant risk for kidney-related issues. Specifically, the measurements for blood urea and creatinine showed a significant increase ($P \leq 0.05$) in the tested groups compared to the control group, highlighting potential kidney dysfunction in these individuals.

Table 3: Kidney function tests among study groups.

Biochemical parameters	Group A (mean± SD)	Group B (mean± SD)	Group C (mean± SD)	p-value (mean± SD)
Urea	19.567 ± 16.34	31.30 ± 21.16	103.60 ± 57.43	0.012*
Creatinine	0.57 ± 0.09	0.85 ± 0.34	2.24 ± 1.18	0.014*

*P≤0.05

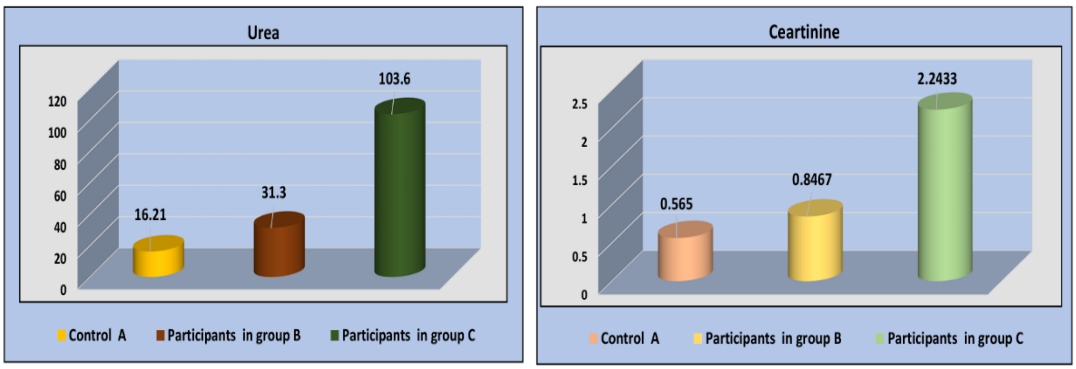


Fig .7: Kidney function tests

C. Symptoms appeared after application black henna:

In Table.6 and fig.8 most symptoms itching 13.3% in group B and 40 % in of group C, and the skin darkening was complained by 6.7 % for group B and 10 % in group C, while the darkening of urine no appeared in group B and 13.3 % in of group C. in contrast, 86.7% from group B and 60 % group C did not have any reaction on skin after application of hair dye, but 13.3 % group B and 40 % group C they had itch on the skin immediately after application of hair dye complained by 23.3 % of group B and 60 % of group C.

Table 4: Distribution of the study subjects according to previous history of reaction after dye application (n=60).

Reaction after dye use	Group B %	Group C
Darkening of urine	**0.0%**	**13.3%**
Itching	**13.3%**	**40.0%**
No effect	**76.7%**	**40.0%**
Pain/darkening of skin	**6.7%**	**10.0%**

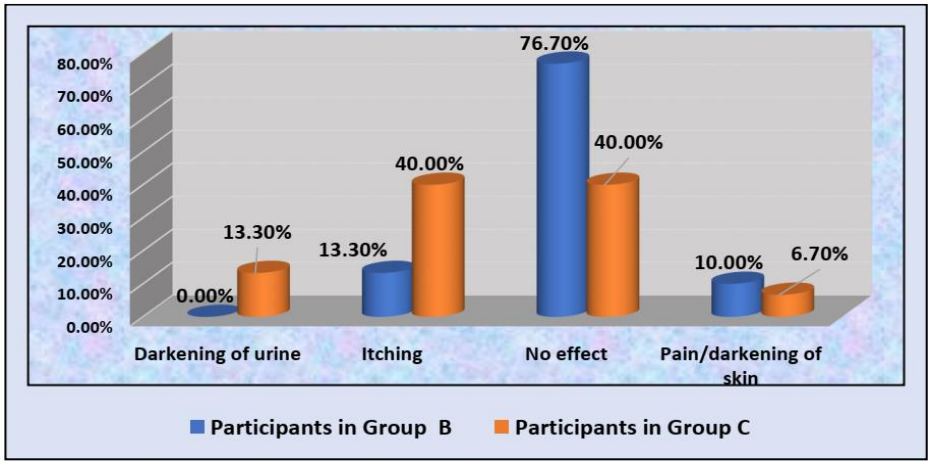


Fig.8: Distribution of the study subjects according to previous history of reaction after dye application (n=60).

IV. Discussion

A. The Hematological Assays

The skin acts as a protective barrier, preventing external substances from disrupting the body's intricate internal processes but applied some compounds such hair dye can penetrate the skin layer without resistance. therefore, the human exposure to PPD during hair dyeing process, regular hairdressers, handling hair dye, industrial workers etc. after absorption during layer skin PPD can reach systemic circulation, where some studies reported the causal association of hair dye usage may be cases: leukaemia, bladder cancer risk and a strong allergen [24].

In this study the effect of the exposures to PPD contented in black henna dye was significantly correlated with blood parameters, the tested groups B, and C showed significant decrease ($P \leq 0.05$), mainly Hb, RBCs compared to control group. The MCV was found to be elevated in the Group B and C, but no statistical significance was observed. While that of MCH, and MCHC measurements of tested group B did not show significant difference ($P \geq 0$ .05), whereas the tested C showed significant decrease ($P \leq 0.05$), while that of WBCs in the blood of tested C showed significant decrease ($P \leq 0.05$) compared to control group, with little decrease in group B. On the other hand, PLT in the blood of tested group B but no statistical significance was observed, whereas the tested C showed significant decrease ($P \leq 0.05$). while that of mean Neutrophil percentages in the blood of tested B, and C little decrease ($P \leq 0.05$) compared to control group. On the other hand, the lymphocytes percentages in the blood of tested C showed significant increase ($P \geq 0.05$) compared to control group, with little decrease in group B. Despite the significant increases ($P \leq 0.05$) in the percentage of neutrophils count, the lymphocytes showed significant ($P \leq 0.05$). We observed from all results the group C more consider effect of the dye from group B counts and compared to the control group.

One of the results obtained in this study is increased in lymphocytes percentages, this result identical with Sieben et al., (2002) where illustrated that is both PPD and its auto-oxidation lead to motivate lymphocyte diffusion which may cause of inflammation or allergies reactions [25]. Our result showed decreasing in WBC this corroborated with Zhang et al., (2023) that is decease WBC probably for exposed to higher PPD concentration [20]. The decreasing in in RBC count, Hb and WBC that reported by [24,26]. The effect of hair dye containing PPD on decrease Hb, RBCs, MCH, MCV and PLT that is possible to associated with macrocytic anemia and iron deficiency, while the significant increase in lymphocyte and decreased in neutrophil and in WBCs indicates to the effect of PPD lead to may injurious the immune system and reduces its efficiency [27]. From hematological assays for all groups group B and C with compare control group shown that the more effect is group C from B which applied black stone dye that the high concentration of PPD and this result compatible with information above. Therefore, the continue application of henna mixed with PPD may be led to severe hemolytic anemia and posing a risk of systemic toxicity [26].

B. The Biochemical assay

The traditional henna used as form a paste henna leaf for cosmetic purposes like: colouring skin, hair, and hand or body, this natural henna appears to be non-toxic effects and biosafe. But before being used, the dye is frequently combined with PPD. While natural henna reactions are rare, allergic contact dermatitis is more likely to occur when PPD is added to natural henna. The idea that drugs applied to the skin can be readily absorbed through the dermal blood vessels and may have an impact on the tissue of the liver and kidneys is supported by several research [28]. Therefore, biomarkers are important tools for measurement, knowing, evaluating the toxicity and adverse health effects that reflect of the health disease as a result of exposure to hair dyes. When exposed to hair dye, chemicals or their biological metabolisms are absorbed and distributed throughout the organism. The toxicity of these substances is assessed by urine, saliva, blood, serum, etc., which reflect the severity of exposure, risks and adverse effects on the skin, kidney, bladder, liver and other organs along with the immune system [29].

a. Liver Function Test:

Our study observed elevated liver enzymes for participants (AST, ALT, ALP) and the increased in Bilirubin with decreases in total protein and albumin, after applied black henna hair that containing of PPD, this increases in liver enzymes and albumin and deceased in total protein and albumin indicating potential poisoning liver or liver dysfunction due to exposure for black henna hair. This study indicated that PPD significantly increased liver enzymes while decreasing total protein and albumin levels, consistent with previous research [30]. it is supported this finding, noting similar higher enzyme behavior in cases of PPD exposure to the cause's hepatotoxicity due to muscle damage associated with PPD toxicity, this results consistent with [31]. The decreased in total protein synthesis by hepatic cells that is reflects the synthetic function of the liver, elevated liver enzymes and in increased total bilirubin due to a breakdown of red blood cells indicate sensitivity and toxicity to liver tissue, even following brief exposure of PPD may lead to hepatotoxicity [32]. El-Amin et al., (2014) when their administration of PPD to rats revealed a significant increase in liver enzymes GOT, GPT, and ALP, and a significant decrease in the total protein, albumin associated with the increase of the commercial hair dye dosage [27]. Workers exposed to high PPD concentrations exhibited increased AST, glutamyl transpeptidase, and total bilirubin levels, while albumin and total protein levels decreased, indicating potential liver dysfunction [33].

b. Kidney function tests

Our result indicate that use of black henna dyes adversely affects urea and creatinine levels, with increases observed in two groups compared to the control group. furthermore, the group C higher levels in urea and creatinine then group B these influences may be due to correlating the concentration of PPD in hair dyes on urea and creatinine, therefore the PPD's toxicity on kidney tissues, leading to reduced renal function and potential renal failure [33,34]. Our result consistent with Amiri (2023), was study effect of mixed henna with PPD users from the relationship between serum creatinine measurement and kidney function, where the study was revealed a high creatinine in blood with significant effect on the of kidney function which indicated possibility to be caused the developing disease Henna-induced pigment nephropathy by consuming hair dye [35]. Moreover, very high serum levels of creatinine and urea making PPD the only cause of nephrotoxicity, this occurs due to the aromatic

structure of PPD which makes it readily absorbable and concentrated in the tubules which cause of acute tubular necrosis in PPD poisoning is due to concentration of PPD in renal tubules [36]. Therefore, the applied of black henna hair dye containing on PPD for long times probably to related with acute or chronic renal disease, which results to cause arrhythmia, myocarditis, and ventricular tachycardia that may causes death [19].

C. Symptoms appeared after application black henna

Patients who are allergic to PPD are usually for the people who have no experience using hair dyes at home, and usually take longer to apply the dye to the scalp, and the risk of allergic contact dermatitis may appear within for as 5-30 minutes, the symptoms are usually very dramatic, and they take the form of acute eczema on the scalp, face, neck and sometime cases of hair loss [21]. In our study In Table.6 and fig.8 most symptoms itching 13.3% in group B and 40 % in of group C, and the skin darkening was complained by 6.7 % for group B and 10 % in group C, while the darkening of urine no appeared in group B and 13.3 % in of group C. in contrast, 86.7% from group B and 60 % group C did not have any reaction on skin after application of hair dye, but 13.3 % group B and 40 % group C they had itch on the skin immediately after application of hair dye that is effect may be due to allergic contact dermatitis, this result same results reported by Han et al., (2018), where study to identify the clinical characteristics of hair dye contact allergy and to assess the relationships between hair dye contact allergy, exposure time to PPD and PPD positivity. Hair dye allergy was appeared on most patients 80% where the diagnosis associated skin diseases, involved Allergic contact dermatitis area and pruritus was the most common symptom; erythematous macules and patches and with the face and non-specific eczema and urticaria [2]. Similar result to irritant reaction and change morphology that looks on skin with Young, et al., (2019), that was found 88% subjects study reacted to PPD at 1.0%, while 69% reacted to 0.32% PPD. Therefore, they recommended to reduce the risk of active sensitization and elicitation or decreased effect of contact allergy may be lowering the patch test concentration of PPD from 1.0% to 0.32%, it also found the elicitation and sensitization dependent on the concentration on the dose per unit area of allergen delivered to the skin and exposure time [37]. This was consistent with the studies of Asghar et al,. (2022) The dark color urine (82.5%) with declining urine output and inclining serum creatinine, it was suggestive of acute kidney injury which was 22.3% in this study. Renal replacement therapy in the form of hemodialysis was done in 4.8% of our patients. where rhabdomyolysis was evident in (74.5%) patients, acute kidney injury in 30% out of which 16 patients required renal replacement therapy [18].

V. Conclusions

- Our results the changes in the biological parameters indicate possible hepatic toxicity, pointed out by significant a decrease in the protein and albumin and increases in liver enzymes (GOT, GPT, ALP) and, in groups C more than groups B compared with the control groups associated with the increase of PPD the hair dye.
- Our result underscores the use of black henna dyes adversely affects in urea and creatinine levels, as the increased urea and creatinine are observed in groups B and C compared to the control group, which may be due to the of PPD concentration in hair dyes therefore the PPD's toxicity on kidney tissues, leading to reduced renal function and potential renal failure
- The exposure to black hair dye effect of PPD on the changes in hematological where the result appearance significant decrease in Hb, RBCs, MCH, MCV, RBCs, MCHC and PLT that is possible cause the hemolytic extend to bone marrow leading to iron deficiency and insufficient produced of red blood cells lead to anemia. Furthermore, the increase in lymphocyte cells and decrease in neutrophil cells associated with significant decrease WBCs count, where observed in group C more group B compared with the control groups that is due to effect of PPD concentration on the immune system.
- The contact allergy related to exposure hair dye with PPD observed from the symptom of itching and dark spots.
- Our study experimental exhibit the correlation between concentration of PPD in black hair dye with in biochemical and hematological parameters and the resulting indication to an incidence of liver and kidney functions which can cause serious health problems.

VI. Recommendations

- Increase public awareness regarding the toxicity of black hair dyes containing PPD.
- The hairdressers should use protective plastic gloves to prevent direct skin contact and the user black henna dyes for the use long time and conduct routine health tests for individuals to reduce the medical effects of PPD on their bodies.
- Encourage further investigations and research to explore the health impacts of PPD and other ingredients in black henna and hair dyes.
- This study recommended to minimize of PPD to the premesibale levels in black henna dyes, and the decreased usage of hair dyes containing PPD, to safe to use and potentially reduced the hazardous.
- The PPD poisoning due to easy access and low priced, especially in poor and popular and rural areas living in rural areas, so the study reflects the importance of public awareness regarding dyeing, and government legislation must restrict the sale of commercial hair dyes and control the use of PPD in hair dye products.

VII. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that have influenced, or could be perceived to influence, the work reported in this article.

References

[1] Naveed, N. (2014), The perils of cosmetics, Journal of Pharmaceutical Sciences and Research; 6(10): 338-341.

[2] Han, H. J, Lee, J. H, Bang, H. C, Lee, H. J, Park, M. Y, Lee, Y. J, (2018) P-Phenylenediamine Hair Dye Allergy and Its Clinical Characteristics, Ann

Dermatol Vol. 30, No. 3, pISSN 1013-9087 • eISSN 2005-3894.

[3] Al-Enezi, M. H., & Aldawsari, F. S. (2022). Study of p-phenylenediamine concentrations after hair dye mixing: A call for safety reassessment. *Cosmetics, 9*(2), 41.

[4] Abouhadaf, R., Elderbi, M., Elburi, A., & Salih, A. (2024). Determination of para-phenylenediamine in henna cosmetic products in Benghazi. *GSJ, 12*(1).

[5] Elmanfe, G. M., Khreit, O. E., Abduljalil, O. A., & Abbas, N. M. (2022). Determination of para-phenylenediamine in henna samples collected from Libyan local markets using HPLC. *Al-Mukhtar Journal of Sciences, 37*(1), 13-21.

[6] Mahmud, S. A. S., Ferdous, T., Alam, M. M., Hossain, M. S., Shozib, H. B., Khalil, F., & Hossain, M. N. (2022). Para-phenylenediamine in commercially available henna preparations in Bangladesh. *Bioresearch Communications (BRC), 8*(1), 1088-1092.

[7] Shihata, A. (2018). Comparison study of toxicity of kohl and black stone hair dye. *J Environ Anal Toxicol, 8*(539), 2161-0525.1000539.

[8] Srhayri, R., Naimi, Y., Takky, D., Motaouakkil, S., & Habti, N. (2016). Chemical and biological analyses of the role of poly-para-phenylenediamine in erythrocyte resistance in a hypotonic environment. *J Membra Sci Technol, 6*(141), 2.

[9] Elyoussoufi. Z, Habti. N, Motaouakki. S, Cadi. R, (2024), The role of paraphenylenediamine in triggering apoptosis in murine myeloma cells through reactive oxygen species generation International Journal of Urology and Nephrology Vol. 12 (3), pp. 001-009. ISSN 2756-3855.

[10] Gupta, S., Kalode, D., & Doifode, C. (2023). A review on cosmetics and their associated adverse effects. *International Research Journal of Modernization in Engineering Technology and Science*. https://doi.org/10.56726/irjmets33011.

[11] Spornraft-Ragaller, P., Kämmerer, E., Gillitzer, C., & Schmitt, J. (2012). Severe allergic reactions to para-phenylenediamine in children and adolescents: Should the patch test concentration of PPD be changed? *JDDG: Journal der Deutschen Dermatologischen Gesellschaft, 10*(4), 258-263.

[12] Abdel-Hady. H. R, Yassa. A. H, Elkabsh. M. M, Mahmoud. M, M, Shaltout. S. E, (2023), Clinical features and outcome of acute hair dye poisoned cases presented in Aswan University Hospital, Aswan University Medical Journal, volume 3 / No.1/ (91-102) Online ISSN: 2735-3117.

[13] Kamboj. K, Sinha. A, Nigam. N, Ahuja. N, (2024), A retrospective study of clinical profile of para-phenylenediamene poisoning (hair dye poisoning) in cases admitted in medical college, Panacea Journal of Medical Sciences 14(1):164–167.

[14] Sayed. A. A, Abdelrahman. H. A, Sayed. A. Z, Abdelhameid. A. M, (2024). Incidence and outcome of laryngeal edema and rhabdomyolysis after ingestion of black rock, International Journal of Emergency Medicine 17:2.

[15] Morel, O. J., & Christie, R. M. (2011). Current trends in the chemistry of permanent hair dyeing. *Chemical Reviews, 111*(4), 2537-2561.

[16] Seo, J.-A., Bae, I.-H., Jang, W.-H., Kim, J.-H., Bak, S.-Y., Han, S.-H., & Lim, K.-M. (2012). Hydrogen peroxide and monoethanolamine as key causative ingredients for hair dye-induced dermatitis and hair loss. *Journal of Dermatological Science, 66*(1), 12-19.

[17] Abdel-Moneim, A. (2017). Acute toxicity of hair dye in Upper Egypt. *International Journal of Forensic Science & Pathology, 5*, 305-311.

[18] Asghar. S, Mahbub. S, Shahid. S, (2022), Paraphenylenediamine (Hair Dye) Poisoning: A Prospective Study on the Clinical Outcome and Side Effects Profile, Cureus 14(9): e29133. DOI 10.7759/cureus.29133.

[19] Beshir. L, Kaballo. B, Young. D, (2017), Attempted suicide by ingestion of hair dye containing pphenylenediamine : a case report, Annals of Clinical Biochemistry, Vol. 54(4) 507–510.

[20] Zhan. M, Zhan. M, Zeng. Q, Lin. D, Zhang. N, (2023), Association of *p*-phenylenediamine exposure with alterations of pulmonary function, pruritus and health-related quality of life in hair dye factory workers: a cross-sectional study, Scientifc Reports | 13:2623 https://doi.org/10.1038/s41598-023-29721-7.

[21] B. Encabo Durán,∗ D. Romero-Pérez, J.F. Silvestre Salvador, (2018), Allergic Contact Dermatitis Due to Paraphenylenediamine: An Update, Actas Dermosifiliogr. 109(7):602---609.

[22] El-Sarnagawy. N. G, Ghonem. M. M, Abdelhameid. A. M, Ali. M. O, Ismail. M. A, El Shehaby. M. D, (2023), Accuracy of Rapid Emergency Medicine Score and Sequential Organ Failure Assessment Score in predicting acute paraphenylenediamine poisoning adverse outcomes, Environmental Science and Pollution Research, 30:32489–32506.

[23] Salah. S. A. A, Mofleh. Q. M. S, (2025), HPLC-UV and FTIR Analytical Techniques to Determination of Paraphenylenediamine in Some Black hair dyes, Journal of Science and Technology, Vol. 30, Issue (1).

[24] Bharali, M. K., Basumatary, R., Rahman, T., & Dutta, K. (2012), Repeated topical application of para-phenylenediamine induces renal histopathological changes in rats, Toxicology International, 19(2), 132.

[25] Sieben, S, Kawakubo. Y, Al Masaoudi. T, Merk. H. F, Blomeke. B, (2002), Delayed-type hypersensitivity reaction to paraphenylenediamine is mediated by 2 different pathways of antigen recognition by specific alphabeta human T-cell clones. *J. Allergy Clin. Immunol.* 109, 1005–1011.

[26] Bharali, M. K., Dutta. K, Hematotoxic and Genotoxic Effect of Paraphenylenediamine after Subchronic Topical Application in Rats (2010), Bioscan an International Quarterly Journal of Life Sciences, 5(4), 585-590.

[27] El-Amin, E. I. S., Gah-Elnabi, M., Ahmed, W. A. M., Ahmed, R. G., & Khalid, K. E. (2014). Toxicity effects of hair dye application on liver function in experimental animals, J Clin Toxicol, V. 4, Issue 4, pp.2-5.

[28] Jen. E. S, Adam. A. M, Garba, H. S, Attah. O. O. M, Dibal. I. N, (2023), *Lawsonia inermis* and paraphenylenediamine exerts an insidious effect on the epidermal, renal, and hepatic tissue with repeated application: a histological and morphometrical study, Eur J Anat, 27 (4): 463-474 .

[29] 29. Kumara, S., Yadavb, S., Vishwakarmaa, V., Yadava, B., Guptac, R., Aggarwald, N., & Yadava, A. (2021). A review on toxicological hazards of p-Phenylenediamines: a primary ingredient of hair dye and potential biomarker-based risk assessment, Nat. Volatiles & Essent. Oils, 8(6): 754-781

[30] Abd-ElZaher, M. A., Fawzy, I. A., Ahmed, H. M., Abd-Allah, A. M., & Gayyed, M. F. (2012). Some toxicological health hazards associated with subchronic dermal exposure to paraphenylene-diamine (PPD): An experimental study. *Egyptian Journal of Forensic Sciences, 2*(3), 105-111.

[31] Naqvi, R., Akhtar, F., Farooq, U., Ashraf, S., & Rizvi, S. A. H. (2015). From diamonds to black stone; myth to reality: Acute kidney injury with paraphenylene diamine poisoning. *Nephrology, 20*(12), 887-891.

[32] Eissa, U. B., Khogali, F., Mohamed, E. E., Osman, E. E., Elmahi, O. M., & Waggiallah, H. A. (2021). Potential effects of henna (stone dye) extract paraphenylenediamine on human blood cells, liver, and renal function. *J Pharm Res Int, 33*, 192-200.

[33] Fan, L., Zhang, M., Liu, B., Liu, J., Tang, H., Zhu, B., & Fang, J. (2018). Effects of p-phenylenediamine on liver and kidney functions of occupationally exposed workers. *Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi, 36*(12), 923-926.

[34] Kallel, H., Chelly, H., Dammak, H., Bahloul, M., Ksibi, H., Hamida, C. B., & Bouaziz, M. (2005). Clinical manifestations of systemic paraphenylene diamine intoxication. *Journal of Nephrology, 18*(3), 308.

[35] Amiri. S. F, (2023), Effect of Henna-Induced Pigment Nephropathy on Kidney Outcomes: A Systematic Review and Meta-Analysis, Yemeni Journal of Medical and Health Research, Vol.12, No (1&2) , 106-175.

[36] Amira. D, Gana. I, Nouioui. A, Khlifi. F, Ben Salah. D, Masri. W, Belwaer. I, Ghorbel. H, (2015), Paraphenylenediamine Poisoning in Tunisia: A Case Report, Arab Journal of Forensic Sciences and Forensic Medicine, Volume 1 Issue (1), 138-142

[37] Young, E., Andersen, K. E., Bruze, M., Giménez-Arnau, A., Ross-Hansen, K., Johansen, J. D., Svedman, C. (2019). Twenty-eight-day follow-up of patch test reactions to p-phenylenediamine and p-phenylenediamine dihydrochloride: a multicentre study on behalf of the European Environmental and Contact Dermatitis Research Group. Contact Dermatitis, 81(1), 1-8.

Identifying Cybersecurity and Information Privacy Challenges During Digital Transformation in Industry (Study: Ministry of Communications and Information Technology)

K. A. Al-Masouri [1,*]

Received: 23/05/2025
Revised: 04/08/2025
Accepted: 05/08/2025

[1] Department of Computer Science, College of Computer and Information Sciences, National Institute of Administrative Sciences, Sana'a, Yemen.
*Corresponding Author's Email: Khalidahmed2009@gmail.com

Identifying Cybersecurity and Information Privacy Challenges During Digital Transformation in Industry (Study: Ministry of Communications and Information Technology)

Khaled Ahmed Al-Masouri
Department of Computer Science, College of Computer and Information Sciences, National Institute of Administrative Sciences,
Sana'a, Yemen.
Khalidahmed2009@gmail.com

Abstract— The world's increasing focus on new technologies and the emergence of smart cities, despite their potential capabilities and benefits, has raised serious concerns about cybersecurity threats and citizen privacy challenges. Sana'a is no exception to this rule in its smart city transformation process. This research paper, through a descriptive survey, aims to provide a framework for managing cybersecurity and privacy challenges in Sana'a's transition to a smart city. In this research, these challenges were identified through in-depth desk studies and the application of the Fuzzy Delphi method among a sample of experts (including ten senior managers and relevant officials from the Ministry of Communications and Information Technology, the Digital Transformation Department, and the Digital Library in Sana'a). Furthermore, the importance (weight) of each challenge was determined using the Fuzzy Best-Worst Method (FBWM). The results of this study indicated the high ability of the proposed framework to accurately identify and weighthese challenges. The research results indicated the high ability of the proposed framework to accurately identify and assess these challenges, as well as identify the three "challenges—"legislative challenge," "lack of secure communications," and "insecure APIs and protocols"—as "the most significant cybersecurity and privacy challenges in the development of a smart city in the capital, Sana'a. Preventive and corrective measures were presented for each challenge accordingly.

Keywords—smart city, cybersecurity, privacy, Fuzzy Delphi method, best approach, worst case.

I. Introduction

With the global movement toward urbanization and the widespread use of information technology, the concept of smart cities has emerged and has received significant attention from researchers in recent years. The primary goal of smart cities is to improve people's lifestyles, encourage development without compromising the resources of future generations (with a focus on sustainable development), and achieve progress in urban functions. Despite the benefits, smart cities pose hidden risks, including information leaks, cybersecurity threats, privacy challenges, and malicious cyberattacks (Manchanda et al., 2020). Given the uncertainty of the real world and the importance of detecting cyberattacks, the cybersecurity of future smart cities and their smart networks is critical. The current evolution of cybersecurity has not kept pace with the enthusiastic adoption of smart city technologies, so appropriate design based on deep learning methods is essential to protect smart cities from cyber risks (Faotu et al., 2024). In addition to cybersecurity, there are specific privacy issues, particularly those related to direct objectivity, which can be classified into three categories: communication, individual, and business. Phishing, fraud, data breaches, eavesdropping, attacks on networks and websites, and so on, are examples of privacy-related challenges and violations (Aljrad, M.J.H. & Al-Dhlan, 2023).

Sana'a is no exception to these trends in its journey toward becoming a smart city. Providing protection in a smart city involves identifying and assessing threats and challenges and implementing appropriate measures and solutions to address priority challenges. In light of the impact of cybersecurity threats and privacy challenges on reducing the efficiency of smart city development in Sana'a, this study seeks to provide a systematic framework for identifying and analyzing cybersecurity threats and privacy challenges in Sana'a's transition to a smart city. The rest of this article is organized as follows. The next section reviews the research background and defines the initial theoretical framework for the challenges. The research methodology and proposed methodological framework are then presented. The proposed framework for identifying and assessing cybersecurity threats and privacy challenges in the smart city of Sana'a is implemented below. The final section of the article is also dedicated to discussion and conclusions.

II. Review of Research Background and Theoretical Framework

Given the emerging nature of the current study area, and despite the unique capabilities and functions of smart cities, the number of local and international studies specifically examining the successful implementation of smart cities and the challenges they face is limited. We will present the most important of these studies below.

Alromaihi et al. (2018) analyzed the issue of security and privacy in smart city healthcare applications. In this regard, they reviewed various Internet of Things (IoT) applications and their cyber vulnerabilities and provided a comprehensive assessment of methods for combating cyber attacks. They then presented techniques for countering cyber attacks on smart city healthcare IoT devices and explained the different types of attacks and their associated security needs.

In their research, Berkel et al. (2017) combined threat analysis and enterprise architecture modeling to examine and mitigate these challenges from a holistic perspective. They presented an information security architecture that can help

stakeholders in smart city projects build safer smart cities. Gunduz & Das (2020) analyzed the threats and possible solutions to IoT-based smart grids. They focused on the types of cyberattacks, examined the cybersecurity status of smart grids in depth, and discussed and examined network vulnerabilities, countermeasures, and security requirements. Zhao et al. (2021) presented a systematic review of research published in the field of smart cities between 2011 and 2019. The aim of their study was to create a comprehensive picture of research progress in the field of smart cities, as well as to identify important issues and identify research gaps.

By examining previous domestic research, we note that the limited number of studies conducted in this field have focused more on basic concepts and compiling material in the form of review papers. Among foreign research, some studies focused on introducing new material in this field, as well as examining the shortcomings of smart city implementation in developed countries in the form of case studies and presenting lessons learned, while other studies have only advanced to the stage of identifying smart city security challenges. Without a thorough evaluation and analysis, these studies have simply offered fragmented solutions to address these challenges. To complement previous research and fill research gaps, this study aims to comprehensively examine the research and achievements of researchers in the field of cybersecurity and privacy in smart cities. This is achieved through an analytical study, identifying and converging expert opinions on the cybersecurity threats and privacy challenges facing Sana'a during the transition to a smart city, and determining the degree of importance of these challenges. Through a review of relevant research, a list of cybersecurity threats and privacy challenges in smart cities was compiled. The list, drawn from a set of previous research findings and based on their frequency in articles, is presented in Table 1.

Table 1. A preliminary theoretical framework for cybersecurity threats and privacy challenges in smart cities.

References	Challenges	Factors
Khatoun & Zeadally, 2017; Aldairi, 2017; Alromaihi et al., 2018; Berkel et al., 2018; Braun et al., 2018; Baig et al., 2017; Thing, 2024; Arabo, 2022; Pelton & Singh, 2019;	1. Increased volume of digital transactions 2. Increase in mobile applications and communications 3.Increased use of artificial intelligence in digital networks and machine-to-machine communications 4. Vulnerable software and hardware products 5. Cyberwarfare and terrorism 6. Cyber espionage 7. Data manipulation and phishing attacks 8. Data loss 9. Virus and malware penetration of smart city systems 10. Legislative challenges 11. Theft of data, information, and physical devices 12. Hardware and software inefficiencies 13. Challenges in accessing data 14. Insecure APIs and protocols 15. Denial of service (DoS) 16. Sensor failure 17. Lack of secure communications 18. Challenges in data management and storage 19. Disruption of critical infrastructure 20. Cloud security 21. Threats to artificial intelligence	Cybersecurity Challenges
Aldairi, 2017; Braun et al., 2018; Baig et al., 2017; Thing, 2024;	1. Privacy Threats in Data Extraction and Sharing 2 .Privacy Threats in Mash-Up Data 3 .Eavesdropping 4. Data Access Challenge 5. Risk to Confidentiality and Integrity 6 .Risk of Fraud and Data Leakage 7 .Identity Forgery 8. Fake Information 9. Side Channel Attacks 10. Secondary Use of Collected Data 11. IP Address Forgery 12. Attacks on Data Integrity	Privacy challenges

III. Research Methodology and Proposed Methodological Framework

The current study is classified as applied research in terms of its objective. It is descriptive survey research based on a data collection method. The statistical population for this study includes senior managers and relevant officials in the Ministry of Communications and Information Technology, the Digital Transformation Department, and the Digital Library in the Sana'a Municipality. Since decision-making regarding the assessment of the smart city and its security challenges resides at the strategic level of these institutions, and the data required for this research is expected to be available to a small number of organizational managers and experts, ten organizational experts were selected using a purposive sampling method. The demographic information of the expert group in this study is presented in Table .2

Table 2. Demographic Information of the Expert Panel Members

Date	Work area	Education Level	Experts
9years	Ministry of Communications and Information Technology	PhD	1
15years		Master's Degree	2
8years		Bachelor's Degree	3
10years		Master's Degree	4
7years		PhD	5
3years	Digital Transformation Department	PhD	6
6years		Master's Degree	7
16years		Bachelor's Degree	8
8years	Digital Library and Information Exchange	Master's Degree	9
5years		PhD	10

In this study, a two-part questionnaire was designed by the researcher to collect data. Content validity was used to determine the validity of the questionnaires. The questionnaires were presented to a number of university professors and organizational experts, and the questionnaire components and structure were approved. To measure the reliability of the first questionnaire, the expert opinion convergence threshold (α) was used, which represents the difference in expert consensus over two consecutive iterations of the fuzzy Delphi method. According to the agreement in this study, the minimum expert opinion convergence threshold value is as follows: 0/1= α. The second questionnaire was used to collect the data needed to determine the weights of cybersecurity threats and privacy challenges using the fuzzy best-worst multi-criteria (F-BWM) method. The consistency ratio method was used to measure reliability (Guo & Zhao, 2017). Figure 1 schematically illustrates the structure and steps of the proposed methodological framework.

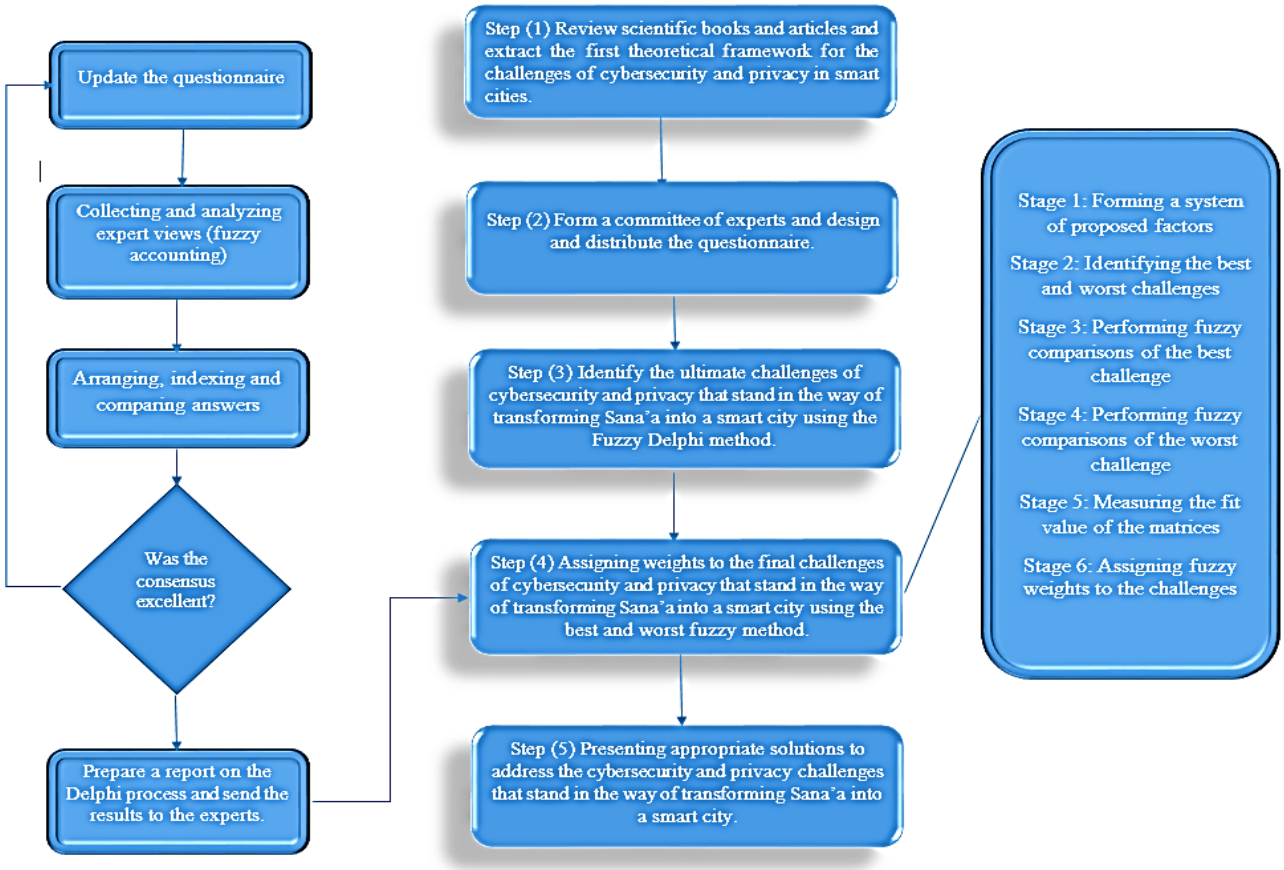


Figure 1. Proposed Methodological Framework

Considering the general research framework, this study utilizes the Fuzzy Delphi method and the Fuzzy Best-Worst method to analyze the data. Excel software was used to define and weigh the cybersecurity and privacy challenges in the Sana'a Smart City, respectively. In the current study, the proposed methods are implemented in an ambiguous environment, using group decision-making. The group decision-making strategy will prevent bias in the results and, by adhering to collective wisdom, will increase decision-making accuracy.

Applying the above methods in an ambiguous environment makes it possible to reduce uncertainty in subjective expert judgments and increase decision-making accuracy by using three-point estimates and considering the likelihood functions of their opinions.

The best-worst method was first introduced by Rezaei (2022). In this method, the best and worst are indexed by the decision maker, and a pair of factors can be compared. Guo & Zhao (2017) presented the fuzzy best-worst method in an article. The fuzzy best-worst method has the salient feature of obtaining weights from indicators with fuzzy numbers rather than specific numbers. Based on the method proposed by the researchers, the algorithm for solving the indicator weighting problem includes five main steps: 1) Forming a system of decision-making indicators, 2) Identifying the best (most important) and worst (least important) indicators, 3) Performing fuzzy preferential comparisons for the best indicator, 4) Performing fuzzy preferential comparisons for the worst indicator, 5) Measuring the fit rate of the matrices, and Determining the optimal fuzzy weights for the indicators (Guo & Zhao, 2017).

IV. Data Analysis

By conducting desk studies and surveying the cybersecurity and privacy challenges in Table1, the Fuzzy Delphi method was adopted to adapt and localize the initial theoretical framework to the prevailing social, economic, cultural, political, legal, and technological infrastructure environment in Sana'a. The Fuzzy Delphi method, a combination of the Delphi method and fuzzy logic, was first introduced by Kaufman and Gupta in the 1980s. Experts used fuzzy numbers to make three-point estimates of phenomena (Chen, 2022). In this method, membership functions are used to represent expert opinions. This avoids the need for experts to constantly revise their opinions. Furthermore, since all opinions are described in terms of membership degrees, no useful information is lost.

With the distribution of the first questionnaire, the opinions of experts during the various rounds of this method regarding cybersecurity threats and privacy challenges were presented and analyzed. It should be noted that, based on the organizational experts' perspective, the level of acceptance of each challenge after consensus was considered to be 5. The threshold level of disagreement between the two survey phases, which represents the stopping state of the Fuzzy Delphi algorithm, was considered to be 0.1. After the Fuzzy Delphi questionnaire was sent to the experts and completed using Table 3, the results of their opinion statistics were compiled. Based on the opinions of 10 experts, their combined opinions were calculated using the arithmetic mean, and then the fuzz was removed.

Table 4presents the results of the experts' opinions on cybersecurity threats and privacy challenges in Sana'a's transition to a smart city in the first phase of the survey.

Table 3. Verbal expressions of the level of agreement on the presence of cybersecurity threats and privacy challenges, particularly in the Sana'a Smart City, and corresponding triangular fuzzy numbers.

The corresponding triangular fuzzy number	Verbal phrase
(1 ،1 ،3)	Very little
(1 ،3 ،5)	A little
(3 ،5 ،7)	Average
(5 ،7 ،9)	much
(7 ،9 ،10)	Too much

Table 4. Average expert opinions in the first phase of the survey

Fuzzy Average Comments	Average Comments	Cybersecurity Threats and Privacy Challenges	Dimension
6.33	(2.5 ,2.7 ,7.8)	Increasing volume of digital transactions	Cybersecurity Challenges
6.73	(6.5 ,6.7 ,1.9)	Increasing applications and mobile communications	
5.88	(8.3 ,8.6 ,5.8)	Increasing use of artificial intelligence in digital networks and machine-to-machine communications	
6.93	(8.5 ,8.7 ,3.9)	Vulnerable software and hardware products	
3.75	(3 ,6.5 ,3.7)	Cyberwar and terrorism	
5.93	(5 ,8.6 ,5.8)	Cyber espionage	
3.68	(8.3 ,6.5 ,5.7)	Data tampering and phishing attacks	
5.38	(3.3 ,3.6 ,1.8)	Data loss	
6.5	(3.5 ,3.7 ,9)	Virus and malware penetration of smart city systems	
6.75	(6.5 ,6.7 ,9)	Legislative challenges	
3.6	(6.3 ,6.5 ,6.7)	Data, information, and physical device theft	
3.33	(3.4 ,3.5 ,3.7)	Hardware and software inefficiencies	
5.48	(3.4 ,3.6 ,1.8)	Challenges in accessing data	
6.08	(5 ,7 ,7.8)	Insecure APIs and protocols	
5.1	(2.3 ,6 ,8.7)	Denial of service (DoS)	

3.05	(2.3 ,5 ,7)	Sensor failure	
5.53	(6.3 ,3.6 ,1.8)	Lack of secure communications	
3.88	(8.3 ,8.5 ,5.7)	Challenges in data management and storage	
5.68	(6.3 ,6.6 ,3.8)	Disruption of critical infrastructure	
5.88	(8.3 ,8.6 ,5.8)	Cloud security	
3.7	(8.3 ,6.5 ,3.7)	AI threats	
6.5	(3.5 ,3.7 ,9)	Privacy Threats in Data Extraction and Sharing	Privacy challenges
5.88	(8.3 ,8.6 ,5.8)	Privacy Threats in Data	
6.13	(5 ,7 ,8.5)	Eavesdropping	
3.65	(6.3 ,6.5 ,3.7)	Data Access Challenge	
6.08	(5 ,7 ,7.8)	Confidentiality and Integrity Risk	
6.18	(2.5 ,7 ,5.8)	Fraud and Data Leakage Risk	
3.9	(3 ,8.5 ,6.7)	Identity Forgery	
5.9	(8.3 ,8.6 ,3.8)	Fake Information	
5.3	(3.3 ,2.6 ,8)	Side Channel Attacks	
5.68	(6.3 ,6.6 ,3.8)	Secondary Use of Collected Data	
5.55	(6.3 ,3.6 ,8)	IP Address Forgery	
5.15	(2.4 ,6 ,6.7)	Attack on data integrity	

In the second phase of the survey, after updating the first questionnaire and applying each expert's opinion to the average of the opinions, experts were given the opportunity to revise their previous opinions if they wished. In accordance with the first phase, after the results of the new expert opinions were tallied, the average opinions were calculated and simplified. Finally, the difference between the average opinions in these two phases was calculated according to Table 5. Indicators for which the difference between the averages of these two phases was less than the minimum (0.1) were excluded from the survey process or included in the third phase of the survey.

Table 5. Differences in Expert Opinions in the First and Second Phases of the Survey

Differences in Steps 1 & 2	**Average Comments in Step 2**	**Average Comments in Step 1**	**Cybersecurity Threats and Privacy Challenges**	**Dimension**
0.17	6.5	6.44	Increasing volume of digital transactions	Cybersecurity Challenges
0.20	6.94	6.74	Increasing applications and mobile communications	
0.15	5.74	5.88	Increasing use of artificial intelligence in digital networks and machine-to-machine communications	
0.02	6.95	6.94	Vulnerable software and hardware products	
0.08	4.84	4.75	Cyberwar and terrorism	
0.12	6.05	5.94	Cyber espionage	
0.24	4.45	4.68	Data tampering and phishing attacks	
0.64	4.85	5.48	Data loss	
0.77	5.74	6.5	Virus and malware penetration of smart city systems	
0.20	6.95	6.75	Legislative challenges	
0.28	4.88	4.6	Data, information, and physical device theft	
0.45	4.88	4.44	Hardware and software inefficiencies	
1.04	4.45	5.48	Challenges in accessing data	
0.02	6.1	6.08	Insecure APIs and protocols	
0.45	5.45	5.1	Denial of service (DoS)	
0.20	4.25	4.05	Sensor failure	
0.50	5.04	5.54	Lack of secure communications	
0.04	4.85	4.88	Challenges in data management and storage	
0.17	5.85	5.68	Disruption of critical infrastructure	
0.04	5.85	5.88	Cloud security	
0.45	4.25	4.7	AI threats	
0.02	6.48	6.5	Privacy Threats in Data Extraction and Sharing	Privacy challenges
0.05	5.84	5.88	Privacy Threats in Data	
0.42	6.45	6.14	Eavesdropping	

0.05	4.6	4.65	Data Access Challenge
0.00	6.08	6.08	Confidentiality and Integrity Risk
0.14	6.05	6.18	Fraud and Data Leakage Risk
0.40	5.2	4.9	Identity Forgery
0.05	5.84	5.9	Fake Information
0.50	4.8	5.4	Side Channel Attacks
0.05	5.64	5.68	Secondary Use of Collected Data
0.08	5.64	5.55	IP Address Forgery
0.08	5.24	5.15	Attack on data integrity

According to the results, it can be noted that the expert panel reached consensus on 79challenges whose level of disagreement in the first and second rounds of the survey was below the minimum (1.7) (the challenges are highlighted in bold in the table above). Therefore, the survey process for these challenges was discontinued. Of these challenges, three—"Cyberwar and Terrorism," "Data Management and Storage Challenge," and "Data Access Challenge"—had an acceptance level below 5. Therefore, they were removed from the final theoretical framework for cybersecurity threats and privacy challenges in Sana'a Smart City. A further 71 challenges were included in the final theoretical framework. By repeating the survey process for four stages, consensus was reached on all challenges, and the final theoretical framework for cybersecurity threats and privacy challenges in Sana'a's transition to a smart city was obtained, as shown in Table .6

Table 6. Final Theoretical Framework for Cybersecurity Threats and Privacy Challenges in Sana'a's Transition to a Smart City

Cybersecurity Challenges and Privacy Challenges	**Dimensions**
Increasing volume of digital transactions (C1)	Cybersecurity
Increasing mobile applications and communications (C2)	Challenges
Existence of vulnerable software and hardware products (C3)	
Cyber espionage (C4)	
Virus and malware penetration of smart city systems (C5)	
Legislative law (C6)	
Insecure APIs and protocols (C7)	
Denial of service attacks (C8)	
Lack of secure connectivity (C9)	
Disruption of critical infrastructure (C10)	
Cloud security (C11)	
Privacy threats in data extraction and sharing (P1)	Privacy Challenges
Privacy threats in data (P2)	
Mixed eavesdropping (P3)	
Confidentiality and integrity risks (P4)	
Fraud and data leakage risks (P5)	
Identity theft (P6)	
Fake information (P7)	
Secondary use of collected data (P8)	
IP address spoofing (P9)	
Attacks on data integrity (P10)	

Next, to determine the weights of these threats and challenges using the fuzzy best-worst multi-criteria decision-making method, a questionnaire was first designed and distributed among the members of the regulatory expert panel. In this questionnaire, to determine the best and worst indicators, the average acceptance value of these indicators using the fuzzy Delphi method, in addition to the direct views of the expert panel members, was used as criteria. The challenges "Legislative Challenge (C6)" and "Identity Theft (P9)" were identified as the best and worst indicators, respectively. The results are shown in Table 7.

Table 7. Average Expert Opinions on the Priority of the Best (Most Important) Challenge in "Legislative Challenge" (C6) "relative to Other Challenges and Other Challenges relative to the Worst (Least Important) Challenge, "Identity Theft (P9)"

Worst Indicator	**Best Indicator**	**Indicators**	**Dimensions**
(5.3, 6.2, 17.2)	(7.2, 2.2, 72.1)	C1	Cybersecurity Challenges
(2.2, 7.2, 25.2)	(8.2, 2.2, 82.1)	C2	
(1.2, 7.2, 2.2)	(1.2, 6.2, 12.2)	C3	
(9.2, 4.2, 9.2)	(4, 5.2, 2)	C4	
(9.2, 4.2, 92.2)	(6.2, 1.2, 62.2)	C5	
(95.2, 5.2, 5.2)	(1, 1, 1)	C6	
(25.2, 9.2, 45.2)	(2, 5.2, 2.2)	C7	
(5.2, 2, 52.2)	(8.2, 2.2,82.1)	C8	
(55.2, 1.2, 65.2)	(9.2, 5.2, 12.2)	C9	

(55.2, 1.2, 67.2)	(7.2, 2.2, 7.2)	C10	
(8.2, 2.2, 8.2)	(25.2, 9.2, 47.2)	C11	
(2.2, 7.2, 22.2)	(2.2, 7.2, 22.2)	P1	Privacy Challenges
(7.2, 2.2, 72.2)	(6.2, 1.2, 65.1)	P2	
(25.2, 8.2, 25.2)	(2.2, 7.2, 22.2)	P3	
(45.2, 2, 55.2)	(7.2, 2.2, 7.2)	P4	
(1.2, 6.2, 12.2)	(1.2, 6.2, 12.2)	P5	
(95.2, 5.2, 7.2)	(4.2, 9.2, 42.2)	P6	
(21.2, 8.2, 2.7)	(2, 5.2, 2)	P7	
(2.2, 7.2, 22.1)	(85.2, 4.2, 97.1)	P8	
(1, 1, 1)	(95.2, 5.2, 5.2)	P9	
(22, 7.2, 22.2)	(2, 5.2, 2.2)	P10	

By entering the resulting values into a linear programming model, an expanded model consisting of 63 variables and 184 constraints was introduced into the GAMS program. By solving the model, the optimal value for the challenge weight vector and function was determined, and the goal was as shown in Table 1. Given that, according to the standard, the concordance index (CI) for is 6.69, and for the value is 8.04. Since this study obtained then, by making this number certain, the resulting value is 3.5. By calculating the arithmetic means of 6.69and $\tilde{a}_{BW} =$ 8.04, the concordance factor becomes .7.36

$$\tilde{a}_{BW} = (3/05,3/5,3/95)$$

Table 8. Final weights for cybersecurity threats and privacy challenges in Sana'a during the transition to a smart city.

w_j^*	$\widetilde{w}_j^*$	**Challenges**
0.0435	(0.033 ,0.033 ,0.062)	C1
0.0355	(0.034 ,0.034 ,0.063)	C2
0.0380	(0.036 ,0.037 ,0.063)	C3
0.0323	(0.035 ,0.032 ,0.051)	C4
0.0370	(0.038 ,0.036 ,0.06)	C5
0.0930	(0.09 ,0.093 ,0.098)	C6
0.0515	(0.036 ,0.051 ,0.069)	C7
0.0390	(0.038 ,0.037 ,0.068)	C8
0.0537	(0.039 ,0.053 ,0.071)	C9
0.0328	(0.033 ,0.032 ,0.056)	C10
0.0393	(0.038 ,0.038 ,0.066)	C11
0.0352	(0.034 ,0.034 ,0.061)	P1
0.0347	(0.035 ,0.034 ,0.057)	P2
0.0362	(0.035 ,0.035 ,0.062)	P3
0.0338	(0.033 ,0.033 ,0.058)	P4
0.0378	(0.033 .0.037 .0.066)	P5
0.0348	(0.033 ,0.034 ,0.059)	P6
0.0370	(0.031 ,0.037 ,0.063)	P7
0.0362	(0.038 ,0.035 ,0.059)	P8
0.0200	(0.019 ,0.002 ,0.021)	P9
0.0388	(0.037 ,0.038 ,0.063)	P10
	1.286	$\bar{\xi}^*$ Quantity
	7.36	Compatibility Index
	0.175	Compatibility Rate

It was noted that five challenge indicators—legislative law (C6), lack of secure communications (C9), insecure protocols (C7), cloud security (C11), and denial-of-service (DoS) attacks (C8)—were identified as the most significant challenges in improving Sana'a's smart city. These five indicators belong to the "cybersecurity threats" dimension, demonstrating the importance of this dimension, from the experts' perspective, to the successful implementation of Sana'a's smart city infrastructure and system. Among the privacy challenges, "attack on data integrity (P10)" was also identified as a significant challenge in improving Sana'a's smart city. The identification of the proposed indicators does not indicate the unimportance of the other indicators; rather, each indicator has a different degree of importance in the emergence of challenges in the Sana'a improvement process, according to the resulting weights. Furthermore, the results obtained for the consistency rate indicate a high degree of consistency in the results and reliability of the questionnaire for this step of the methodology. Since determining the importance of cybersecurity threats and privacy challenges depends on the subjective judgments and expressed preferences of the members of the expert panel in an environment of uncertainty, predicting the behavior of experts' risk tolerance levels using the alpha cut approach based on the weights obtained for these threats and challenges can provide the basis for a sensitivity analysis of this issue.

In other words, this approach will determine the extent to which increasing or decreasing the risk tolerance of experts when presenting their judgments on the relative priority of indicators over each other changes the resulting weights of these indicators. Each complete fuzzy set is uniquely defined by its alpha cutoffs. The alpha cutoffs for any fuzzy number, for any alpha value in the interval [0, 1], are a closed interval of real numbers. The closer the alpha cutoffs in the proposed range are to zero, the lower the risk level of experts in presenting their judgments on the indicators. The results of the sensitivity analysis of experts' opinions regarding the relative priority of indicators at different alpha cutoff levels are presented in Table .9

Table 9. Prediction, sensitivity analysis, expert opinions, and final weights of cybersecurity threats and privacy challenges in smart city development in Sana'a

Alpha Surface										Factor
α=0/1	**α=0/2**	**α=0/3**	**α=0/4**	**α=0/5**	**α=0/6**	**α=0/7**	**α=0/8**	**α=0/9**	**α=1**	
1 / 1418	/ 1463 1	1 / 466	/ 1469 1	/ 1463 1	1 / 1431	/ 1434 1	1 / 1457	1 / 1441	1 / 1445	C1
1 / 1417	/ 1413 1	/ 1416 1	/ 1439 1	/ 1411 1	1 / 1431	/ 1434 1	1 / 1467	1 / 1451	1 / 1455	C2
1 / 1431	/ 1436 1	/ 1434 1	/ 1437 1	/ 1431 1	1 / 1411	/ 1436 1	1 / 1414	11417	1 / 1411	C3
1 / 1433	/ 1433 1	/ 1431 1	/ 1471 1	/ 1431 1	1 / 1436	/ 1475 1	1 / 1475	1 / 1474	1 / 1479	C4
1 / 1411	/ 1416 1	/ 1414 1	/ 1417 1	/ 1411 1	1 / 1431	/ 1416 1	1 / 1414	1 / 1417	1 / 1411	C5
1 / 1341	/ 1341 1	/ 1341 1	/ 1341 1	/ 1341 1	1 / 1331	/ 1341 1	1 / 1341	1 / 1331	1 / 1341	C6
1 / 1574	/ 1579 1	/ 1537 1	/ 1537 1	/ 1571 1	1 / 1573	/ 1571 1	1 / 1571	1 / 1536	1 / 1575	C7
1 / 1576	/ 1577 1	/ 1531 1	/ 1574 1	/ 1571 1	1 / 1516	/ 1517 1	1 / 1431	1 / 1444	1 / 1431	C8
1 / 1543	/ 1541 1	/ 1536 1	/ 1545 1	/ 1549 1	1 / 1547	/ 1547 1	1 / 1593	1 / 1571	1 / 1571	C9
1 / 1447	/ 1447 1	/ 1441 1	/ 1471 1	/ 1471 1	1 / 1475	/ 1477 1	1 / 1477	1 / 1471	1 / 1471	C10
1 / 1571	/ 1575 1	/ 1577 1	/ 1513 1	/ 1511 1	1 / 1514	/ 1517 1	1 / 1433	1 / 1436	1 / 1437	C11
1 / 1417	/ 1411 1	/ 1461 1	/ 1466 1	/ 1467 1	1 / 1467	/ 1453 1	1 / 1456	1 / 1454	1 / 1457	P1
1 / 1453	/ 1451 1	/ 1456 1	/ 1455 1	/ 1457 1	1 / 1457	/ 1457 1	1 / 1443	1 / 1441	1 / 1441	P2
/ 21417 1	/ 1411 1	/ 1411 1	/ 1416 1	/ 1417 1	1 / 1417	/ 1463 1	1 / 1466	1 / 1464	1 / 1467	P3
1 / 1457	/ 1457 1	/ 1451 1	/ 1441 1	/ 1441 1	1 / 1445	/ 1447 1	1 / 1447	1 / 1441	1 / 1471	P4
1 / 1437	/ 1437 1	/ 1431 1	/ 1411 1	/ 1411 1	1 / 1415	/ 1417 1	1 / 1417	1 / 1411	1 / 1411	P5
1 / 1467	/ 1467 1	/ 1461 1	/ 1451 1	/ 1451 1	1 / 1455	/ 1457 1	1 / 1457	1 / 1451	1 / 1441	P6
1 / 1411	/ 1411 1	/ 1411 1	/ 1411 1	/ 1411 1	1 / 1411	/ 1411 1	1 / 1411	1 / 1411	1 / 1411	P7
1 / 1417	/ 1411 1	/ 1411 1	/ 1416 1	/ 1417 1	1 / 1417	/ 1463 1	1 / 1466	1 / 1464	1 / 1467	P8
1 / 1711	/ 1711 1	/ 1711 1	/ 1711 1	/ 1711 1	1 / 1711	/ 1711 1	1 / 1711	1 / 1711	1 / 1711	P9
1 / 1517	/ 1517 1	/ 1511 1	/ 1431 1	/ 1431 1	1 / 1435	/ 1437 1	1 / 1437	1 / 1431	1 / 1411	P10

Figure 2. Predicting changes in the weights of cybersecurity threats and privacy challenges in smart city development in Sana'a, based on experts' risk tolerance levels and their subjective judgments regarding the relative priority of indicators.

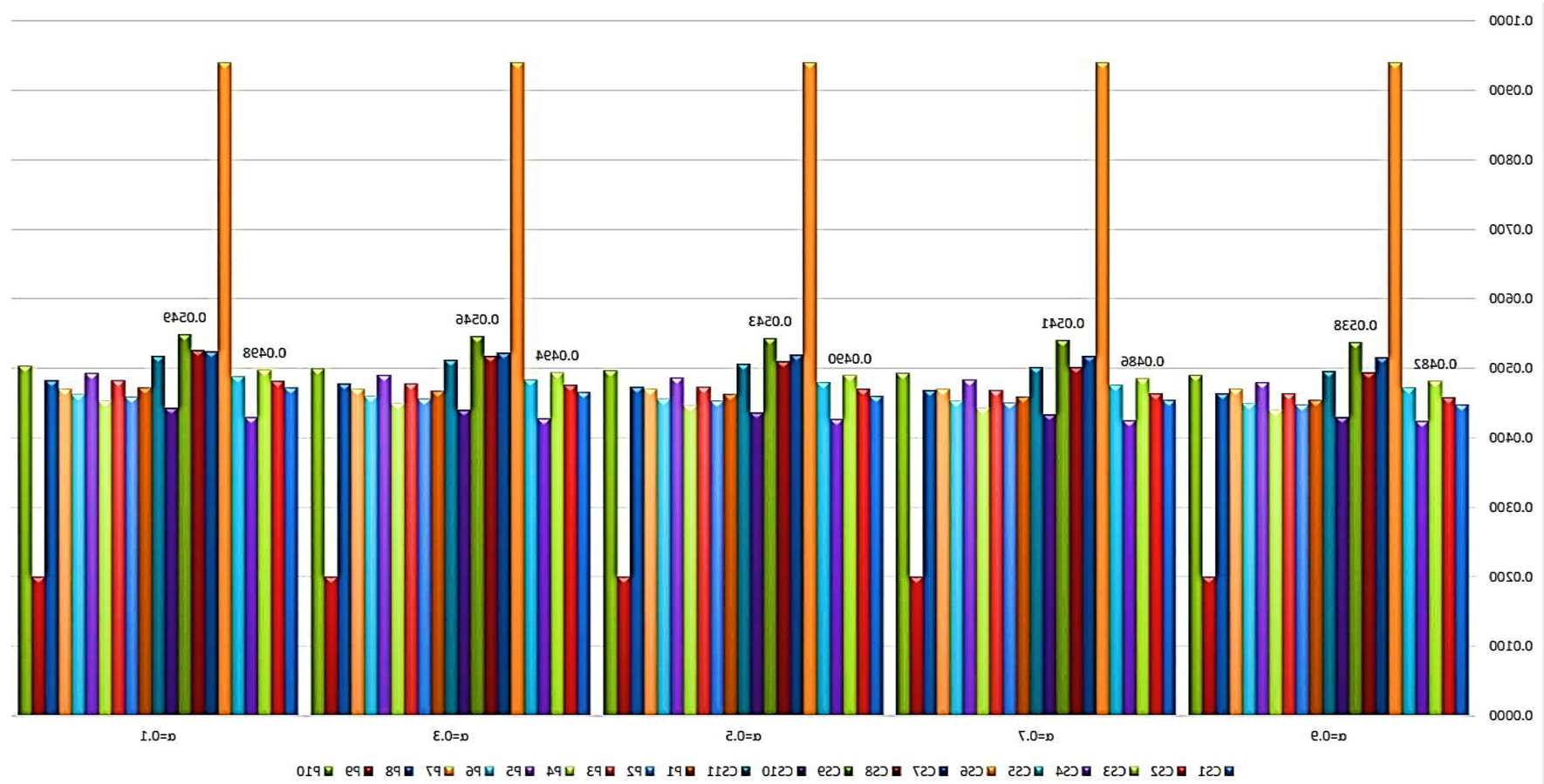


Figure 2. Predicting the weights of cybersecurity threats and privacy challenges in smart city development in Sana'a, based on experts' risk awareness levels

https://doi.org/10.20428/jst.v30i9.3037

As it turns out, as the experts' risk level increased in their judgments regarding the relative importance of cybersecurity threats and privacy challenges (by decreasing the alpha threshold), the weight assigned to that threat and challenge increased. Given that the fuzzy probability functions for all challenges are equal (i.e., the triangular functions are considered linear), the increasing weight of threats and challenges was achieved by decreasing the alpha threshold, according to the slope of the triangular boundary line of that indicator, by a certain factor.

V. Discussion and Conclusion

Since the successful implementation of the Sana'a Smart City project depends on identifying and assessing cybersecurity threats and privacy challenges and providing a documented plan to address the prioritized threats and challenges, the current study was written with the aim of presenting a model for analyzing cybersecurity threats and privacy challenges in Sana'a's transition to a smart city. This research, both objectively and in terms of communication with the academic community, is expected to address the priorities in the source documentation system, as well as in terms of research methodology and data analysis tools. In this research, a unique model for assessing cybersecurity threats and privacy challenges was implemented using a mixed-method approach, using a fuzzy decision-making approach. Having a plan for addressing cybersecurity and privacy threats, obstacles, and challenges is an important step toward the successful implementation of the Sana'a Smart City project. Based on the research findings, the following strategies and operational plans were proposed to address the threats and challenges identified in the development of the Sana'a Smart City project. These strategies include creating a culture of respect for the legal aspects of security, implementing safety standards, following the recommendations of national cybersecurity agencies and IT security actors, promoting appropriate practices for the use of ICT, developing performance standards, and educating specialists, technicians, programmers, and experts with experience in the fields of networking, ICT, and computer science to include technical measures in their agendas to address cybersecurity threats and privacy challenges. These measures include closing intrusion routes by establishing secure protocols, providing physical security for equipment, network cables, and servers, encrypting network traffic using stable symmetric algorithms, enhancing cloud security, using secure connections such as VPNs for remote access, securing wireless networks with WPA 2protocols, deploying firewalls at every transmission point, etc.

References

[1] A. AlDairi, "Cyber Security Attacks on Smart Cities and Associated Mobile Technologies," *Procedia Computer Science*, vol. 109, pp. 1086–1091, 2022, doi: 10.1016/j.procs.2017.05.391.

[2] S. Alromaihi, W. Elmedany, and C. Balakrishna, "Cyber Security Challenges of Deploying IoT in Smart Cities for Healthcare Applications," in *Proc. 6th Int. Conf. Future Internet of Things and Cloud Workshops (FiCloudW)*, 2018, pp. 140–145, IEEE, doi: 10.1109/WFiCloud.2018.00028.

[3] M. J. H. Aljrad and K. Al-Dhlan, "The Effect of Using Social Engineering for Cybersecurity on the Internet of Things Environment," *مجلة العلوم والتكنولوجيا*, vol. 27, no. 2, pp. 43–47, 2023, doi: 10.20428/jst.v27i2.2055.

[4] A. Arabo, "Cyber security challenges within the connected home ecosystem futures," *Procedia Computer Science*, vol. 61, pp. 227–232, 2022, doi: 10.1016/j.procs.2015.09.201.

[5] Z. A. Baig et al., "Future challenges for smart cities: Cyber-security and digital forensics," *Digital Investigation*, vol. 22, pp. 3–13, 2020, doi: 10.1016/j.diin.2017.06.015.

[6] A. R. Berkel, P. M. Singh, and M. J. van Sinderen, "An Information Security Architecture for Smart Cities," in *Int. Symp. Business Modeling and Software Design*, Cham: Springer, 2018, pp. 167–184.

[7] T. Braun, B. C. Fung, F. Iqbal, and B. Shah, "Security and privacy challenges in smart cities," *Sustainable Cities and Society*, vol. 39, pp. 499–507, 2018, doi: 10.1016/j.scs.2018.02.039.

[8] D. Chen, P. Wawrzynski, and Z. Lv, "Cyber security in smart cities: A review of deep learning-based applications and case studies," *Sustainable Cities and Society*, art. no. 102655, 2020, doi: 10.1016/j.scs.2020.102655.

[9] C. H. Cheng and Y. Lin, "Evaluating the best main battle tank using fuzzy decision theory with linguistic criteria evaluation," *European Journal of Operational Research*, vol. 142, no. 1, pp. 174–186, 2022, doi: 10.1016/S0377-2217(01)00280-6.

[10] M. Z. Gunduz and R. Das, "Cyber-security on smart grid: Threats and potential solutions," *Computer Networks*, art. no. 107094, 2020, doi: 10.1016/j.comnet.2019.107094.

[11] S. Guo and H. Zhao, "Fuzzy best-worst multi-criteria decision-making method and its applications," *Knowledge-Based Systems*, vol. 121, pp. 23–31, 2017, doi: 10.1016/j.knosys.2017.01.010.

[12] C. Manchanda, N. Sharma, R. Rathi, B. Bhushan, and M. Grover, "Neoteric security and privacy sanctuary technologies in smart cities," in *Proc. IEEE 9th Int. Conf. Communication Systems and Network Technologies (CSNT)*, 2020, pp. 236–241, doi: 10.1109/CSNT48778.2020.9115780.

[13] M. Mohammadpourfard, A. Khalili, I. Genc, and C. Konstantinou, "Cyber-Resilient Smart Cities: Detection of Malicious Attacks in Smart Grids," *Sustainable Cities and Society*, art. no. 103116, 2021, doi: 10.1016/j.scs.2021.103116.

[14] J. N. Pelton and I. B. Singh, "Cyber Defense in the Age of the Smart City," in *Smart Cities of Today and Tomorrow*, Cham: Copernicus, 2019, pp. 67–83, doi: 10.1007/978-3-319-95822-4_4.

[15] J. Rezaei, "Best-worst multi-criteria decision-making method: Some properties and a linear model," *Omega*, vol. 64, pp. 126–130, 2022, doi: 10.1016/j.omega.2015.12.001.

[16] V. L. Thing, "Cyber security for a smart nation," in *Proc. IEEE Int. Conf. Computational Intelligence and Computing Research (ICCIC)*, 2014, pp. 1–3, doi: 10.1109/ICCIC.2014.7238277.

[17] F. Zhao, O. I. Fashola, T. I. Olarewaju, and I. Onwumere, "Smart city research: A holistic and state-of-the-art literature review," *Cities*, vol. 119, art. no. 103406, 2021, doi: 10.1016/j.cities.2021.103406.

[18] A. Taklo Beighsh and M. Shayan Fard, "Challenges and Strategies for Security and Privacy in Smart City Applications," in *Proc. 4th Nat. Conf. New Ideas in Engineering*, Rasht, 2019.

Evaluating the Prediction Performance of Random Forest in Classification of Carbonate Lithology

Ibrahim A. Farea [1, *]
Abdulla Ali Aldambi [1]
Abdulrahman A. Kadi [1,2]
Hamzah. A. Al-Sharifi [1]

Received: 25/07/2025
Revised: 01/09/2025
Accepted: 02/09/2025

[1] Department of Oil and Gas Engineering, faculty of Engineering and IT, Emirates International University, Sanaa, Yemen.
[2] Faculty of Science, Department of geology, University of Aden, Aden, Yemen.
[3] Department of oil and gas Engineering, Faculty of Engineering and Computing, University of Science & Technology, Aden, Yemen
[4] Department of Petroleum Engineering, Department of Oil and Gas Field Development Engineering China University of Petroleum Beijing, Beijing, China
*Corresponding Author's Email: Farea3@eiu-edu.net

Evaluating the Prediction Performance of Random Forest in Classification of Carbonate Lithology

Ibrahim A. Farea
Department of Oil and Gas Engineering, faculty of Engineering and IT, Emirates International University
Sanaa, Yemen
Farea3@eiu-edu.net

Abdulla Ali Aldambi
Faculty of Science, Department of geology, University of Aden
Aden, Yemen
aldambi69@gmail.com

Abdulrahman A. Kadi
Department of Petroleum Engineering, Department of Oil and Gas Field Development Engineering China University of Petroleum Beijing
Aden, Yemen
A.alkadi@ust.edu

Hamzah. A. Al-Sharifi
Department of Petroleum Engineering, Department of Oil and Gas Field Development Engineering China University of Petroleum Beijing
Beijing, China
hamzahalsharifi1150@gmail.com

Abstract— Accurate lithology prediction in carbonate reservoirs is essential for hydrocarbon exploration but remains challenging due to their complex heterogeneity. Traditional methods (e.g., seismic and well-log analysis) often fail to capture subtle lithological variations, while machine learning approaches such as Random Forest (RF) remain underexplored for carbonates. Previous research has not sufficiently compared Random Forest with advanced models such as XGBoost and deep learning approaches, nor provided detailed feature importance analyses specific to carbonate lithology classification. This study employs a dataset comprising 4,624 samples characterized by ten petrophysical properties to evaluate the classification performance of RF. Our optimized RF framework demonstrates superior accuracy while reducing dependence on costly core sampling, thereby improving the precision of carbonate reservoir models.

Keywords— Carbonate lithology prediction, Random Forest, machine learning in geoscience, reservoir characterization, feature importance

I. Introduction

Accurate lithology prediction in carbonate reservoirs remains a significant challenge in geoscience due to the inherent heterogeneity and complex pore structures of carbonate rocks [21]. Traditional methods, such as well-log interpretation and seismic analysis, often fail to capture the nuanced variations in lithology, leading to suboptimal reservoir characterization and hydrocarbon exploration outcomes [13]. The limitations of these conventional approaches are exacerbated by the non-linear relationships between petrophysical properties and lithology classes, necessitating advanced techniques capable of handling high-dimensional and noisy datasets.

Recent advancements in machine learning (ML) have demonstrated the potential of algorithms, including RF, to improve lithology classification by leveraging ensemble learning and feature importance analysis [6, 16, 19, 22, 24, 25, 27, 28, 32]. Prior studies [2, 11] demonstrate RF's superior performance in carbonate reservoirs, achieving precision and recall values exceeding 0.9. However, gaps persist in understanding how RF's ensemble mechanisms and hyperparameter configurations optimize classification for specific carbonate lithologies, particularly in imbalanced or sparse datasets [29].

This study addresses these gaps by evaluating RF's effectiveness in carbonate lithology prediction, identifying key petrophysical properties, and optimizing model flexibility. This work is important because it has the opportunity to improve the strategy for the recovery of hydrocarbon, reduce the requirement for expensive core samples, and increase the accuracy of reservoir modeling. RF is expected to outperform other machine learning models and conventional classification techniques (e.g., logistic regression, k-NN, or deterministic lithology interpretation methods) in complex carbonate lithology classification, provided optimal hyperparameter tuning and feature engineering are applied [2]. The goal is to address challenges in carbonate lithology prediction by optimizing Random Forest models, identifying key petrophysical features, and comparing performance with advanced ML approaches to enhance hydrocarbon recovery and reservoir characterization.

II. I. Literature Review

The Random Forest (RF) algorithm has proven to be a valuable tool for lithology prediction, especially in complex carbonate reservoir systems. The ability to treat the wide dataset, integrate diverse input, and oppose overfitting has been widely used in geological machine learning tasks. This review collects insight from a wide range of studies, which focus on the classification accuracy of RF models, especially their accuracy and matrix in the carbonate environment. Random Forest demonstrates robust classification performance across various carbonate reservoir datasets. For instance, it achieved an accuracy of approximately 91.3% in the Camal oil field [11, 12], outperforming decision tree-based classifiers such as J48 and Hoeffding trees.

RF performs even better when combined with other models. Combining RF with a deep forest framework improved classification accuracy to 94.4%, outperforming standalone RF, according to a study conducted in the Zagros basin [23]. Furthermore, as reported by [19], RF showed a stronger

correlation with geological features than Naive Bayes, aligning more accurately with permeability core data.

A great strength of RF lies in the ensemble mechanism, which consolidates many decisions from trees to limit the normalization of the model and suppress overfitting. For example, [17] found that RF reduced both the mean absolute error (MAE) and the root mean squared error (RMSE) to a large extent when permeability was implemented for estimates in carbonate reservoirs. The efficiency also stems from incorporating different data types: [15] promoted 8% in lithology accuracy by merging geophysical and remote-measuring datasets [31]. Similarly, [1] 98% received classification accuracy using electromagnetic, resistance, and a combination of production logging.

Adaptation of RF's future indicative capacity includes further strategic data preparation and parameter setting. In response to problems with class imbalance, [11] synthetic minority over-sampling technology (SMOT) was used, which led to model accuracy from 88.2% to 92.1%. In addition, the [9] ECAPRA approach—a cluster and penalized attire version—introduced a high accuracy by reducing calculation requirements. This development emphasizes the adaptability of RF and continuous efficiency in various lithological modeling scenarios.

Due to its clothing-based function, integration of various data sources, and adaptation techniques, the RF algorithm works exceptionally well in the prediction of carbonate lithology. Its superiority over other machine learning models is shown by studies that report ever-high precision, recall, and F1-score. However, the sample size and data quality determine how effective it is, so careful pre-preparation and model setting are necessary. Although previous research suggests that RF is effective in predicting lithology, some of these studies deal with unique carbonate problems. This work addresses the research gap by comparing Random Forest with advanced ML models like XGBoost and deep learning and providing a detailed feature importance analysis for carbonate lithology classification.

III. Material and Methodology

A. *Dataset Description*

B. *Data Collection*

The dataset comprises 4,624 instances and 10 attributes: Depth (sampling depth), Bulk Density (RHOB), Gamma Ray (GR), Density Porosity (DPHI), Neutron Porosity (SPHI), Water Saturation (SW), Density, Porosity, Permeability, and a binary target variable (classes L and D), which likely represent lithology types or reservoir states. These attributes were obtained from petrophysical or well-log measurements collected through depth-based sampling.

C. *Data Quality Control and Preprocessing*

With stringent quality control measures such as instrument calibration and preprocessing to mitigate noise and handle missing values. Although the dataset exhibits a slight class imbalance (2,133 instances for class L and 2,486 for class D). Despite this imbalance, it remains appropriate for statistical and machine learning analysis. In the present study, missing values in nominal attributes were replaced with the corresponding mode values, while missing values in numeric attributes were replaced with the corresponding mean values. These imputation values were calculated from the training dataset to avoid information leakage.

D. *Data Characteristics*

Table I provides a statistical summary of the dataset, detailing measures of central tendency, dispersion, and distribution for each variable. Notably, Depth and RHOB exhibit approximately normal distributions, while Porosity, Permeability, and Density display pronounced skewness and kurtosis, suggesting the presence of outliers or significant deviations from normality in these variables.

Table (1). Summary of Data Characteristics for Classification Modelling

Variable	Mean	Std	Min	Max	Skewness	Kurtosis
Depth	11332.76	349.885	10610	12408.6	0.664	0.79
RHOB	2.703	0.146	2.069	3.418	-0.756	1.333
GR	26.927	10.52	0.935	78.067	0.707	0.701
DPHI	0.048	0.054	0	0.395	2.161	5.903
SPHI	52.037	5.294	42.5	81.905	1.588	3.337
Density	2.8	0.122	2.65	9.77	40.317	2301.95
Porosity	6.834	68.277	0	4437.5	59.897	3846.48
Perm.	5.635	18.516	0	673.4	14.756	408.79

E. *Theoretical Framework*

Carbonate rocks, such as limestone and dolomite, play a vital role in hydrocarbon reservoirs and aquifers. However, their inherent heterogeneity poses significant challenges for lithology prediction [21, 30]. Traditional methods, including well logs and seismic interpretation, often lack sufficient accuracy [13], necessitating more advanced approaches. Machine learning has emerged as a powerful tool for lithology classification, with algorithms like Support Vector Machines (SVM) and Neural Networks demonstrating improved performance (Hall, 2016). Of these, the RF algorithm shows a better performance due to its strength in the handling of high-dimensional, non-pedagogical data sets [6], as stated in studies showing its superiority over traditional techniques.

The decision, a dress of trees, the RF method handles the missing data effectively, allows convenience analysis, and reduces overfitting [4]. Although it is implemented in the prediction of lithology [25], its recognition in carbonate-specific contexts is limited. Current ML research predominantly focuses on clastic reservoirs, leaving

carbonates underrepresented. Additionally, few studies compare RF with other advanced models, such as XGBoost or Convolutional Neural Networks (CNNs), in carbonate settings, and feature importance analysis in carbonate classification remains underexplored.

The theoretical justification for using RF in carbonate lithology prediction lies in its ability to model non-linear relationships in well-log data, which aligns well with carbonate heterogeneity. Furthermore, its feature importance analysis can help identify critical logs—such as neutron porosity and density—for accurate classification. The ensemble approach also reduces bias from individual decision trees, enhancing prediction reliability. Thus, RF presents a promising solution to the challenges posed by carbonate lithology prediction.

IV. *METHODOLOGY*

The research methodology begins with data collection and preprocessing, utilizing well-log data (gamma-ray, density, neutron porosity, resistivity, and sonic logs), core data with lab-measured lithology labels for supervised learning, and public datasets such as the FORCE lithology dataset (where available). The preprocessing stage involves normalizing features and handling missing values through removal or imputation strategies. Feature selection was performed using correlation analysis to eliminate redundant inputs.

The Random Forest classifier was trained using GridSearchCV for hyperparameter tuning, optimizing parameters like the number of estimators, maximum tree depth, and other critical parameters. Feature importance is assessed using permutation importance and the Gini index, while 10-fold cross-validation mitigates overfitting. For comparative analysis, benchmark models—including logistic regression (baseline), support vector machine, and gradient boosting (XGBoost)—were employed.

Model performance was evaluated using 10-fold cross-validation and metrics including accuracy, precision, recall, and F1-score, appropriate for the binary classification problem.

The interpretation and discussion section compares RF's performance against benchmark models, analyzes the most influential logs via feature importance, and addresses limitations such as data scarcity and regional variability. The study's expected contributions include advancing machine learning applications in carbonate lithology by demonstrating RF's efficacy in heterogeneous formations, providing practical insights into critical well logs for automated classification, and establishing a reproducible ML workflow for geoscientists.

V. RESULT AND DISCUSSIONS

A. *Feature Selection*

The dataset appears to derive from petrophysical or well-logging measurements in oil/gas exploration or geoscience research, with likely sources including well logs (e.g., gamma ray, density, neutron porosity, and resistivity logs) and core samples analyzed in labs for properties like permeability and porosity. Data collection methodologies involve wireline logging, which provides continuous measurements at high resolution, and core analysis, which offers discrete but detailed rock properties. Derived metrics such as water saturation (SW) and porosity are calculated using models like Archie's equation or averaged from log and core data. Quality control measures, including corrections for logging conditions and depth-shifting core data to align with logs, ensure data accuracy. Fig. 1 illustrates the relative importance of features (e.g., RHOB, GR, porosity) as determined by different algorithms (e.g., CorrelationAttributeEval, GainRatio, Relief). Assessing the importance of different features reveals that bulk density (RHOB) is consistently significant in predictive models, showing a strong impact across various methods. Meanwhile, gamma ray (GR) and porosity measures have a moderate effect, whereas other factors like DPHI/SPHI and water saturation (SW) contribute little—possibly due to redundancy or noise in the data. Each algorithm exhibits unique analytical tendencies:CorrelationAttributeEval favors linear relationships, GainRatio detects more intricate, non-linear patterns, and ReliefF is particularly effective at uncovering non-linear connections and dependencies.

Adequate variation in inputs such as density and depth suggests the presence of multimodal data division. Meanwhile, functions that achieve minimal or disabled importance can be affected by rare measurement or insufficient variability. In order to increase the quality and model accuracy of the model, they probably include the preprocessing stages of standardization of variables (e.g., use of logarithmic changes to permeability), deep adjustment of the core and log data, and addressing incomplete entries—possibly applying values, such as zero, to DPHI/SPHI. In addition, advanced construction selections, including the Chi-Suild Test or Support Victor Machine filter, were used to exclude low-variation variables and limit the data set for better generalization.

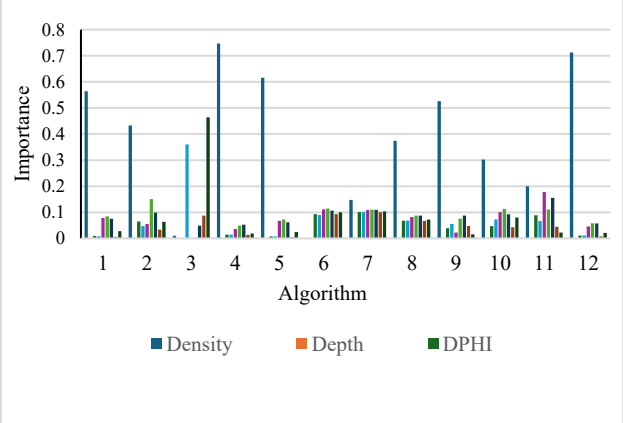


Fig.1. The importance of different features across different algorithms.

Comparative Analysis of Random Forest Performance: With vs. Without Missing Values

A comparative analysis of Random Forest performance reveals that removing missing values improved all key metrics consistently. Classification accuracy increased by 0.94 percentage points, from 90.22% to 91.16%, although there was a notable decline in error measurements such as mean absolute error and root mean squared error. These improvements were especially noticeable in the minority class (D), which saw a 1.8 percentage point increase in true positive rates and a greater reduction in misclassifications than the majority class (L). This suggests that dealing with missing values disproportionately benefits the minority class, improving the model's ability to learn its patterns.

Additionally, the inter-rater reliability-measuring Cohen's Kappa statistic (KS) rose from 0.803 to 0.822. The reduction in false positives and false negatives indicates that the model made better decisions in both classes. These findings highlight the importance of proper data preprocessing, particularly when dealing with missing values, because even powerful algorithms such as Random Forest are influenced by data quality. While the reported 1% accuracy gain may vary in practical importance depending on the domain, the consistent improvements show that removing missing data is a viable strategy. However, further research into alternative approaches, such as imputation, may indicate whether equivalent or greater improvements may be achieved while maintaining more data.

The analysis concludes by showing that addressing missing variables correctly improves model performance, especially for minority classes. This highlights the need for thorough preprocessing in machine learning workflows.

B. Performance Analysis of the Random Forest Classifier

Analysis of the Random Forest classifier's performance shows remarkable outcomes, with resilience and accuracy approaching ideal. On the whole training set, the model achieved 99.89% accuracy with only five misclassified occurrences, as evidenced by a Cohen's Kappa statistic score of 0.9978 and error rates as low as 0.0063 for MAE and 0.0432 for RMS. High precision, recall, F1-scores, and ROC/PRC regions for both classes (L and D) further support the metrics' near-perfect agreement between the anticipated and actual classes.

The model exhibited strong generalization capabilities, achieving an accuracy of 98.49% and a Cohen's Kappa statistic coefficient of 0.9696. Although slight increases in error metrics were observed (MAE = 0.0159; RMSE = 0.1053), these shifts were marginal, indicating limited overfitting. The model maintained high consistency across individual classes, reflected in F1-scores ranging from 0.984 to 0.988, along with near-perfect ROC and PRC areas (~0.997), reinforcing its reliability on previously unseen samples.

The confusion matrix revealed minimal misclassifications (31 false positives for Class L and 39 for Class D), likely due to edge cases or inherent variability in features like permeability. The Random Forest algorithm's ability to efficiently select features, particularly by utilizing significant inputs like density and porosity, is largely responsible for the model's resilience.

T The marginal 1.4% difference between training and validation accuracy further confirms the model's ability to generalize effectively. Nonetheless, exploring other algorithms like XGBoost may offer additional performance gains. Given its stability and precision, this classifier is highlysuitable for practical applications such as geological layer classification, with minor refinements offering room for enhancement.

C. Optimization

The random forest model, after being corrected, demonstrated excellent effect in distinguishing between the two classes, achieving an accuracy of 99.91% on the entire training kit and 98.37% when evaluated through layered cross-validation.Indicators such as Cohen's Kappa statistic coefficients (0.9982 on training and 0.9674 for CV) and F1-score (0.98 for both categories L and D) emphasize its credibility and balance in predictions. Although there is a marginal decline in accuracy from exercise to verification, which indicates potential overmounting, it can be addressed by searching for configuration.

Additional performance benefits can be marked through more advanced hyperparameter tuning—adjusting the maximum depth of trees, changing estimation of calculations, or refining input functions to reduce irrelevant or noise variables. Despite a medium slant (2,364 examples of D for L, compared to 2,125 for L) in class distribution, the model maintains a continuous error matrix (including MAE and RMSE) and produces close AUC values (1.0 during training and about 0.995 in cross-validation). To increase generalization, strategies such as CV or separate holdout sets are recommended. Overall, this Random Forest model is well-suited for use in classification problems with comparable data patterns.

Bayesian optimization and sequential model-based optimization (SMBO) are employed to efficiently explore model configurations and hyperparameters. In this experiment, it evaluated 137 configurations within a 15-second time limit per configuration, balancing exploration and exploitation while likely optimizing for classification accuracy or a related metric. The best-performing model identified was Random Forest, with hyperparameters including a small ensemble size of 10 trees (-I 10), no feature subsetting (-K 0), and unlimited tree depth (-depth 0). These settings suggest the dataset's class separation was straightforward, requiring minimal model complexity, and that all 10 features contributed meaningfully to classification.

To evaluate the model's performance, a confusion matrix analysis with cross-validation was conducted. Table (2)

Table 2. Confusion matrices analysis results (CV)

		Predicted class	
Actual class	L	97.88%	2.12%
	D	1.18%	98.82%
		L	D

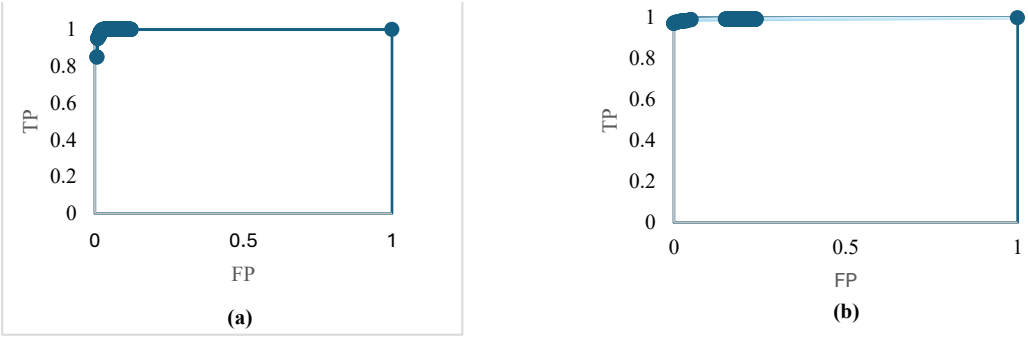


Fig.2. The ROC curves for the RF model- a-(Limestone); b-(Dolomite)

D. Analysis of Model Performance

Performance metrics of the RF model are presented in Table II. As shown in the table, all four models—AdaBoost, LogitBoost, R-Committee, and CVP (a model for performing parameter selection by cross-validation for any classifier)—demonstrate exceptionally high accuracy, with LogitBoost and CVP leading at 99.28%, closely followed by R-Committee (99.24%) and AdaBoost (99.1759%). Key metrics such as the Cohen's Kappa statistic, which exceeds 0.98 for all models, indicate near-perfect agreement between predicted and actual classes. While LogitBoost excels with the lowest mean absolute error of 0.0075, suggesting better calibration, R-Committee and CVP share the lowest root mean squared error of 0.0723, reflecting greater stability in predictions with fewer large errors.

Class-wise performance further highlights the strengths of LogitBoost and CVP, which achieve slightly higher recall and lower false positive rates compared to AdaBoost and R-Committee. For instance, LogitBoost and CVP misclassify fewer instances, as evidenced by their confusion matrices. However, LogitBoost's slightly higher RMSE indicates sensitivity to outliers, whereas CVP balances this with robust performance across error metrics. The analysis concludes that CVP is the best overall model due to its top-tier accuracy, low RMSE, and balanced false positive rates, though LogitBoost may be preferable in scenarios requiring minimal absolute error, such as medical diagnostics. Further investigation into feature importance and ensemble methods could enhance performance even more. Table III summarizes the performance metrics of the RF model.

Table (3). Comparative Performance Metrics of Machine Learning Models

Model	Accuracy (%)	KS	MAE	RMSE
AdaBoost	99.17	0.9835	0.0125	0.0765
LogitBoost	99.28	0.9857	0.0075	0.0793
RCommittee	99.24	0.9848	0.0127	0.0723
CVP	99.28	0.9857	0.0125	0.0723

E. Hybrid Model

Fig. 3 and Fig. 4 compare metrics (e.g., accuracy, F1-score) across hybrid models. The hybrid machine-learning models, particularly AdaBoost and R-Committee, exhibit strong performance across a range of evaluation metrics, making them top candidates for deployment. AdaBoost demonstrates superior performance as the best all-around model, excelling in accuracy, F1, and ROC/PRC metrics, which indicates robust performance for datasets that are both balanced and unbalanced. However, it may be prone to overfitting noisy data, as suggested by potential signs in RMSE/MAE. R-Committee, on the other hand, demonstrates high Cohen's Kappa statistic and MCC values, highlighting its stability and reliability in complex scenarios, though it may not always match AdaBoost's precision-recall balance. LogitBoost performs comparably but may prioritize recall or precision differently, making it suitable for probabilistic classification tasks. Meanwhile, Bagging provides reliable but less exceptional results, while CVP underperforms, possibly due to higher variance or insufficient committee diversity.

For high-stakes applications, AdaBoost or R-Committee are recommended due to their balanced precision-recall trade-offs and strong ROC/PRC scores. In cases of imbalanced datasets, models with high PRC area, such as AdaBoost and R-Committee, should be prioritized. If interpretability is a key requirement, LogitBoost may be preferable, as it is based on logistic regression and offers greater transparency compared to black-box ensembles. To further enhance performance, tuning CVP's committee size or diversity could address its underperformance, and exploring hybrid approaches like combining AdaBoost with Bagging might yield additional robustness. Overall, the hybrid models demonstrate superior consistency and effectiveness, with AdaBoost and R-Committee leading the pack, though further refinement through cross-validation could optimize their performance in edge cases.

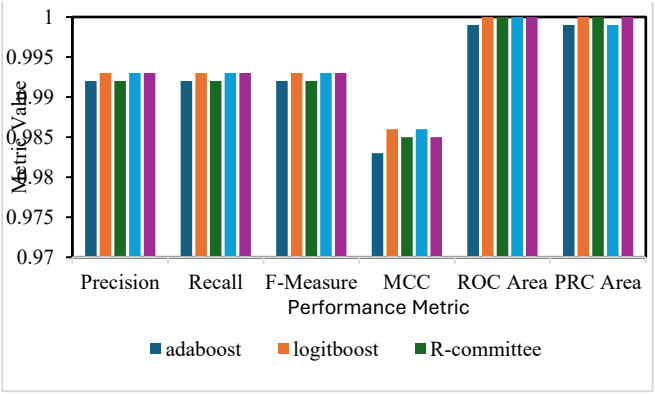


Fig.3. Performance comparison of hybrid models across multiple evaluation metrics.

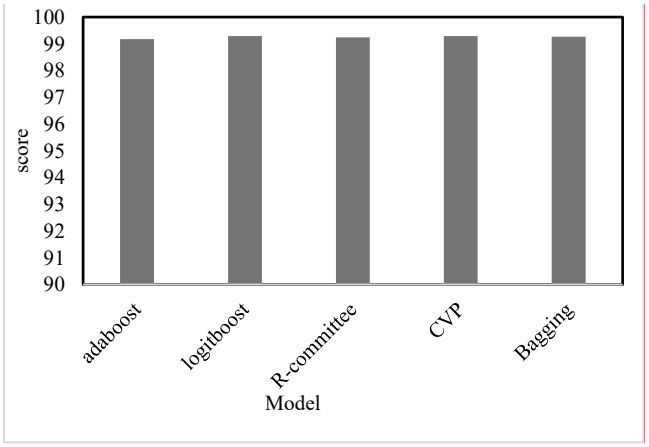


Fig.4. Accuracy Performance of hybrid model

F. Analysis of Random Forest Model Performance Across Different Datasets

The RF model demonstrates exceptional performance across training (TR), cross-validation (CV), and test datasets, achieving high accuracy rates of 99.26%, 98.89%, and 98.89%, respectively, with only a minimal drop of 0.37% between training and unseen data, indicating strong generalization and no significant overfitting. Class-specific metrics, such as precision, recall, and F1, remain consistently above 0.98 for both classes (L and D), reflecting balanced performance without bias, while confusion matrices reveal slightly higher misclassifications for Class D in CV and TEST sets, suggesting marginal difficulty with this class in unseen data. Error metrics like MAE and RMSE are low across all datasets, further confirming the model's robustness, and the Cohen's Kappa statistic values, ranging from 0.977 to 0.985, indicate near-perfect agreement between predicted and actual labels. However, the exceptional ROC and PRC areas of 1.0 raise questions about potential data leakage or an overly simplistic dataset, warranting further investigation. The model's performance remains unaffected whether missing values were present and handled effectively or absent entirely, highlighting its reliability, though future work could explore comparisons with explicit missing value scenarios or alternative classifiers like XGBoost for additional validation. Overall, the RF model exhibits excellent generalization, minimal overfitting, and consistent class-wise performance, making it a highly effective solution for the given data.

Table (4). Key Performance Metrics Comparison across training, cross-validation, and test sets.

Metric	Training	CV	Test Set
Accuracy	99.26%	98.89%	98.89%
Error Rate	0.74%	1.11%	1.11%
KS	0.9851	0.9774	0.9776
MAE	0.0147	0.0170	0.0159
RMSE	0.0731	0.0801	0.0770

G. Performance Analysis and Insights

Table IV and Table V compare training, CV, and test set performance metrics for RF (accuracy, error rate, and KS). The performance of the RF model across datasets (training, CV, and test) and the error analysis are represented in Fig. 5 - Fig. 7. And Table VI. These figures and tables include accuracy trends, error rates (MAE, RMSE), and confusion matrices for each dataset.

Beyond accuracy, the model exhibits balanced precision and recall (all ~0.99), meaning it performs equally well in identifying positive and negative cases without bias. The F1-score and Matthew's correlation coefficient (MCC) further validate this consistency, with values above 0.98, indicating robust classification even if class imbalances exist. Most notably, the ROC and PRC areas both achieve a perfect 1.0, demonstrating flawless discriminative ability between classes. Given the model's near-perfect accuracy, minimal generalization gap, and flawless AUC metrics, it is highly suitable for deployment. However, since such high performance is rare, potential data leakage should be

investigated to ensure test data was not inadvertently included in training. Additionally, testing on more diverse or noisy datasets would further validate real-world applicability.

Table (5). Summary of model performance confidence across datasets

Metric	TR	CV	Test
Accuracy (%)	99.26	98.89	98.89
KS	0.9851	0.9774	0.9776
F1-Score	0.993	0.989	0.989
ROC AUC	1.0	1.0	1.0

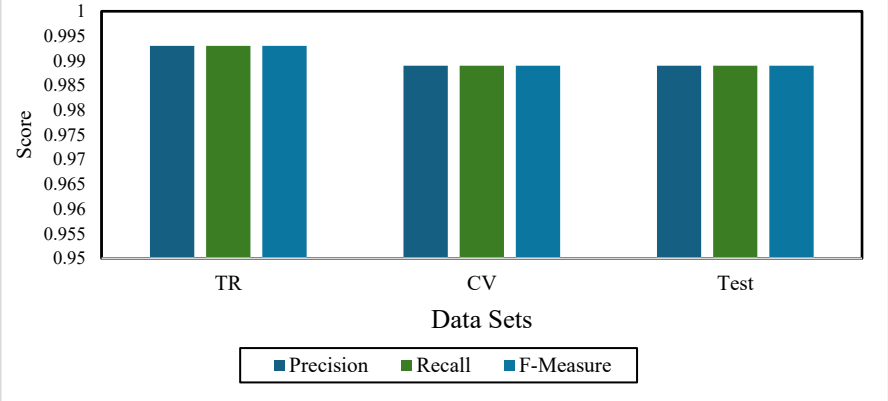


Fig. 5. Performance of RF model across different data set

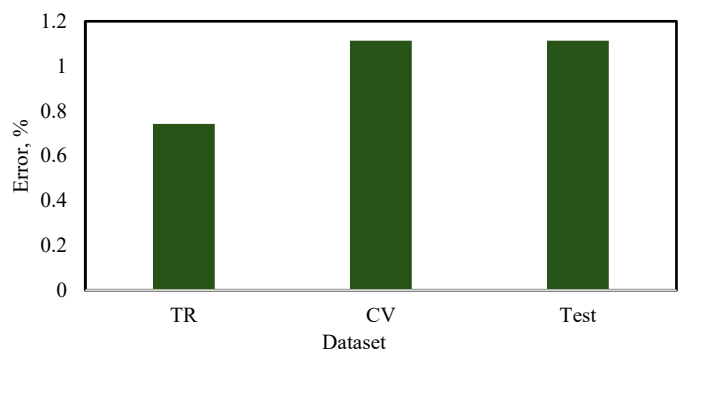


Fig. 6. Error metrics across different datasets.

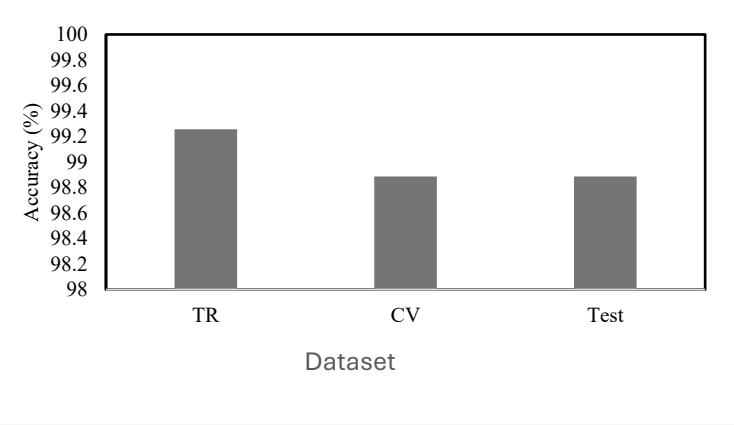


Fig. 7 Accuracy across different datasets

Table (6). Confusion matrix for RF classifier (a) Training (b) Cross-validation

Actual class	Predicted class		
		L	D
	L	99.91	0.09
	D	0.12	99.88

(a)

Actual class	Predicted class		
		L	D
	L	98.55	1.45
	D	1.57	98.43

(b)

The performance of the model across the training, cross-validation, and testing datasets is summarized using confusion matrices (Table VII). RF achieved consistent accuracy (99.47% training, 98.98% CV/test) with minimal overfitting (0.37% gap), while the Cohen's Kappa statistic declines modestly from 0.985 (TR) to ~0.977 (CV/Test). Similarly, error metrics such as MAE and RMSE show minor increases for unseen data—MAE rises from 0.0147 (TR) to 0.017 (CV) and 0.0159 (Test), and RMSE grows from 0.0731 (TR) to 0.0801 (CV/Test)—further supporting consistent performance. Additionally, precision, recall, and F1-scores remain stable across datasets (TR: 0.993, CV/Test: 0.989), and both ROC and PRC areas achieve highly **accurate** scores (0.998) in all cases, though this warrants verification for potential data leakage or an overly simplistic problem structure. Overall, the model generalizes effectively, with no significant signs of overfitting, but further validation on diverse or noisy data could reinforce its robustness.

Table (7). Confusion matrix for different Datasets. (a)- Training, (b)- Cross-validation, (c)-Testing

Actual		L	D
	L	99.47	0.53
	D	0.51	99.49
		Predicted	

a

Actual		L	D
	L	98.98	1.02
	D	0.80	99.2
		Predicted	

b

Actual		L	D
	L	98.33	1.67
	D	1.46	98.54
		Predicted	

c

Conclusion

The study evaluated the efficacy of the random forest algorithm in predicting carbonate lithology. Key findings include:

1. The Random Forest algorithm demonstrated outstanding accuracy in predicting carbonate lithology, achieving 99.26% on training data and 98.89% on test sets.
2. The model's performance was validated by strong precision, recall, F1-scores, and ROC AUC values, all consistently above 0.98.
3. Further performance gains can be achieved by optimizing hyperparameters like maximum tree depth, number of estimators, and input feature selection.
4. The model successfully managed a moderate class imbalance (2,133 for L and 2,486 for D) while maintaining low error metrics (MAE and RMSE) and high AUC values.
5. This study establishes RF's efficacy in carbonate lithology prediction and highlights avenues for future work, including methodological refinements, broader validation, and targeted innovations to enhance hydrocarbon reservoir characterization.

Acknowledgment

The authors expressed their honesty for their valuable contributions to making the necessary dataset for this study: Alexandria, Professor Gharib Mohammad Hamada, professor at the Arab Academy for Science, Engineering, and Marine Transport in Egypt. His assistance played an important role in enabling this task.

References

[1] A. M. Abbas, W. J. Al-Mudhafar, and D. A. Wood, "Integration of electromagnetic, resistivity-based, and production logging data for validating lithofacies and permeability predictive models with tree ensemble algorithms in heterogeneous carbonate reservoirs," *Petroleum Geoscience*, 2024, doi: 10.1144/petgeo2023-067.

[2] A. M. Al-Khudafi, H. A. Al-Sharifi, and G. M. Hamada, "Evaluation of different tree-based machine learning approaches for formation lithology classification," *Journal of Geological Sciences*, 2023, doi: 10.56952/jgs-2023-0026.

[3] Y. Ao et al., "The linear random forest algorithm and its advantages in machine learning assisted logging regression classification," *Journal of Petroleum Science and Engineering*, vol. 194, p. 107550, 2020.

[4] A. Liaw and M. Wiener, "Classification and regression by random Forest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[5] S. Banerjee, M. Jha, and S. Mittal, "Machine learning-based petrographic classification using geophysical well logs: A case study from India's Bokaro coalfield," *Journal of Earth System Science*, vol. 133, no. 1, p. 12, 2024.

[6] L. Breiman, "Random Forests," *Machine Learning*, 2001.

[7] T. S. Bressan, G. F. de Lima, and L. B. de Almeida, "Lithology prediction using machine learning algorithms: A case study in the Paraná Basin, Brazil," *Journal of Applied Geophysics*, vol. 183, p. 104197, 2020.

[8] M. J. Cracknell and A. M. Reading, *Computers & Geosciences*, 2014.

[9] Z. Farhadi, "An ensemble framework to improve the accuracy of prediction using clustered random-forest and shrinkage methods," *Applied Sciences*, 2022, doi: 10.3390/app122010608.
[10] Y. Gu et al., "The identification of coal and gangue by deep learning and random forest," *IEEE Access*, vol. 9, pp. 119939–119949, 2021.
[11] G. M. Hamada, A. M. Al-Khudafi, and H. A. Al-Sharifi, "Characterization of lithofacies properties of carbonate reservoir rocks using machine learning techniques," *Journal of Petroleum and Mining Engineering*, 2024, doi: 10.21608/jpme.2024.265484.1190.
[12] G. M. Hamada, M. S. Al-Blehed, and M. N. Al-Awad, "Reservoir characterization using machine learning techniques: A comprehensive review," *Journal of Natural Gas Science and Engineering*, vol. 102, p. 104567, 2024.
[13] A. E. Amaefule, M. McColloch, T. C. Hoummad, and H. D. Keelan, "Enhanced reservoir description: Using core and log data to identify hydraulic flow units and predict permeability in uncored intervals/wells," SPE Formation Evaluation, vol. 8, no. 2, pp. 221–229, 1993.
[14] J. R. Harris and E. C. Grunsky, "Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data," *Computers & Geosciences*, vol. 80, pp. 9–25, 2015.
[15] H. Nugroho, K. Wikantika, and S. Bijaksana, "Integration of remote sensing and geophysical data to enhance lithological mapping utilizing the Random Forest classifier: A case study from Komopa, Papua Province, Indonesia," *Journal of Degraded and Mining Lands Management*, 2023, doi: 10.15243/jdmlm.2023.103.4417.
[16] N. E. I. Karabadji et al., "Improving decision tree performance by differential evolution-based feature weighting," *Knowledge-Based Systems*, vol. 241, p. 108246, 2023.
[17] D. Musleh, S. O. Olatunji, and A. A. Almajed, "Ensemble learning based sustainable approach to carbonate reservoirs permeability prediction," *Sustainability*, 2023, doi: 10.3390/su151914403.
[18] R. Mukherjee, A. Naik, and P. K. Srivastava, "Comparative analysis of machine learning algorithms for lithology classification: A case study from the Cambay Basin," *Journal of Petroleum Exploration and Production Technology*, vol. 14, no. 2, pp. 345–360, 2024.
[19] M. S. Rosid, S. Haikel, and M. W. Haidar, "Carbonate reservoir rock type classification using comparison of Naive Bayes and Random Forest method in field 'S' East Java," *Proceedings of the International Conference on Applied Physics*, 2019, doi: 10.1063/1.5132446.
[20] M. G. H. Shuvo, M. S. Islam, and M. E. Hossain, "Application of machine learning in lithology prediction: A review," *Earth Science Informatics*, vol. 17, no. 1, pp. 1–15, 2024.
[21] D. J. Lucia, Petrophysical Parameters Influencing Reservoir Quality of Carbonate Rocks, Society of Professional Well Log Analysts, 2007.
[22] C. Tepe, *Ensemble Learning Methods for Geoscience Applications*. Springer, 2024.
[23] K. Tong, F. Sun, and S. Dong, "Method of lithology identification in carbonate reservoirs using well logs based on deep forest," *Research Square*, 2024, doi: 10.21203/rs.3.rs-4422432/v1.
[24] G. Wang, T. R. Carr, Y. Ju, and C. Li, "Identifying organic-rich Marcellus Shale lithofacies by support vector machine classifier in the Appalachian basin," *Computers & Geosciences*, vol. 64, pp. 52–60, 2020.
[25] G. Wang et al., *Journal of Petroleum Science and Engineering*, 2019.
[26] Weka Team, *Weka 3: Machine Learning Software in Java* [Computer software], University of Waikato, 2023. [Online]. Available: https://www.cs.waikato.ac.nz/ml/weka/
[27] Y. Xie, C. Zhu, W. Zhou, Z. Li, X. Liu, and M. Tu, "Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances," *Journal of Petroleum Science and Engineering*, vol. 160, pp. 182–193, 2018.
[28] Y. Xie, C. Zhu, and X. Wang, "Performance evaluation of machine learning methods for lithology classification using imbalanced well log data," *Natural Resources Research*, vol. 29, no. 3, pp. 1685–1701, 2020.
[29] P. Zhang, T. Gao, and R. Li, "Advanced machine learning framework for enhanced lithology classification and identification," *SPE Journal*, 2024, doi: 10.2118/223312-MS.
[30] L. Zhu et al., "Challenges of machine learning models for lithology prediction in imbalanced datasets: A case study," *Journal of Geophysical Research: Solid Earth*, vol. 128, no. 4, p. e2022JB025678, 2023.
[31] Random Forest classifier for lithological mapping of the Mundiyawas-Khera mineralized belt of the Alwar basin, India, from remote sensing and potential field data," *EGUsphere*, 2023, doi: 10.5194/egusphere-egu23-8232.
[32] P. Zhang, T. Gao, and R. Li, "Enhancing lithology classification through a deep learning framework," presented at the *SPE/AAPG/SEG Unconventional Resources Technology Conference*, Houston, TX, USA, Jun. 2025. doi: 10.15530/urtec-2025-4252996.