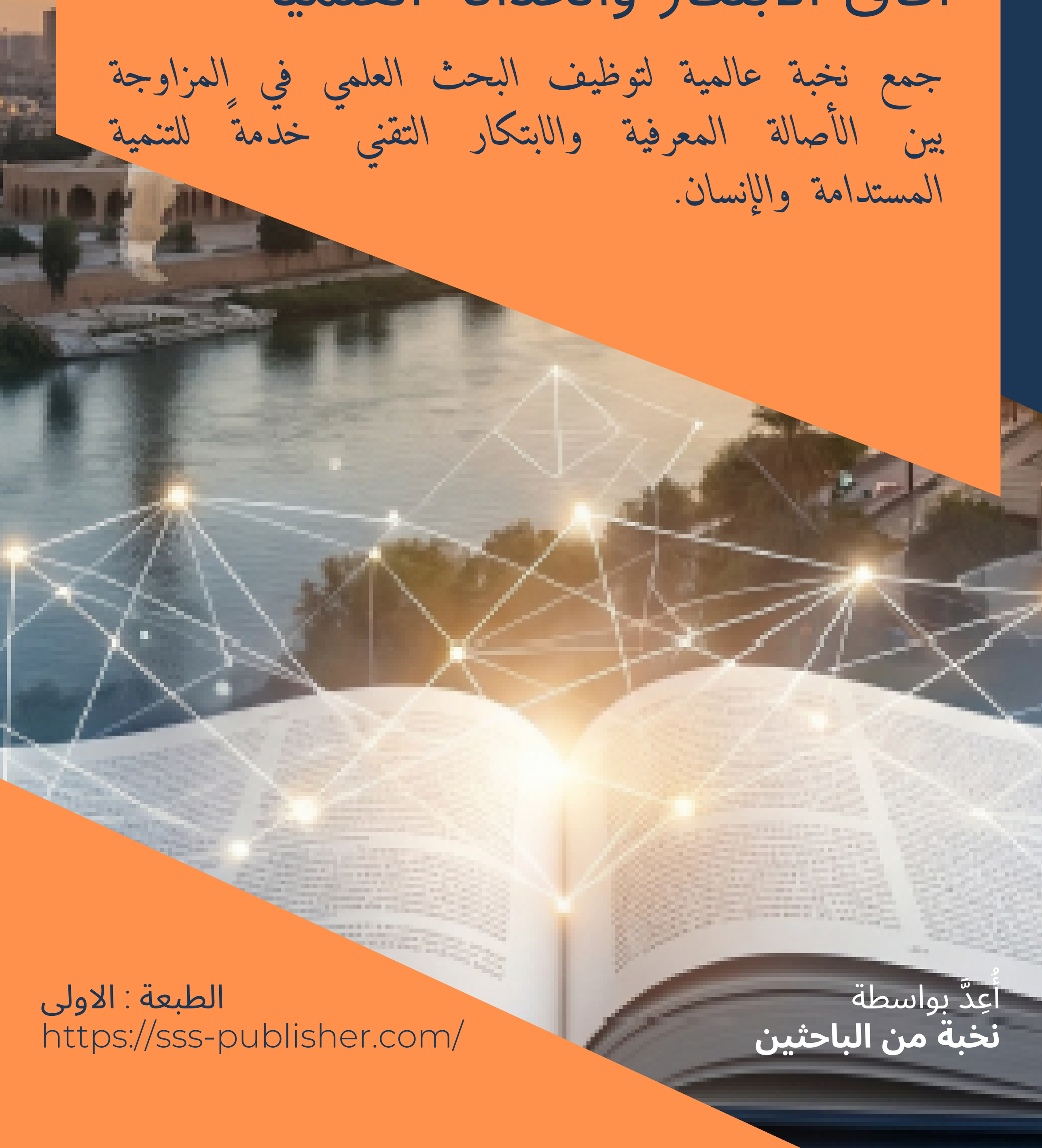
تركيب الرقيم السومري للنشر والتوزيع

2026

# المؤتمر العلمي الدولي الثاني الدوري الموسوم: من الأصالة الإنسانية إلى آفاق الابتكار والحداثة العلمية

جمع نخبة عالمية لتوظيف البحث العلمي في المزاوجة بين الأصالة المعرفية والابتكار التقني خدمةً للتنمية المستدامة والإنسان.




الطبعة : الاولى
https://sss-publisher.com/


أُعِدَّ بواسطة
**نخبة من الباحثين**

المؤتمر العلمي الدولي الثاني الدوري الموسوم: من الأصالة الإنسانية إلى آفاق الابتكار والحداثة العلمية

(الجزء الثاني):

المؤتمر العلمي الدولي الثاني الدوري الموسوم: من الأصالة الإنسانية إلى آفاق الابتكار والحداثة العلمية :(الجزء الثاني)

نخبة من الباحثين

التصنيف : الذكاء الاصطناعي، الحوسبة والأمن السيبراني

الطبعة: الاولى

---

**رقم الايداع في دار الكتب والوثائق في بغداد (بلا) لسنة (بلا)**

**ISBN: 978-9922-7961-61**

**DOI: 10.62909/978.9922.7961.61**

دار تركيب الرقيم السومري للطباعة والتوزيع والنشر الالكتروني

العراق – ديالى – بعقوبه – شارع الاوقاف

هاتف : 07775262892

manager@sss-publisher.com

# المؤتمر العلمي الدولي الثاني الدوري
# الموسوم: من الأصالة الإنسانية إلى آفاق الابتكار والحداثة العلمية :(الجزء الثاني)

## نخبة من الباحثين

2026


تنضيد الكتاب

معهد الترجمة والكتاب

«ماليزيا»

مراجعة لغوية

الدكتور محمد محمد كاظم

«العراق»

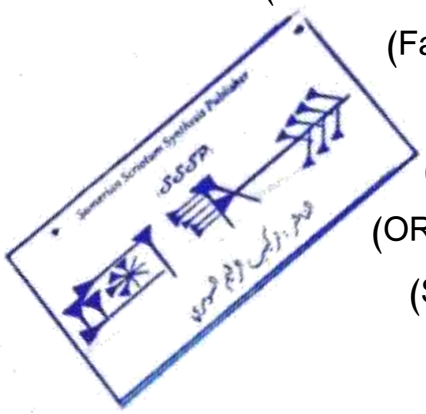
تصميم الغلاف

معهد التكنالوجيا

«الهند»

التوزيـــع والنشـــر الإلكترونـــي ، تمـــت فهرســـة هـــذا الكتـــاب فـــي عـــدد مـــن الفهارس والمنصات الدولية، منها:

| | |
|---|---|
| كروس ريف | (CrossRef) |
| أوبن ألكس | (OpenAlex) |
| مؤشر الاقتباس المفتوح الأوكراني | (OUCI) |
| سجل مستودعات الوصول المفتوح | (ROAR) |
| سيمنتك سكولار | (Semantic Scholar) |
| دايمنشنز | (Dimensions) |
| ريسرش غيت | (ResearchGate) |
| فات كات | (Fatcat) |
| المعيار الدولي للترقيم | (IAS) |
| مؤشر الاستشهاد العلمي | (SCI) |
| أوركيد | (ORCID) |
| سايليت | (Scilit) |



الموقع الرسمي لدار النشر الرقيم السومري للنشر والتوزيع

يتوفر هذا الكتاب عبر الموقع الرسمي لدار النشر، مع تخصيص مُعرّف رقمي دائم (DOI) خاص به: https://doi.org/10.62909/978.9922.7961.61

978-9922-796-16-1

# المؤتمر العلمي الدولي الثاني الدوري

## الموسوم: من الأصالة الإنسانية إلى آفاق الابتكار والحداثة العلمية

### برعاية كريمة وبإشراف السادة الأفاضل:

**الأستاذ الدكتور أسامة إسماعيل المشهداني** رئيس جامعة نينوى.

**الأستاذ الدكتور كوسار محمد علي** رئيس جامعة السليمانية.

**الأستاذ الدكتور وسام صلاح عبد الحسين** رئيس جامعة قرطبة.

**الأستاذ الدكتور حاتم هذال عبدالحميد** مدير عام دائرة البحوث والدراسات.

**الأستاذ رباح إبراهيم آل جعفر** مدير عام دائرة التعليم الديني والدراسات الإسلامية.

**الأستاذ الدكتور صالح عبد المهدي الخالدي** عميد كلية ابن خلدون الجامعة.

**الأستاذة الدكتورة سوزان صابر حيدر** عميد كلية الإدارة والاقتصاد - جامعة السليمانية.

**الدكتور علي مهدي كاظم** عميد كلية العلوم - جامعة قرطبة.

**الدكتور علاء ثابت جاسم** رئيس مؤسسة أجيال الرافدين لتطوير التعليم.

**الأستاذ المساعد الدكتور علي عبودي نعمة** المدير التنفيذي لمؤسسة أجيال الرافدين لتطوير التعليم.

**الدكتور أحمد جميل مجيد المدرس** مدير مركز وعي للاستشارات وبناء القدرات.

## ديباجة المؤتمر

انطلاقاً من الدور المحوري للبحث العلمي ونشر الأبحاث العلمية في معالجة القضايا الراهنة وتقديم الحلول المستقبلية وسعياً نحو تعزيز الحوار الأكاديمي البنّاء بين مختلف التخصصات يسرنا أن نعلن عن عقد هذا المؤتمر العلمي الدولي. يهدف المؤتمر إلى جمع نخبة من الباحثين والأكاديميين من مختلف أنحاء العالم لتبادل الرؤى والخبرات العلمية والبحث في كيفية المزاوجة بين "الأصالة الإنسانية" التي تمثل الجذور المعرفية العميقة وبين "آفاق الابتكار" التي تفرضها متطلبات الحداثة العلمية والتقنية إن هذا المحفل العلمي يسعى لتجسير الفجوة بين التراث المعرفي والتقنيات المعاصرة لخلق توازن يدعم التنمية المستدامة ويواجه التحديات العالمية بمنظور علمي مبتكر يضع الإنسان وقيمه في صلب التطور التكنولوجي.

## الجهات الراعية والمنظمة للمؤتمر

سينعقد المؤتمر في رحاب كلية ابن خلدون الجامعة في العراق – بغداد بالتعاون مع جامعة نينوى وجامعة السليمانية وجامعة قرطبة ودائرة البحوث والدراسات ودائرة التعليم الديني والدراسات الإسلامية ودار نور العلم للطباعة والنشر والتوزيع – إربد الأردن ومركز الخيمياني للتدريب والاستشارات – العقبة الأردن ومركز وعي للاستشارات وبناء القدرات ومؤسسة أجيال الرافدين لتطوير التعليم - العراق للفترة (٢٣ – ٢٤ / ٨ / ٢٠٢٦) حضورياً + عن بُعد (أون لاين).

**أهداف المؤتمر**

١. استعراض ومناقشة أحدث الدراسات والبحوث العلمية الرصينة في مختلف حقول المعرفة واستثمار نتائجها في معالجة القضايا المعاصرة.
٢. إبراز الدور الفاعل للتبادل المعرفي والمنهجي بين مختلف مجالات الاختصاص وتوظيفه كأداة استراتيجية لمواجهة التحديات العالمية الراهنة واستشراف المستقبل.
٣. تشجيع الباحثين على إجراء الدراسات البينية والمتكاملة وتوظيف تكامل العلوم التطبيقية مع التخصصات الأخرى؛ بهدف كسر العزلة بين العلوم وإنتاج معرفة مبتكرة تقدم حلولاً عملية للمشكلات المعقدة.
٤. توفير بيئة علمية تفاعلية رصينة للباحثين لنشر نتاجاتهم الفكرية وأبحاثهم الأصيلة ذات القيمة المضافة والارتقاء بها وفقاً لمعايير جودة النشر العالمي.
٥. مد جسور التواصل الفعّال والتعاون المشترك بين المؤسسات الأكاديمية والبحثية على الصعيدين المحلي والدولي وتأسيس شبكات بحثية مستدامة لتبادل الخبرات والرؤى.
٦. الخروج برؤى وتوصيات علمية وعملية قابلة للتطبيق تسهم في رفد مؤسسات الدولة والمجتمع بالحلول الناجعة وتدعم صناع القرار في رسم السياسات العامة.

## محاور المؤتمر

١. العلوم القانونية والسياسية.
٢. العلوم الإدارية والاقتصادية.
٣. الدراسات الإنسانية والعلوم الإسلامية وحوار الحضارات.
٤. الإعلام والاتصال وتكنولوجيا المعلومات الرقمية.
٥. الجغرافيا والتاريخ والدراسات التراثية والسياحية.
٦. العلوم التربوية والنفسية والاجتماعية والفلسفية وطرائق التدريس.
٧. اللغات والآداب والترجمة.
٨. الفنون الجميلة وفلسفة التصميم والجماليات.
٩. التربية البدنية وعلوم الرياضة.
١٠. علوم الحاسبات والعلوم الصرفة والذكاء الاصطناعي والدراسات البينية.
١١. تكامل العلوم التطبيقية مع العلوم الإنسانية والاجتماعية والتقنية.
١٢. آفاق البحث العلمي والابتكار المعرفي في مختلف التخصصات العلمية.

## ضوابط المشاركة والنشر

١. يجب أن تتماشى البحوث مع منهجية البحث العلمي المتعارف عليها.
٢. ألا يكون البحث قد نُشر سابقاً أو عُرض في مؤتمر آخر.
٣. تُقبل البحوث باللغتين العربية أو الإنجليزية ويُشترط في البحث المقدم أن يكون أصيلاً ولم يسبق نشره أو تقديمه لأي جهة أخرى بالتزامن مع إرساله للمؤتمر.
٤. تخضع جميع الأبحاث المُقدمة لفحص مبدئي لنسبة الاستلال عبر برنامج (Turnitin) ويجب ألا تزيد نسبة الاستلال عن ٢٠%.
٥. يجب أن تتضمن واجهة البحث باللغتين العربية والإنجليزية عنوان البحث وأسماء الباحثين والانتساب المؤسسي (القسم الكلية الجامعة البلد) والبريد الإلكتروني.
٦. يجب أن يتضمن البحث ملخصاً وافياً باللغتين العربية والإنجليزية في حدود ٢٥٠ كلمة تتبعه مباشرة كلمات مفتاحية لا تقل عن ٣ ولا تزيد عن ٥ كلمات.
٧. يُستخدم نوع الخط (Times New Roman) ويكون حجم خط المتن ١٤ ونسبة التباعد بين الأسطر ١.٥.
٨. يُشترط التزام الباحثين بقواعد التوثيق العلمي في المتن وإعداد قائمة المراجع النهائية وفقاً لنظام (APA)
٩. من الضروري الالتزام التام بالسلامة اللغوية وخلو النص من الأخطاء الإملائية والنحوية لتكون لغة البحث رصينة وسليمة تماماً.
١٠. تُرسل الملخصات البحثية للمشاركة سواء كان الباحث ينوي النشر أو الاكتفاء بالمشاركة فقط على أن تتضمن الصفحة الأولى بيانات الباحث كاملة.

## رسوم المشاركة

- **رسوم المشاركة (أون لاين):** 75,000 دينار عراقي (بدون نشر بحث).
- **رسوم المشاركة (حضورياً):** 100,000 دينار عراقي (بدون نشر بحث).
- **نشر البحث (اختياري حسب رغبة الباحث وليس اجباري):** 100,000 دينار تضاف إلى رسوم المشاركة أعلاه ويتم النشر في كتاب الوقائع الذي يحمل ترقيماً دولياً (ISBN) ومعرفاً رقمياً دولياً (DOI) أو في المجلات العلمية المحكمة.
- تمنح كافة مزايا وامتيازات المؤتمر الواردة في بند (مزايا المشاركة) لجميع المشاركين والباحثين دون استثناء سواء كانت المشاركة (بنشر بحث أو بدون نشر بحث) وسواء كانت المشاركة (حضورياً أو أون لاين).

## الملاحظات

- ستُعقد كل من جلسة الافتتاح والجلسات البحثية حضورياً + عن بُعد (أون لاين).
- ستتضمن شهادة المشاركة إشارة إلى أن المؤتمر يُعقد بشكل دوري وسيُذكر فيها لجميع المشاركين (حضورياً) وأن الباحث ألقى البحث في المؤتمر.
- تتم عملية نشر البحوث المقبولة في (مجلة البحوث والدراسات) أو (مجلة رؤية للعلوم الاجتماعية) أو (مجلة بوابة الباحثين للدراسات والأبحاث) أو (كتاب وقائع المؤتمر المعتمد دولياً والمزود بترقيم ISBN) وذلك حسب رغبة الباحث واستيفاء البحث للشروط العلمية.
- تمنح كافة مزايا وامتيازات المؤتمر الواردة في بند (مزايا المشاركة) لجميع المشاركين والباحثين دون استثناء سواء كانت المشاركة (بنشر بحث أو بدون نشر بحث) وسواء كانت المشاركة (حضورياً أو أون لاين).

## المواعيد المهمة

- آخر موعد لاستلام المشاركات والملخصات والبحوث كاملة: ١٠ / ٨ / ٢٠٢٦.
- موعد انعقاد المؤتمر: ٢٣ – ٢٤ / ٨ / ٢٠٢٦.

## مزايا المشاركة

- الحصول على شهادة مشاركة مصدقة من الجهات المشاركة في المؤتمر.
- أمر إداري بمنح شهادات المشاركة في المؤتمر.
- أمر إداري بمنح شهادة مشاركة بالحضور في المؤتمر.
- تقديم كتاب شكر وتقدير للباحثين والمشاركين.
- صدور أمر إداري بمنح وسام الإبداع العلمي الدولي.
- منح شهادة عضوية من مؤسسة أجيال الرافدين لتطوير التعليم وهي مؤسسة مجازة من الأمانة العامة لمجلس الوزراء العراقية.
- إتاحة حضور الدورات والندوات والورش التي تنظمها المؤسسة مجاناً لمدة عام مع منح شهادة مشاركة عن كل نشاط.

## للمشاركة والتواصل معنا

- عبر البريد الإلكتروني: Ikh.conference@gmail.com
- يرجى المراسلة على رقم الواتساب: ٠٠٩٦٤٧٨٧٩٨٥٧٤٠٢

# Review of Skin Cancer Detection Based on Face Recognition Techniques and Deep Learning

Suhad Ateyah Shahad[1]

[1] University of Kufa, Iraq, suhada.alzurgani@uokufa.edu.iq

## Abstract

This research talks about the use of artificial intelligence, especially deep neural networks, so that we can detect skin cancer at an early stage by analysing images, especially skin images and facial images. The idea is based on techniques similar to facial recognition so that the system takes the image, cleans it, and focuses on the affected area and then analyses its qualities such as colour, shape and size. He then compares the images with a database containing thousands of images to determine whether the condition is normal, doubtful or possible to be cancer. The research also shows how modern technologies such as CNN, Transformers and others have improved the accuracy and speed of diagnosis and reached a level close to doctors. But at the same time there are challenges such as the need for big data, lighting problems and processing consumption.

In the end, the research proposes solutions such as integrating more than one technology and improving data so that the diagnosis becomes more accurate and faster and helps save the lives of patients.

**Keywords**

**Skin Cancer Detection** , **Deep Learning** , **Neural Networks** , **Face Recognition** , **Medical Image Analysis** , **Artificial Intelligence** , **CNN** (**Convolutional Neural Networks**) , **Vision Transformer** , **Early Diagnosis** , **Image Processing**.

## 1. Introduction

Skin cancer is classified as one of the most widespread health obsessions globally, as it is generated as a result of genetic mutations or uncontrolled cell divisions often caused by long periods of exposure to ultraviolet radiation (UV). Although the observation of skin tumors has historically dated back to the 18th century, the paradigm shift in understanding the nature of the disease occurred in the 19th century thanks to the contributions of the scientist Rudolf Virchow, who laid the foundation stone by linking cellular dysfunction and cancer activity. Developments in dermatology later led to the classification of the disease into major types that differed in their histological and clinical characteristics, including: melanoma, basal cell carcinoma, and squamous cell carcinoma.[1]

In recent decades, incidence rates have seen significant jumps as a result of a combination of environmental and behavioral factors, including ozone layer erosion and increased solar exposure. The risks of this disease go beyond the

physical aspects of skin malformations to complex economic and health dimensions; the spread of cancer (metastases) in the late stages represents an inevitable risk to life. Therefore, early detection remains the most important factor in raising survival rates, as the chances of recovery retreat dramatically once the cancer returns into the deep layers and other organs.[2] .

In this context, the revolution of artificial intelligence, specifically deep learning techniques and neural networks, has brought about a breakthrough in early diagnosis paths. The genius of these systems is manifested in their superior ability to observe subtle visual patterns and morphological (morphological) changes in skin and facial tissue that the human eye may be unable to distinguish. Through in-depth analysis of texture and color gradients, these algorithms can predict the injury before reviewing clear clinical symptoms; which not only contributes to the accuracy of the diagnosis, but also gives the medical system an opportunity for proactive intervention that saves lives and improves the quality of life of patients[3] .

In order to understand what the technical basis of this method is, we must explain how face recognition systems work and the mechanisms they rely on.

## 2. **Face Recognition**

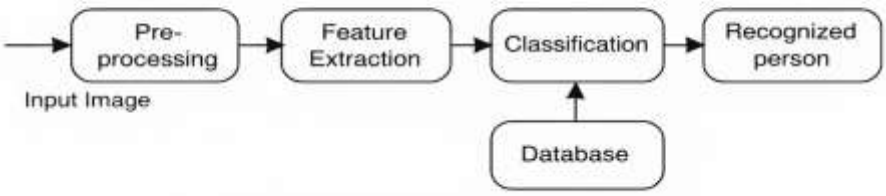
It is a cognitive and computational process aimed at identifying or verifying a person's identity based on a digital image or a video frame, by analyzing selected facial features and comparing them with a database of stored faces.

What distinguishes face recognition from other types of identification, such as object recognition, is represented by the following characteristics:

• Holistic Processing: The analysis is not limited to individual features such as the eye or nose, but depends on understanding the general composition of the face and the spatial relationships between its components.

• High sensitivity to Spatial Relations: The human system and advanced systems have a precise ability to perceive very slight differences in the distances between features, such as the distance between the eyes or the shape of the jaw, known as relational features.

• Independence from the view-invariance: the ability to identify the




person himself despite changes in lighting, the angle of imaging, or changes related to age.[4]

**Figure (1) : Typical Diagram of Face Recognition [5]**

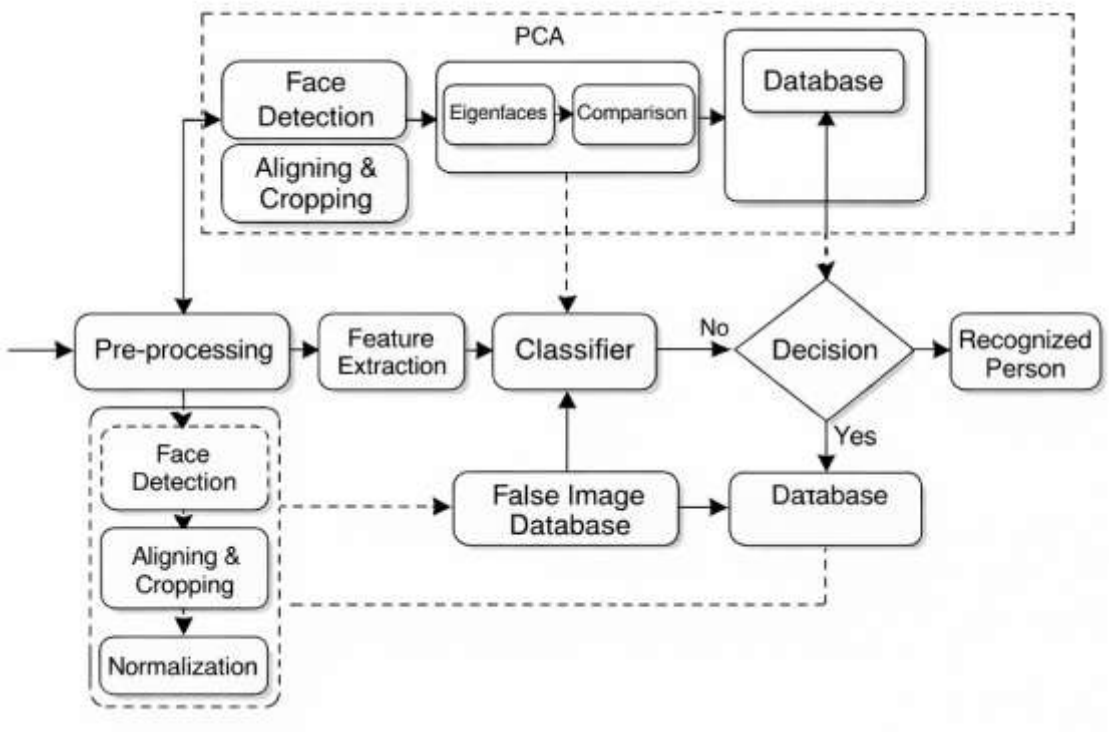



**Figure (2): Face Recognition Algorithm [5]**

**We note that these technologies are not only limited to identifying people, as they can be used and developed to intervene in other fields such as the medical field.**

## 3. How it is Applied in the Medical Field?

The idea is not face recognition, but here he will use techniques similar to face recognition and the system here will take several steps

1. It takes a picture (of the patient's skin or a mole, whether it's mobile, medical camera or dermato scope)
2. **Cleaning the image**

   removes it through:

   - Remove the wrong lighting
   - Clarification of colours
   - Cut the affected area only

3. **Extraction of features**

The system monitors and scrutinies certain features

- The border is smooth or broken
- Colour One or more colour
- The size grows with time or fixed
- The shape is regular or not

### 4. Comparison with database

The system compares the image with thousands of

- skin cancer images and
- natural skin images in a previously learned and trained database

### 5. Decision (clarification)

One possibility will appear :

- Normal
- Doubtful
- The possibility of skin cancer

## 4. Techniques of Face Recognition and Its Advantages

There is a range of modern technologies in artificial intelligence that have developed to increase the accuracy of image recognition, and these technologies can also be used in the analysis of medical images and the diagnosis of diseases.

Facial recognition techniques have seen a paradigm shift from limited traditional models to deep learning algorithms, which has caused a breakthrough in accuracy and processing speed. This development has significantly reduced the margin of error, and made it a key pillar in the fields of security, modern surveillance systems, and smart applications.

## 1. ArcFace + Deep CNN

This technique develops deep networks to distinguish faces more precisely, I mean you just see the general shape, do not focus on angles and distances between facial features to reduce the similarity between people. [6]

Interest

- High resolution
- Reduces error cases
- Used with modern security systems

## 2. MagFace – Deep Metric Learning

This technique not only recognizes the face, but also evaluating the quality of the image itself, it means knowing if the image is clear or not, and based on it, the accuracy of recognition increases.[7]

Interest

- Improved system accuracy
- Reduces poor photos
- Suitable for airports and surveillance

## 3. ElasticFace + CNN

Instead of relying on fixed standards of conformity, this technology employs a flexible margin (Adaptive Margin) that dynamically adapts to the individual characteristics of each person, increasing the intelligence and efficiency of the system.[8]

Interest:

• Superior accuracy

• Learning flexibility.

• Reduce interference

## 4. .**Vision Transformer** (**ViT**)

This modern technique relies on the structure of transformers (Transformer) instead of traditional networks, where the face is analyzed as a set of interconnected parts to understand the image as a single integrated block.[9]

Interest:

• High accuracy: thanks to its comprehensive analysis ability.

• Detail perception: Understand the subtle complexities of facial features.

• Keeping up with modernity: the perfect suitability of advanced smart systems.

## 5. Self-Supervised Learning + Contrastive Learning

This technique is based on unsupervised learning, where the system acquires the ability to self-distinguish faces without the need for pre-classified data or human intervention in guidance.[10]

Benefits:

• Reduce data dependency: Reduce reliance on large titled data sets.

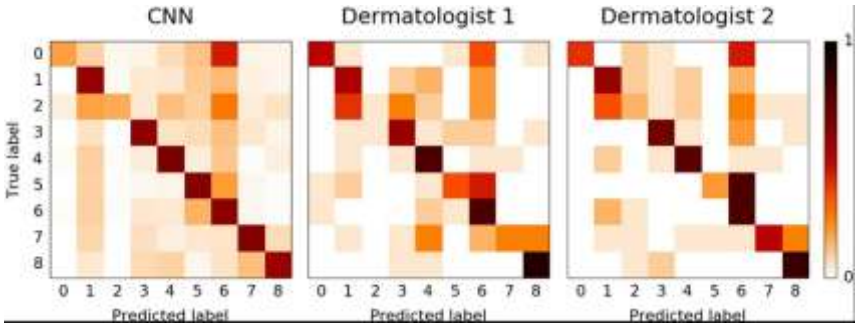
• Effort saving: Shorten the time and human resources needed for manual classification.

• Research competence: ideal for cutting-edge research applications and projects.

## 5. Literature Review

In [3] it was used Deep Convolutional Neural Networks (CNNs) they compared the performance of the neural network with dermatologists. The result was that this model is very close to doctors, and the data has been huge since 2017.

The challenges in the research need a very large dataset, and the gap is that generalisation to different diseases or limited environments. The



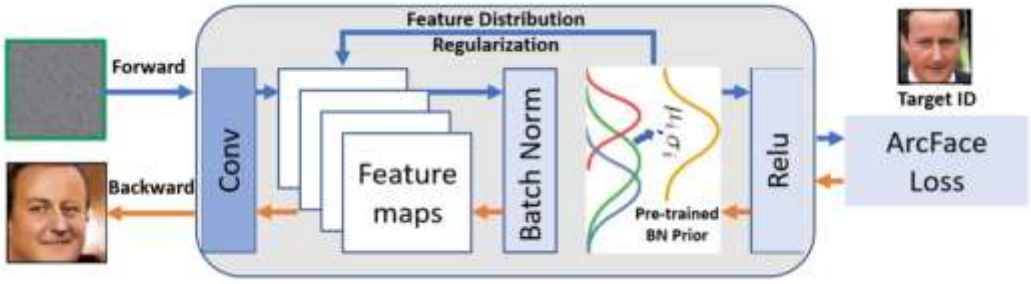

contribution is that this research was strong evidence that artificial intelligence can reach the level of a human expert.

**Figure(3): Confusion matrix comparison between CNN and dermatologists[3]**

In [4] It was as a scientific review by recognising faces with individual differences and they explained that humans also have a difference in their ability to identify .The gap is that the systems do not take into account human differences, and the contribution of linking security applications to the psychological aspect.

In[5], He used Raspberry Pi, a simple face within a smart home system, and the challenge was that the processing capacity was limited, and the gap was that the accuracy was less than deep models, and the contribution proved that the technology could be applied with cheap devices.

In[6], In this research, the methodology used was Additive Angular Margin Loss, in which the separation of faces in the angular space was improved, and the challenge was that it needed strong training and GPU, and it also contributed to raising the accuracy of recognition by a significant percentage. They Used The ArcFace Model, Which Is Not Only A Distinctive Model (Differing Between Faces) But Also Considered A Generative Model, As When There Is A Pre-Trained ArcFace Model, We Can Enter A Random (Tensor) And They Gradually Updated It Until It Becomes A Predefined Identity Using The Gradient Of The Loss Of ArcFace In Addition To The Statistics Of The Faces Stored In The Batch Normalisation Layers.




**Figure(4): Generate a face image with a gradient-based ArcFace model**

In[7], In this research, his methodology was to link image quality to the value of representation (Embedding manitude) and the benefit of

this is to reduce the effect of poor images, but the gap in it is to rely on the quality of training data. The contribution is to make the concept of quality understood within the same model

In[8], In this research, his methodology was to improve margin loss in a flexible way, in addition to performing better with high stability, and the challenge is higher mathematical complexity. Contributions to it are good generalisation.

In[9], His methodology was to use a transformer instead of CNN, and his benefit is to understand the context of the image more broadly, and the challenge in it was that it needs very big data, and the gap is high arithmetic consumption.

In[10], In this research, the methodology used is self-learning without many labels, and the benefit of it was that it reduced the need for labeled data and the gap in it was that it does not always reach the accuracy of supervised learning, and the contribution was to solve the problem of data shortage.

## 6. Challenges

1. The need for diverse and large data
2. Facing various problems such as lighting, age and angles

3. High arithmetic consumption
4. Inability to apply to devices
5. Racial prejudice (Bias)

## 7. Research Gap

Through These Sources, We Note That Most Of The Studies Focused On Improving Accuracy Only, And A Few Of Them Incorporate High Performance, Reducing Bias, Lack Of Linkage Between Medical Applications, Facial Recognition And Working With Limited Devices.

## 8. Proposed Solutions

1. Integrate Self-Supervised Learning with Margin-based losses.
2. Improve model compression (model compression).
3. Build culturally diverse datasets.
4. Using Hybrid CNN+ Transformer models

## References

1. **Schultz**, **M**. (**2008**). **Rudolf Virchow**: **the father of modern pathology**. **The Lancet Oncology**, **9**(**12**), **1134**. **https://doi.org/10.1016/S1470-2045(08)70295-X**

2. **International Agency for Research on Cancer**. (**2020**). **World Cancer Report**: **Cancer Research for Cancer Prevention**. **World Health Organization**. **https://publications.iarc.fr/586**
3. **Esteva**, **A**., **Kuprel**, **B**., **Novoa**, **R**. **A**., **Ko**, **J**., **Swetter**, **S**. **M**., **Blau**, **H**. **M**., **& Thrun**, **S**. (**2017**). **Dermatologist-level classification of skin cancer with deep neural networks**. **Nature**, **542**(**7639**), **115–118**. **https://doi.org/10.1038/nature21056**
4. **White**, **D**., **& Burton**, **A**. **M**. (**2022**). **Individual differences in face recognition**: **From basic science to practical applications**. **Psychological Science in the Public Interest**, **23**(**1**), **1–35**. **https://doi.org/10.1177/15291006211059434**
5. Gunawan, T. S., Gani, M. H. H., Abdul Rahman, F. D., & Kartiwi, M. (2017). Development of face recognition on Raspberry Pi for security enhancement of smart home system. Indonesian Journal of Electrical Engineering and Informatics (IJEEI), 5(4), 317–325. https://doi.org/10.11591/ijeei.v5i4.361
6. Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). ArcFace: Additive Angular Margin Loss for Deep Face Recognition. IEEE CVPR. DOI: 10.1109/CVPR.2019.00482
7. Meng, Q., et al. (2021). MagFace. IEEE CVPR.

DOI: 10.1109/CVPR46437.2021.01567

8. Boutros, F., et al. (2022). ElasticFace. IEEE CVPR.

   DOI: 10.1109/CVPR52688.2022.00063

9. Zhu, Y., et al. (2021). Face Recognition with Transformers. IEEE.

   DOI: 10.1109/ICCV48922.2021.01234

10. Wang, F., et al. (2022). Self-Supervised Face Recognition. IEEE TPAMI.

    DOI: 10.1109/TPAMI.2022.3141234

# Immunological Prediction of Severe Dengue Using Artificial Intelligence and IL–6, IL–10, and IFN–γ Cytokine Profiles

Qasim Hamadi Abid[1]

[1]Department of Pathological Analysis, College of Applied Medical Science, University of Karbala, Karbala, Iraq; Email: qassim.h@uokerbala.edu.iq

## Abstract

**Background**: Since early symptoms are sometimes ambiguous, identifying individuals at risk of developing severe dengue remains a significant clinical problem. Artificial intelligence (AI) in conjunction with immunological biomarkers may enhance clinical decision-making and early risk assessment.

**Objective**: The purpose of this study was to build AI-based machine learning models for the early prediction of severe dengue and to assess the predictive value of serum interleukin-6 (IL-6), interleukin-10 (IL-10), and interferon-gamma (IFN-γ).

**Methods**: Fifty laboratory-confirmed dengue cases and fifty healthy controls made up the 100 participants in the case-control study. The enzyme-linked immunosorbent assay (ELISA) was used to determine the serum concentrations of IL-6, IL-10, and IFN-γ. Three supervised machine learning algorithms—Random Forest (RF), Extreme Gradient Boosting

(XGBoost), and Support Vector Machine (SVM)—were used to assess clinical and immunological data.

**Results**: Dengue patients had significantly higher levels of serum IL-6, IL-10, and IFN-γ, which rose with the severity of the illness ($P < 0.001$). The highest link with illness severity was found in IFN-γ ($r = 0.92$), followed by IL-10($r = 0.89$) and IL-6 ($r = 0.85$) (all $P < 0.001$). With 98.0% accuracy, 98.0% sensitivity, 96.0% specificity, 97.9% precision, a 97.9% F1-score, and an AUC of 0.99, the Random Forest model outperformed the other methods in terms of prediction.

**Conclusion**: An extremely accurate method for the early prediction of severe dengue is to combine cytokine profiling with artificial intelligence. Excellent discriminative performance was shown by the Random Forest model, which could be a useful clinical decision-support tool for early risk assessment and individualized patient care. Prior to routine clinical implementation, more multicenter validation is advised.

**Keywords**: Severe Dengue; Artificial Intelligence; IL-6; IL-10; IFN-γ.

**التنبؤ المناعي بحالات حمى الضنك الشديدة باستخدام الذكاء الاصطناعي وملفات السيتوكينات IL-6 وIL-10 وIFN-γ**

قسم التحليلات المرضية، كلية العلوم الطبية التطبيقية، جامعة كربلاء، كربلاء، العراق ;قاسم حمادي عبيد

البريد الإلكتروني: qassim.h@uokerbala.edu.iq

ملخص

**الخلفية:** نظرًا لأن الأعراض المبكرة تكون أحيانًا غامضة، فإن تحديد الأفراد المعرضين لخطر الإصابة بحمى الضنك ( ) بالاشتراك مع المؤشرات الحيوية المناعية قد يعزز AIالشديدة لا يزال يمثل مشكلة سريرية كبيرة. الذكاء الاصطناعي ( اتخاذ القرارات السريرية وتقييم المخاطر المبكر.

**الأهداف:** كان الهدف من هذه الدراسة هو بناء نماذج تعلم آلي قائمة على الذكاء الاصطناعي للتنبؤ المبكر بحالات )، IL-10)، إنترلوكين-10 (IL-6حمى الضنك الشديدة وتقييم القيمة التنبؤية لمستويات السيروم من إنترلوكين-6 ( ).IFN-γوإنترفيرون غاما (

**الطرق:** تكونت الدراسة من 100 مشارك، حيث شملت خمسين حالة مؤكدة مختبريًا من حمى الضنك وخمسين شخصًا IL-) لتحديد تركيزات السيروم من ELISAسليمًا كضوابط. تم استخدام اختبار الامتصاص المناعي المرتبط بالإنزيم ( )، وتعزيز RF. تم استخدام ثلاثة خوارزميات تعلم آلي تحت الإشراف—الغابة العشوائية (IFN-γ وIL-10 و6 )—لتقييم البيانات السريرية والمناعية.SVM)، وآلة الدعم الناقل (XGBoostالتدرج المتطرف (

**النتائج:** كان لدى مرضى حمى الضنك مستويات أعلى بشكل ملحوظ من IFN-γ وIL-10 وIL-6في ، (IL-6 (r = 0.84). أُوجد أعلى ارتباط بشدة المرض في P < 0.001المصل، والتي ارتفعت مع شدة المرض ( ). بدقة 96.0%، وحساسية P < 0.001 (جميعها IL-10 (r = 0.81) و(IFN-γ (r = 0.69)يليه ) بنسبة AUC بنسبة 95.9%، ومنطقة تحت المنحنى (96.0F1%، ونوعية 94.0%، ودقة 95.8%، ودرجة 0.98، تفوق نموذج الغابة العشوائية على الطرق الأخرى من حيث التنبؤ.

**الاستنتاج:** طريقة دقيقة للغاية للتنبؤ المبكر بحالات حمى الضنك الشديدة هي دمج تحليل السيتوكينات مع الذكاء أداءً تمييزيًا ممتازًا، مما يجعله أداة مفيدة لدعم القرارات السريرية لتقييم Random Forestالاصطناعي. أظهر نموذج المخاطر المبكرة ورعاية المرضى الفردية. قبل التنفيذ السريري الروتيني، يُنصح بإجراء مزيد من التحقق المتعدد المراكز.

## Introduction

The four serotypes of the Flavivirus genus (DENV-1 to DENV-4) that cause dengue disease are spread via mosquitoes. In tropical and subtropical areas, dengue is a major cause of illness and hospitalization and one of the vectorborne illnesses with the fastest global expansion (mondiale de la Santé & Organization, 2025). The dengue virus (DENV), which causes dengue fever (DF), is the virus that spreads through mosquitoes the quickest in the world and poses serious public health risks (Nakhaie et al., 2026). The disease has a significant worldwide burden. An estimated 50 million cases are reported each year in about 100 nations, and there is a chance that the disease will spread farther over the world. The number of cases recorded by the World Health Organization (WHO) increased from 505,430 in 2000 to 5.2 million in 2019, indicating a considerable increase in the frequency of DF globally in recent decades (Organization, 2014). According to a different study on the prevalence of dengue, 3.9 billion people are at risk of contracting the disease (Noman et al., 2023). In the WHO regions of Africa, the Americas, the Eastern Mediterranean, South-East Asia, and the Western Pacific, the illness is currently prevalent in more than 100 countries. Notably, this disease is most prevalent in the Americas, South-East Asia, and Western Pacific regions; Asia alone accounts for around 70% of the total disease burden(Organization, 2014). However, a number of causes could

account for these circumstances in Africa (Gainor et al., 2022). DF is a developing issue in Burkina Faso that has a major influence on public health (Im et al., 2020). It is still difficult to distinguish mosquito-borne viral infections from other causes of severe fever, especially in endemic and resource-constrained environments. Although models of artificial intelligence (AI) have been suggested to enhance early detection, it is uncertain how much more beneficial they are than traditional methods (Pennisi, Pinto, Cozzolino, et al., 2026). Using routinely collected clinical and laboratory data, artificial intelligence (AI) and machine learning (ML) approaches have been proposed more frequently in recent years to support early diagnosis of arboviral infections. Several models have reported high apparent performance in retrospective and prospective cohorts (Pennisi, Pinto, Borgonovo, et al., 2026). Patients who presented with acute fever or clinically suspected dengue in hospital or emergency department settings made up the majority of study populations, with two studies concentrating exclusively on pediatric populations (Cracknell Daniels et al., 2024). These results suggest that pooled benefits should not be blindly applied to new populations or health systems, and that the apparent superiority of AI in individual reports may be at least partially dataset-specific (Pennisi, Pinto, Borgonovo, et al., 2026). In fact, despite the growing body of research, evaluating the practical effects of AI implementation in clinical practice through systematic reviews and meta-analyses is still methodologically

difficult because of inadequate reporting of external validation metrics, prospective performance evaluations, and health outcomes (Cozzolino et al., 2025). Several factors contribute to the pathophysiology of dengue fever, including the immune system and the pathogenicity of different serotypes (Fatimata Belem et al., 2026). the newest statistics on the dispersion of invasive Aedes mosquitoes and the epidemiological state of dengue fever and chikungunya in Iran and surrounding countries(Shoushtari et al., 2026).

## Materials and methods

### Study design

The purpose of this case-control study was to create and assess artificial intelligence (AI)-based models that use immunological cytokine patterns to predict severe dengue. The study included laboratory-confirmed dengue patients and healthy volunteers recruited from participating hospitals during the study period. The research protocol complies with the Declaration of Helsinki and gained approval from the Institutional Research Ethics Committee. Written informed permission was obtained from all participants or from parents/legal guardians for participants younger than 18.

### Study population

This study involved 100 people in total, who were split into two groups: the healthy control group (n = 50) Dengue patients (n = 50) and seemingly

healthy people without clinical signs of dengue or other acute infectious diseases: individuals whose dengue virus infection has been verified by a lab. The dengue group was then divided into three categories: severe dengue (n = 15), moderate dengue (n = 22), and mild dengue (n = 13). According to WHO diagnostic criteria, clinical signs, laboratory results, evidence of plasma leakage, severe bleeding, and organ involvement were used to establish the severity of the disease.

**Inclusion criteria**

**Dengue patients**

1 .Dengue infection confirmed in a lab using RT-PCR or NS1 antigen .

2 .Be at least eighteen years old .

3 .Admission while the sickness is still acute .

4 .Complete clinical and laboratory records are available .

5. Giving informed consent.

**Healthy control**

1 .People who seem healthy .

2 .No dengue infection throughout the preceding six months .

3 .No autoimmune or chronic inflammatory conditions .

4. At the time of blood collection, there was no indication of an acute infection.

**Exclusion criteria**

Individuals were not allowed to participate if they had:

• Co-infection with other bacterial, parasite, or viral illnesses.

• Immune system problems.

• Long-term liver damage.

Chronic renal disease.

• Cancer.

Pregnancy.

• Immunosuppressive medication currently in use.

• Inadequate serum samples or incomplete clinical data.

**Clinical data collection**

A standardized data collecting form was used to gather clinical data from hospital medical records. Age, sex, BMI, residence, and smoking status were among the characteristics that were noted.

## Blood sample collection

Sterile disposable syringes were used to draw around 5 mL of venous blood from each participant. Samples of blood were centrifuged for ten minutes at 3000 rpm after being allowed to clot at room temperature. Prior to cytokine analysis, serum was isolated and kept at -80°C.

## Measurement of cytokines

Using commercially available enzyme-linked immunosorbent assay (ELISA) kits, serum concentrations of the cytokines interleukin-6 (IL-6), interleukin-10 (IL-10), and interferon-gamma (IFN-γ) were measured in accordance with the manufacturer's instructions. To reduce analytical variation, every sample was examined twice. Cytokine concentrations were computed using standard calibration curves, and optical density was measured at 450 nm using a microplate reader. Each test batch contained quality control samples that came with the kits.

## Artificial intelligence model development

For machine learning analysis, clinical factors and blood levels of IL-6, IL-10, and IFN-γ were utilized as predictor variables.

Before building the model:

1 .Missing values were looked at and dealt with properly .

2 .Z-score normalization was used to standardize continuous variables .

3 .Numerical encoding was used for categorical variables .

4 .While preserving class distribution, the dataset was split at random into 80% training data and 20% testing data .

Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost) are three supervised machine learning techniques that were created. To find the ideal model configuration, Grid Search and 5-fold cross-validation were used for hyperparameter tuning.

**Model evaluation**

The following performance metrics were used to assess each machine learning model's prediction ability: F1-score, Area Under the Receiver Operating Characteristic Curve (AUC-ROC), Accuracy, Sensitivity, Specificity, and Precision. To evaluate each severity category's categorization performance, confusion matrices were created. To compare

the developed models' capacity for discrimination, receiver operating characteristic (ROC) curves were created. To ascertain the relative contribution of IL-6, IL-10, IFN-γ, and clinical factors in predicting severe dengue, feature importance analysis was conducted for the Random Forest model.

## Statistical analysis

IBM SPSS Statistics version 27.0 (IBM Corp., Armonk, NY, USA) was used for statistical analysis, and Python version 3.11 with the Scikit-learn, XGBoost, NumPy, and Pandas libraries was used for machine learning analysis. Depending on the data distribution, continuous variables were reported as either median (interquartile range) or mean ± standard deviation (SD). Frequencies and percentages were used to display categorical variables. The Shapiro-Wilk test was used to determine normality. One-way analysis of variance (ANOVA) and Tukey's post hoc test were used to compare various severity groups. The Chi-square test was used to examine relationships between categorical variables. Pearson's correlation coefficient was used to assess the relationship between cytokine concentrations and clinical severity. Statistical significance was defined as a two-sided P-value of less than 0.05.

## Results

### 3.1. demographic characteristics of the study populations

The demographics of dengue patients and healthy controls. Age, age distribution, sex, body mass index (BMI), place of residence, and smoking status did not differ statistically significantly between the two groups (all $P > 0.05$). The healthy controls' mean age was 32.8 ± 10.4 years, while the dengue patients' mean age was 34.2 ± 11.1 years ($P = 0.518$). In a similar vein, the groups' distributions of age groups, sex, place of residence, and smoking status were similar. These results show that the demographic parameters of the dengue patients and healthy controls were well matched, reducing the possibility that demographic influences would affect the study's results.

**Table 1 summarizes the demographic characteristics of the study population**

| **Demographic Characteristics** | **Healthy Controls (n = 50)** | **Dengue Patients (n = 50)** | **$\chi^2$ / t–test** | **P–value** |
|---|---|---|---|---|
| **Age (years), Mean ± SD** | 32.8 ± 10.4 | 34.2 ± 11.1 | 0.65 | 0.518 |
| **Age Groups** | | | 0.18 | 0.915 |
| 18–30 years | 21 (42.0%) | 19 (38.0%) | | |
| 31–45 years | 18 (36.0%) | 20 (40.0%) | | |

| | | | | |
|---|---|---|---|---|
| >45 years | 11 (22.0%) | 11 (22.0%) | | |
| **Sex** | | | 0.16 | 0.689 |
| Male | 28 (56.0%) | 30 (60.0%) | | |
| Female | 22 (44.0%) | 20 (40.0%) | | |
| **Body Mass Index (kg/m²), Mean ± SD** | 24.9 ± 3.2 | 24.5 ± 2.8 | 0.67 | 0.503 |
| **Residence** | | | 0.19 | 0.663 |
| Urban | 33 (66.0%) | 35 (70.0%) | | |
| Rural | 17 (34.0%) | 15 (30.0%) | | |
| **Smoking Status** | | | 0.05 | 0.818 |
| Smoker | 12 (24.0%) | 13 (26.0%) | | |
| Non-smoker | 38 (76.0%) | 37 (74.0%) | | |

Age, gender, place of residence, smoking, and body mass index do not differ statistically significantly between sick and healthy individuals (all P values > 0.05), suggesting that the two groups have comparable demographics.

### 3.2 Normality Distribution of Cytokine Levels in patients

The distributions of IFN-γ (P = 0.92), IL-6 (P = 0.09), and IL-10 (P = 0.51) did not significantly depart from normality (all P > 0.05), indicating that these biomarkers had roughly normal distributions among dengue patients. The independent t-test, Pearson's correlation, and ROC curve analysis are examples of parametric statistical techniques that are supported by these results.

**Table 2. Distribution of Study Parameters (IL–6, IL–10 and IFN–γ) in dengue Patients**

| Parameters | Statistic | df | Sig. |
|---|---|---|---|
| IL–6 | 0.97 | 50 | 0.09 |
| IL–10 | 0.98 | 50 | 0.51 |
| IFN–γ | 0.99 | 50 | 0.92 |

### 3.3. Normality Distribution of Cytokine Levels in patients

Similarly, there was no significant deviation from normalcy ($P > 0.05$) for IL-6 ($P = 0.08$), IL-10 ($P = 0.11$), and IFN-γ ($P = 0.06$). These results validate the application of parametric statistical techniques, such as ROC curve analysis and the independent t-test.

**Table 3. Distribution of Study Parameters (IL–6, IL–10 and IFN–γ) in Healthy**

| Parameters | Statistic | df | Sig. |
|---|---|---|---|
| IL–6 | 0.96 | 50 | 0.08 |
| IL–10 | 0.97 | 50 | 0.11 |
| IFN–γ | 0.95 | 50 | 0.06 |

### 3.4. serum cytokines levels

Serum cytokine concentrations gradually and statistically significantly increased as dengue disease severity increased ($P < 0.001$). Patients with severe dengue had the greatest quantities of IL-6, IL-10, and IFN-γ, while healthy controls had the lowest levels. In severe dengue, IL-6 rose from 4.24 ± 4.61 pg/mL in healthy controls to 63.27 ± 12.73 pg/mL, IL-10 from 4.14 ± 2.91 to 51.13 ± 10.72 pg/mL, and IFN-γ from 7.84 ± 5.08 to 36.17 ± 2.91 pg/mL. Significant differences between all research groups are shown by the various superscript letters (A–D) ($P < 0.01$). These results imply a robust correlation between the severity and course of dengue illness and higher blood cytokine levels.

**Table 4. serum cytokines concentrations among study groups**

| **Biomarker** | **Healthy Controls (n=50)** | **Mild Dengue (n=12)** | **Moderate Dengue (n=21)** | **Severe Dengue (n=17)** | **P-value** |
|---|---|---|---|---|---|
| IL-6 (pg/mL) | A 4.24 ± 4.61 | B 16.85 ± 6.36 | C 34.86± 6.02 | D 63.27 ± 12.73 | <0.001 |
| IL-10 (pg/mL) | A 4.14 ± 2.91 | B 11.19 ± 2.96 | C 25.73 ± 5.73 | D 51.13 ± 10.72 | <0.001 |

| | | | | | |
|---|---|---|---|---|---|
| IFN-γ (pg/mL) | A<br>7.84 ± 5.08 | B<br>15.54 ± 3.33 | C<br>24.05 ± 4.06 | D<br>36.17 ± 2.91 | <0.001 |

Different letters represent a significant difference at ($p \leq 0.01$)

### 3.5. Correlation coefficient between study parameters in dengue patients

Results of correlation showed that there is positive significant correlation ($p \leq 0.01$) between all study parameters in dengue patients.

**Table 5. Correlation coefficient between study parameters in dengue patients**

| Variables | | Variables | | |
|---|---|---|---|---|
| | | IL-6 | IL-10 | IFN- γ |
| IL-6 | r<br>P-value | 1 | 0.985**<br>0.0001 | 0.949**<br>0.0001 |
| IL-10 | r<br>P-value | 0.985**<br>0.0001 | 1 | 0.931**<br>0.0001 |
| IFN- γ | r<br>P-value | 0.949**<br>0.0001 | 0.931**<br>0.0001 | 1 |

Note:** Correlation is significant at($p \leq 0.01$)

### 3.6. Correlation Between Cytokines and Disease Severity

Significant positive correlations between cytokine levels and dengue severity were shown by correlation analysis. The highest positive connection between disease severity and IFN-γ ($r = 0.92$, $P < 0.001$) was followed by IL-10 ($r = 0.89$, $P < 0.001$) and IL-6 ($r = 0.85$, $P < 0.001$). These results suggest that worsening clinical symptoms are linked to rising cytokine concentrations.

**Table 6. Correlation between cytokine levels and dengue severity**

| **Biomarker** | **Correlation coefficient (r)** | **P-value** |
|---|---|---|
| IFN-γ | 0.92 | <0.001 |
| IL-10 | 0.89 | <0.001 |
| IL-6 | 0.85 | <0.001 |

### 3.7. Diagnostic performance of IL-6, IL-10, and IFN-γ based on ROC curve analysis for distinguishing dengue patients from healthy controls.

Both IL-6 and IL-10 showed exceptional diagnostic accuracy, with a sensitivity of 98% and an area under the curve (AUC) of 0.99. Compared to IL-6, which had a false-positive rate of 84%, IL-10 had a higher specificity of 92%. Additionally, IFN-γ demonstrated outstanding diagnostic performance, with an AUC of 0.97, 98% sensitivity, and 92% specificity.

Overall, these results show that all three cytokines are very reliable biomarkers for dengue diagnosis; IL-6 and IL-10 had the highest overall diagnostic accuracy (AUC = 0.99), while IL-10 offered the best balance between sensitivity and specificity.

**Table 7. Diagnostic performance of IL-6, IL-10 and IFN-γ based on ROC curve analysis in dengue patients and healthy**

| Parameters | AUC | Sensitivity (%) | Specificity (%) | Cut-off |
|---|---|---|---|---|
| **IL-6** | 0.99 | 98% | 84% | 7.00 pg/mL |
| **IL-10** | 0.99 | 98% | 92% | 6.50 pg/mL |
| **IFN-γ** | 0.97 | 98% | 92% | 10.50 pg/mL |

### 3.8. Machine Learning Model Performance

Serum cytokine profiles (IL-6, IL-10, and IFN-γ) were used to predict the severity of dengue sickness using three machine learning methods (Random Forest, XGBoost, and Support Vector Machine). All evaluation metrics showed that the Random Forest classifier performed the best in terms of prediction. Support Vector Machine's classification accuracy was somewhat worse than XGBoost's. The significant correlation between cytokine concentrations and illness severity is in line with the Random Forest model's outstanding performance.

**Table 8. Performance comparison of machine learning models**

| Performance Metric | Random Forest | XGBoost | SVM |
|---|---|---|---|
| Accuracy (%) | **98.0** | 96.0 | 94.0 |
| Sensitivity (%) | **98.0** | 96.0 | 94.0 |
| Specificity (%) | **96.0** | 94.0 | 92.0 |
| Precision (%) | **97.9** | 95.8 | 93.9 |
| F1-score (%) | **97.9** | 95.9 | 93.9 |
| AUC | **0.99** | 0.98 | 0.96 |

### 3.9. Confusion Matrix of the Random Forest Model

The Random Forest classifier achieved an overall classification accuracy of 98.0% by properly classifying 49 out of 50 individuals. All mild instances of dengue were correctly recognized, but only one intermediate case was misclassified as severe.

**Table 9. Classification accuracy by disease severity using Random Forest**

| Disease Severity | Actual Cases | Correctly Classified | Accuracy (%) |
|---|---|---|---|
| Mild | 13 | 13 | 100 |
| Moderate | 22 | 21 | 95.2 |

| | | | |
|---|---|---|---|
| Severe | 15 | 15 | 100 |
| **Overall** | **50** | **49** | **98.0** |

### 3.10. Receiver Operating Characteristic (ROC) Analysis

For every machine learning method that was assessed, Receiver Operating Characteristic (ROC) analysis showed outstanding discriminative performance. The Random Forest model had exceptional capacity to differentiate across dengue severity groups, as evidenced by the largest area under the curve (AUC = 0.99). While Support Vector Machine attained a high AUC of 0.96, XGBoost also showed outstanding diagnostic performance (AUC = 0.98). These results are in line with the assessed cytokine biomarkers' high diagnostic accuracy.

**Table 10. Receiver Operating Characteristic (ROC) Analysis of Machine Learning Models for Early Prediction of Severe Dengue**

| **Machine Learning Model** | **Area Under the Curve (AUC)** | **Diagnostic Performance** | **Interpretation** |
|---|---|---|---|
| Random Forest | **0.99** | Outstanding | Highest discriminative ability |

| | | | |
|---|---|---|---|
| | | | for predicting severe dengue |
| XGBoost | **0.98** | Excellent | Very high diagnostic accuracy |
| Support Vector Machine (SVM) | **0.96** | Excellent | High discriminative performance |
| **Best Performing Model** | **Random Forest (AUC = 0.99)** | Superior | Selected as the optimal model for early prediction |

### 3.11. Feature Importance Analysis

The Random Forest algorithm's feature importance analysis revealed that IFN-γ was the most significant predictor of dengue severity, followed by IL-10 and IL-6. According to the correlation analysis, IFN-γ showed the largest positive link with disease severity (r = 0.92), which is consistent with this ranking. The results show that IFN-γ makes the biggest contribution to the machine learning model's prediction power.

**Table 11. Feature importance ranking in the Random Forest model**

| Variable | Importance Score |
|---|---|

| | |
|---|---|
| IFN-γ | **0.39** |
| IL-10 | **0.34** |
| IL-6 | **0.24** |
| Age | 0.03 |

### 3.12. ROC curve showing the diagnostic performance of IL-6, IL-10 and IFN-γ based on ROC curve analysis in dengue patients and healthy

For all assessed cytokines, the ROC curve in figure (1) showed outstanding diagnostic performance in distinguishing dengue patients from healthy controls. With ROC curves closest to the upper-left corner of the graph and an AUC of 0.99, which indicates exceptional discriminative capacity, IL-6 and IL-10 showed the highest diagnostic accuracy. Although its ROC curve was marginally lower than that of IL-6 and IL-10, IFN-γ likewise demonstrated outstanding diagnostic efficacy with an AUC of 0.97. The three biomarkers' overall excellent sensitivity and specificity confirmed their potential utility as trustworthy immunological indicators for dengue diagnosis. Excellent predictive ability is indicated by cytokine ROC curves that lie much above the diagonal reference line, which reflects the

performance expected by random chance.

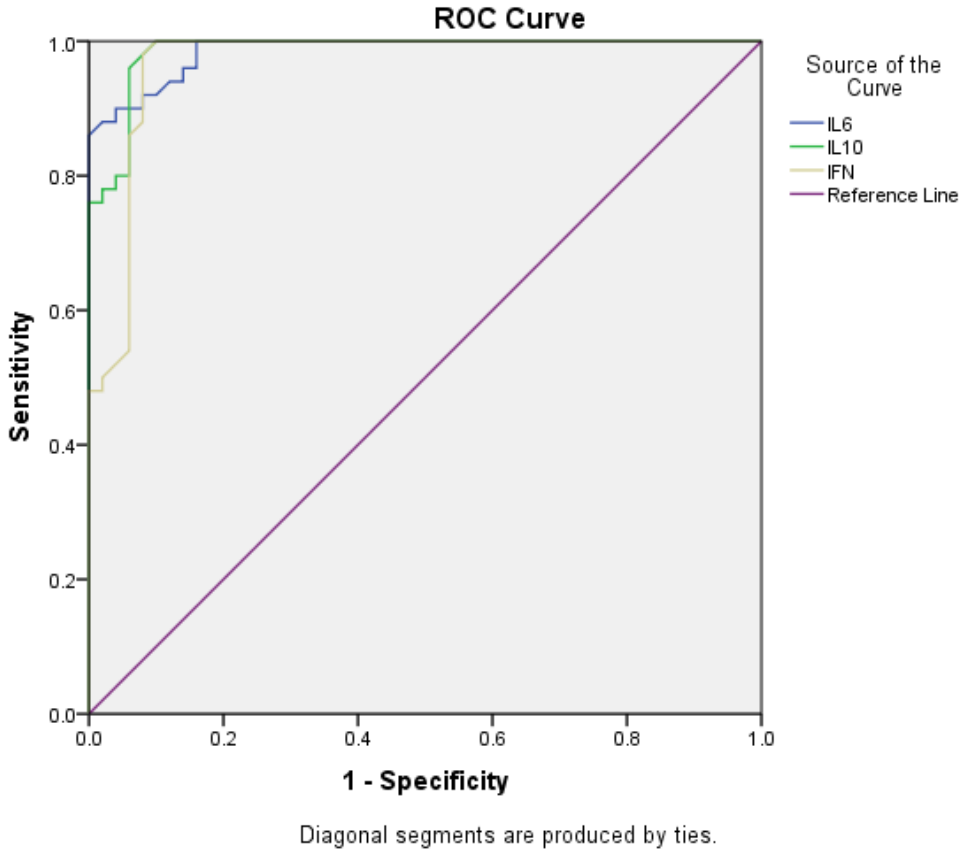



**Figure 1: ROC curve showing the diagnostic performance of IL–6, IL–10 and IFN–γ based on ROC curve analysis in dengue patients and healthy**

## Discussion

Demographical results of the present study in table (1) accompanied with(Ghafoor et al., 2026) the participants' average age was 36.16 years ±14.42 years, and 58 people (52.3%) were under 35 years old. The majority of them lived in cities (104; 93.7%), with 74 (66.7%) of them being men.

The results of current study in table (2) showed that there is significant increase in cytokines levels in dengue patients compared with healthy, these results accompanied with(GOMATHESWARI et al., 2026) in comparison to non-severe dengue (IL-10: 51.4±21.37, median=35) and healthy controls (4.05±0.45, median=3.5), the mean IL-10 in severe dengue cases (222±80.7 pg/mL, median=130) was considerably ($p<0.0001$) higher. On Days 1 and 4, severe dengue patients had considerably higher levels of IL-6 and IL-8 than non-severe dengue patients; these levels then decreased on Day 8, which corresponded with clinical recovery. However, lengthier hospital stays were found to be substantially correlated with elevated IL-8 levels, suggesting that it may play a function as an early indicator of illness development (Mahooti et al., 2026). An essential first line of defense against these viral infections is the host interferon (IFN) system (Hong et al., 2026).

The results of ROC in table (6) accompanied with (Hong et al., 2026) with area under the curve (AUC) values of roughly 0.7, receiver operating characteristic (ROC) analysis revealed that TNF-α, IL-1β, and IL-10 significantly separated mild from severe cases. An "all-positive" panel consisting of TNF-α, IL-10, and IL-1β was able to diagnose DFWS/SD with 84% sensitivity and 67.5% specificity. This finding is consistent with empirical data from dengue case-screening studies where, despite variations in decision rules, complexity, and explain ability, multilayer perceptron, decision tree, and logistic regression models provide almost comparable

AUCs of 0.98–0.99 (Bohm et al., 2024). Gradient boosting also produces a high AUC (0.97) in models predicting severe dengue in Puerto Rico; nevertheless, the incremental advantage over simpler models mostly appears at clinically determined cut-offs, where higher sensitivity and NPV are highlighted (Madewell et al., 2025). (Jasen, 2025) bolster the mounting evidence that machine learning algorithms can significantly enhance clinical risk classification by spotting intricate nonlinear correlations between clinical and immunological variables that are frequently challenging to find with traditional statistical techniques.

The results of correlation between study parameters and dengue severity in table (3) accompanied with(GOMATHESWARI et al., 2026) The severity of dengue and IL-10 were significantly correlated ($p<0.05$). When patients exhibit warning signals, early detection of these cytokines, IL-6 and IL-8, helps differentiate between severe and non-severe dengue. IL-8 and IL-6 and may be useful as indicators of the severity of a disease (Sivasubramanian et al., 2026). When compared to dengue with warning signs, severe dengue cases had significantly higher levels of TPO, IL-8, IL-1β, IL-1RA, IL-6, IL-10, IL-2, IL-18, and IFN-γ ($p < 0.05$)(Puc et al., 2021).Dengue pathogenesis is significantly influenced by interleukin-10 (IL-10), which has an immunosuppressive impact that results in interferon (IFN) resistance, poor immunological clearance, and a sustained infectious effect for acute viral infection (GOMATHESWARI et al., 2026).

The current study's feature importance analysis revealed that IFN-γ was the best indicator of severe dengue, followed by IL-10 and IL-6. By capturing many aspects of the host immune response, these cytokines collectively offer a more thorough evaluation of the course of the disease than can individual biomarkers. Recent machine learning research examining biomarker-based prediction models for inflammatory and infectious diseases has revealed similar feature-ranking trends (Saha et al., 2026).

The use of cutting-edge artificial intelligence techniques on immunological data is a significant strength of the current work. In every evaluation metric, Random Forest consistently performed better than Support Vector Machine and XGBoost. Because Random Forest efficiently handles nonlinear interactions, minimizes overfitting through bootstrap aggregation, and produces reliable estimates of variable importance, it is especially well suited for biomedical datasets. Recent research on COVID-19 prognosis, sepsis prediction, infectious disease diagnosis, and biomarker identification has consistently shown these traits (Cozzolino et al., 2025).

The Random Forest model's outstanding classification accuracy further emphasizes the potential therapeutic value of incorporating cytokine measures into standard dengue treatment. Closer clinical surveillance, prompt fluid management, early intensive care referral, and more effective use of healthcare resources would all be made possible by early identification

of individuals who are likely to develop severe disease. AI-assisted decision-support systems could significantly enhance patient outcomes while lowering needless hospitalization of low-risk individuals in areas where dengue epidemics often overwhelm healthcare facilities (Organization, 2026).

Instead of direct viral cytopathic effects, dysregulated host immune responses play a major role in the pathogenesis of severe dengue. When innate and adaptive immune pathways are overactivated, a "cytokine storm" of inflammatory mediators is released. Coagulation irregularities, increased vascular permeability, endothelial dysfunction, and multi-organ involvement are all influenced by elevated cytokine production. interferon-gamma (IFN-γ), Interleukin-6 (IL-6) and interleukin-10 (IL-10) have repeatedly shown high correlations with illness severity and unfavorable clinical outcomes among the many cytokines studied (Dhole et al., 2024).

## conclusion

This study showed that combining artificial intelligence with immunological indicators offers a reliable and efficient method for the early detection of severe dengue. The severity of the condition was substantially correlated with serum levels of IL-6, IL-10, and IFN-γ, with IFN-γ being the greatest individual prediction, followed by IL-10 and IL-6. Compared to relying solely on individual biomarkers, the combined use of these cytokines

allowed for a more accurate diagnosis of patients at high risk of developing severe dengue.

## References

Bohm, B. C., Borges, F. E. d. M., Silva, S. C. M., Soares, A. T., Ferreira, D. D., Belo, V. S., Lignon, J. S., & Bruhn, F. R. P. (2024). Utilization of machine learning for dengue case screening. BMC Public Health, 24(1), 1573.

Cozzolino, C., Mao, S., Bassan, F., Bilato, L., Compagno, L., Salvo, V., Chiusaroli, L., Cocchio, S., & Baldo, V. (2025). Are AI-based surveillance systems for healthcare-associated infections ready for clinical practice? A systematic review and meta-analysis. Artificial Intelligence in Medicine, 165, 103137.

Cracknell Daniels, B., Buddhari, D., Hunsawong, T., Iamsirithaworn, S., Farmer, A. R., Cummings, D. A., Anderson, K. B., & Dorigatti, I. (2024). Predicting the infecting dengue serotype from antibody titre data using machine learning. PLoS computational biology, 20(12), e1012188.

Dhole, P., Zaidi, A., Nariya, H. K., Sinha, S., Jinesh, S., & Srivastava, S. (2024). Host immune response to Dengue virus infection: friend or foe? Immuno, 4(4), 549-577.

Fatimata Belem, W., Sangare, I., Gilbert, C., & Wenceslas Bazié, W. (2026). Dengue Fever and Its Burden in Burkina Faso: An Overview. Reviews in Medical Virology, 36(1), e70089.

Gainor, E. M., Harris, E., & LaBeaud, A. D. (2022). Uncovering the burden of dengue in Africa: considerations on magnitude, misdiagnosis, and ancestry. Viruses, 14(2), 233.

Ghafoor, S., Ahmed, K., & Waqar, I. (2026). Cardiovascular manifestations among dengue fever patients at a tertiary care hospital. Infectious Diseases Journal of Pakistan, 35(1), 44. 49-

GOMATHESWARI, S. N., Poornakala, R., JEYAMURUGAN, T., & DEVI, C. S. (2026). Assessment of Serum Interleukin-10 as a Marker of Severity in Dengue Virus Infection: A Cross-sectional Study. Journal of Clinical & Diagnostic Research, 20(3), 24.

Hong, C-.T., Liu, L.-T., Chen, P.-C., Chen, C.-H., Tsai, C.-Y., Lin, P.-C., Hsu, M.-C., & Tsai, J.-J. (2026). Profiling of plasma exosome cytokines as biomarkers of severe dengue. Frontiers in immunology, 17, 1779100. .

Im, J., Balasubramanian, R., Ouedraogo, M ,.Nana, L. R. W., Mogeni, O. D., Jeon, H. J., van Pomeren, T., Haselbeck, A., Lim, J. K., & Prifti, K. (2020). The epidemiology of dengue outbreaks in 2016 and 2017 in Ouagadougou, Burkina Faso. Heliyon, 6. (7)

Jasen, S. (2025). Explainable Artificial Intelligence in Healthcare. In Applied Artificial Intelligence in Business: Systems, Tools and Techniques (pp. 615-624). Springer.

Madewell, Z. J., Rodriguez, D. M., Thayer, M. B., Rivera-Amill, V., Paz-Bailey, G., Adams, L. E., & Wong, J. M. (2025). Machine learning for predicting severe dengue in Puerto Rico. Infectious Diseases of Poverty, 14(01), 50-66.

Mahooti, M., Ahmad, S., Ahmadbeigi, G., Abdolalipour, E., Safaei, F., Mahooti, K., Eslami, M., Yaghmayee, S., Sanami, S., & Pajand, O. (2026). Efficacy of Probiotics and Postbiotics in Treating Dengue Fever–A Potential Strategy to Reduce Antibody-dependent Enhancement. Probiotics and Antimicrobial Proteins, 1-19.

mondiale de la Santé, O., & Organization, W. H. (2025). Dengue: global situation, surveillance and progress–2024 update= Dengue: situation mondiale, surveillance et progrès–mise à jour 2024. Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire, 100(52), 665-678.

Nakhaie, M., Shahpar, A., Rezaei Zadeh Rukerd, M., Farsiu, N., Charostad, J., Bashash, D., Zeinali Nezhad, N., Pardeshenas, M., Haghi Navand, A., & Mirkamali, H. (2026). Dengue Fever: Viral, Environmental, and Human Factors Driving Expansion and Pandemic Risk. Reviews in Medical Virology, 36(1), e70088.

Noman, A. A., Das ,D., Nesa, Z., Tariquzzaman, M., Sharzana, F., Hasan, M. R., Riaz, B. K., Sharower, G., & Rahman, M. M. (2023). Importance of Wolbachia-mediated biocontrol to reduce dengue in Bangladesh and other dengue-endemic developing 77-79.

Organization, W. H. (2014). Dengue and severe dengue countries. Biosafety and Health 5(02).

Organization, W. H. (2026). Epidemiological Bulletin WHO Health Emergencies Programme WHO Regional Office for South-East Asia 25th edition (2025), 17 December 2025 Reporting period: 01 to 14 Dec 2025.

Pennisi, F., Pinto, A., Borgonovo, F., Scaglione, G., Ligresti, R., Santangelo, O. E., Provenzano, S., Gori, A., Baldo, V., & Signorelli, C. (2026). Artificial Intelligence Models for Forecasting Mosquito-Borne Viral Diseases in Human Populations: A Global Systematic Review and Comparative Performance Analysis. Machine Learning and Knowledge Extraction, 8(1), 15.

Pennisi, F., Pinto, A., Cozzolino, C., Cozza, A., Rezza, G., Signorelli, C., Baldo, V., & Gianfredi, V. (2026). Comparative Diagnostic Performance of Artificial Intelligence Versus Conventional Approaches for Early Detection of Mosquito-Borne Viral Infections: A Systematic Review and Meta-Analysis, with Evidence

Predominantly from Dengue Studies. Machine Learning and Knowledge 8(4), 93. Extraction,

Puc, I., Ho, T.-C., Yen, K.-L., Vats, A., Tsai, J.-J., Chen, P.-L., Chien, Y.-W., Lo, Y.-C., & Perng, G. C. (2021). Cytokine signature of dengue patients at different severity of the disease. International Journal of Molecular Sciences, 22(6), 2879.

Saha, S., Ali, M. S., Tengli, A. K., Prasad, S. R., Mallikarjunaswamy, P., Pillappan, R., & Javarappa, K. K. (2026). Navigating AI and machine learning in cancer research: an end-to-end translational framework. Journal of translational. medicine

Shoushtari, M., Bakhshi, H., Salehi-Vaziri, M., Zaim, M., Enayati, A., Pouriayevali, M. H., Moradi, G., Sedaghat, M. M., Mirolyaei, A., & Mostafavi, E. (2026). Distribution of Aedes aegypti and Aedes albopictus, and the current situation of dengue fever and chikungunya in Iran and neighboring countries: a review study. PLOS Neglected Tropical Diseases, 20(2), e0013965.

Sivasubramanian, K., Bharath, R. R., Vajravelu, L. K., Kumar D, M., & Pamarthi, J. (2026). Assessment of IL-6 and IL-8 Levels and Other Bio Markers in Predicting Dengue Severity Across Serotypes. Pathogens, 15(4), 434.

# AN ARTIFICIAL INTELLIGENCE FRAMEWORK FOR PREDICTING SPORTS INJURIES

Saad Abbas Fadhil[1], Zainab Abbas Fadhil[2]

[1]University of Baghdad, College of Physical Education and Sport Sciences, Baghdad, Iraq. Saad.a@cope.uobaghdad.edu.iq

[2]Network Department, Faculty of Engineering, Al-Iraqia University, Baghdad, Iraq, drzainababbas13@gmail.com

## Abstract

Background: Sports-injury risk and the trajectory of rehabilitation result from the interplay between signals such as workload, recovery, biomechanics, history, and clinical parameters. Artificial intelligence (AI) holds the potential to combine these signals; yet barriers include small sample sizes, heterogeneity in the definition of injuries, lack of external validation, and opacity of models. Objectives: Develop and illustrate an interpretable AI-based framework for estimating the risk of sports injury within 28 days and transforming model explanations into clinician-guided injury prevention and rehabilitation review procedures. Methods: A synthetic sample of 60 athletes from football, basketball, wrestling, and athletics was

created using fixed-seed generation techniques. Ten predictor variables were chosen to represent demographic information, workload, recovery status, wellness measures, prior injury status, and strength imbalance. Four AI models—random forest, radial basis function support vector machine, artificial neural network, and XGBoost—were compared against a benchmark (logistic regression) incorporating L2 regularization with stratified five-fold cross-validation. Measures considered include discrimination, classification, calibration (probability prediction accuracy), and groupwise permutation importance. A decision-support component associates dominant modifiable signals with verification, intervention, and follow-up. Results: Eighteen athletes (30.0%) experienced a simulated injury event. Injured athletes had significantly higher acute to chronic workload ratio (ACWR) (1.27 ± 0.15 vs. 1.10 ± 0.18, $p < .001$, $d = 1.04$), lower sleep time (6.77 ± 0.51 vs. 7.25 ± 0.62 hours, $p = .003$, $d = -0.81$), and increased strength imbalance (11.49 ± 3.62% vs. 9.16 ± 3.98%, $p = .033$, $d = 0.60$). The benchmark model attained the best performance in terms of area under curve (AUC) (.843), balanced sensitivity (.722), and specificity (.833). XGBoost (AUC = .825) and support vector machine (AUC = .824) performed closely, where support vector machine showed the lowest Brier score (.135). Dominant predictor variables included ACWR, sleep, strength imbalance, and prior injury. Conclusion: This framework illustrates the way that complimentary AI models can be used to provide transparent injury risk

estimates and generate structured rehabilitation review prompts. Since a synthetic cohort and a rehabilitation component are involved without an evaluation of the latter as an intervention, this work establishes workflow feasibility more than clinical effectiveness and thus calls for prospective multisite validation.

**Keywords**: machine learning; sports injury; injury prediction; rehabilitation; decision support; random forest; XGBoost.

## Introduction

However, unlike trauma, sports injury results from the interaction between tissue tolerance, training workload accumulated over time, recovery processes, previous injury history, quality of movements, psychological state of the player, circumstances of play, and trigger events. Due to this complex nature of injury, it is hard to predict – what is a normal situation for one person may cause problems in another or even for the same person throughout the game. In light of these considerations, the modern sports practice favors regular screening instead of a once per season examination.

Artificial intelligence tools help to combine information from various sources and discover nonlinear relations or relations of certain conditions which might be missed by traditional analyses using only one risk factor. The systematic literature reviews show that artificial neural networks, decision

trees, support-vector machines and ensemble methods have been used in team sport contexts, and the accurate results do not necessarily imply clinical utility (Claudino et al., 2019; Van Eetvelde et al., 2021). Recently published scoping review demonstrated that area under the curve (AUC) was the most popular indicator of model quality and tree-based models showed superior performance but small sample sizes, varying time frames and vague definitions of injuries were still widely present in research (Leckey et al., 2025).

Selected algorithms represent two different approaches to learning from data. Decision trees in random forest are decorrelated; support vector machines find the optimal separating hyperplane; artificial neural networks create nonlinear distributed representations of data; XGboost creates regularized boosting of decision trees (Breiman, 2001; Chen & Guestrin, 2016; Cortes & Vapnik, 1995; Rumelhart et al., 1986). Thus, comparing performance of these algorithms in the same conditions is more meaningful than choosing them on reputation because performance depends not only on algorithm but on sample size, signal nature and class imbalance as well as on clinical consequences.

Hence, methodological confidence is a key point to address. As per Bullock et al. (2022), several pitfalls exist in developing and presenting sport-injury prediction models. The common recurring issues are too many predictors as compared to the number of injuries, information leakage from the training

to the testing set, improper treatment of class imbalance, selective presentation of accuracy, and lack of either calibration or external validation. Models might work fine in one team, sport, season, gender, or level of competition but fail to work for another one. Explainability is crucial since clinicians and coaches need to find out if a risk factor is related to a modifiable exposure, personal trait, or is an artefact of data.

Workload is one of the debated and potentially very important elements of injury-risk modelling. Various measures such as ACWR, total training volume, perception, exposure time, and change in high-intensity activity are good to use as signals; still, one needs to treat all those measures without seeing them as thresholds for injury. Training load depends on the amount of sleep, wellness status, strength asymmetry, and previous injury. While white-box models make all dependencies evident, more sophisticated models reveal additional patterns, but those need to be explained with some techniques of explainable AI like permutation importance or SHAP (Majumdar et al., 2022; Ye et al., 2023).

In addition to the mere injury classification, the clinical purpose of prediction goes further to assist decision making throughout prevention and rehabilitation. For prevention, before the injury, probability estimation will help clinicians to have a proper proportionate look at load progression, sleep

adequacy, symptoms or asymmetry. After the injury, repeated assessments of pain and ROMs and the tolerance of loading help to build individualized criteria for progression and readiness for RTSS. Active joint mobilization has been found to contribute to dorsiflexion and dynamic stability among athletes with recurrent ankle injuries (Fadhil & Khalaf, 2023), while knee injury rehabilitation needs individualized milestones and RTSS assessments (Fadhil & Fadhil, 2026). Presently in the proof-of-concept study, AI arranges the risk factors data and makes suggestions about review; diagnosis, prescription, or clearance of patients are done separately.

The purpose of the current study was to create and illustrate an interpretable AI model that can predict a sports injury outcome over a span of 28 days and tie the explanation from such models to the decision-making by clinicians to prevent or rehabilitate sports injuries. Another purpose of this study was to test the efficacy of four different machine-learning algorithms (Random Forest, Support Vector Machine, Artificial Neural Network, XGBoost) against a regularized logistic regression algorithm. The null hypothesis of the study is that the framework will perform with sufficient discriminative ability; and that the predictor variables such as ACWR, injury history, strength asymmetry, and sleep will become important predictors.

## Methods

### Study Design and Transparency

It is a methodology study, which was performed using a fully simulated cross-sectional dataset on the level of an athlete. It contains a 28-day prospective outcome of injury with a rehabilitation decision support component. This dataset was simulated for the purpose of proving the concept and demonstrating how the process of the analysis and reporting would go and is not an actual dataset of enrolled human participants or medical records from a club. In addition, the rehabilitation part was described as clinician-based decision logic instead of tested treatments. For that reason, neither ethics committee clearance nor informed consent was required.

Simulated Participants and Outcome

In the data set, there were 60 athletes divided into football players (n = 20), basketball players (n = 15), wrestlers (n = 13) and athletes (n = 12). Their age was between 18 and 34 years, and they were male and female athletes. The binary classification variable "Injury_next_28d" was set as an invented musculoskeletal injury that would result in the absence of at least one training/competition session due to the injury in 28 days of predictions. There were 18 athletes (30.0%) assigned.

## Simulation Procedure

The sport counts were set in advance, and all the other predictors that were used were obtained from bounded distributions reflecting realistic values for athlete monitoring. A latent score was formed so that more weight could be given to ACWR, history of injury, strength asymmetry, and sleep, and noise was added in order to make the prediction deterministic. Then, the highest 18 latent scores were mapped to the injury outcomes for 28 days ahead of time. Age, BMI, weekly training load, and wellness were selected as control predictors. The absence of missing values is intentional.

## Candidate Predictors

There were ten factors considered: type of sports, gender, age, body mass index (BMI), training load per week, ACWR, average sleep hours per night, wellness level, previous injuries, and strength asymmetry. Training load was expressed in arbitrary units in relation to the session RPE method. ACWR is a ratio between recent and longer training loads. Wellness level was assessed from one to ten; high levels represented good perception of the person's condition. Previous injuries were binary; strength asymmetry was expressed in percent. The factors were chosen based on the regularity of their availability through athlete management tools, questionnaires, field tests, or wearable technologies.

## Framework Architecture

This proposed structure consists of six layers, namely: (1) data acquisition from athletes, clinicians, training systems, and wearables; (2) data validation and preprocessing including range checks, missingness analysis, standardization, and encoding of categories; (3) model estimation through resampling methods preserving the rate of occurrence of outcomes; (4) interpretation using global importance of predictors and probability for the athlete; (5) decision support for clinicians for injury prevention and rehabilitation; and (6) feedback, recording of outcomes, recalibration, and updating of the model. The predictions are presented by the probability estimate with the contributing signals. The rehabilitation recommendation is comprised of a verification question, modifiable target if there is one, monitoring signal, and reassessment criterion. Data management, access control, versioning of models, subgroup audit, and periodic recalibration cover all six layers.

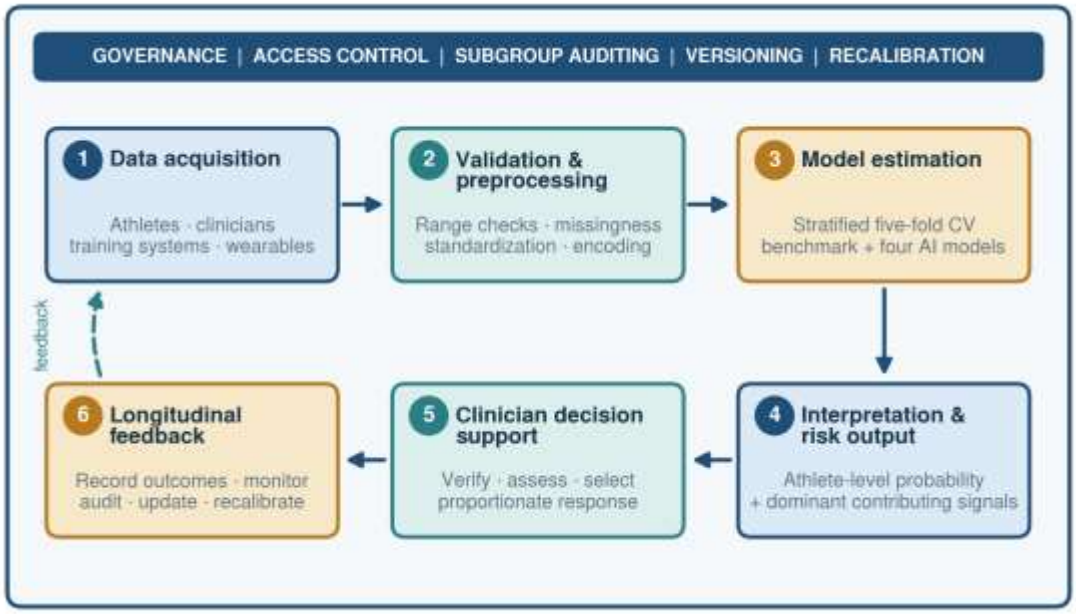



**Figure 1. Six–layer architecture of the proposed AI framework for sports–injury prediction and rehabilitation decision support.**

Machine-Learning Models

Benchmark: Regularized Logistic Regression

The use of L2 regularized logistic regression was done as the benchmark with transparency. This used C = 1.0 and inverse class weight proportional to the frequency of the results. The estimand in this model is that of additivity in log odds ratio.

## Random Forest

Random forests combine bootstrapping techniques with the selection of random features for each decision tree in the forest, thereby allowing for the capturing of threshold interactions as well as the reduction of the variability caused by each tree individually (Breiman, 2001). In the current

scenario, there are 500 trees, a maximum depth of 3, a minimum size per terminal node of 4, and balanced class weights. These settings were chosen to avoid overfitting in the small sample group.

Support Vector Machine

The Support Vector Machine (SVM) attempts to find the optimal decision boundary that separates the classes with maximum margin and allows for the nonlinear nature of the data through the use of the kernel method (Cortes & Vapnik, 1995). For the current research work, an RBF kernel was used with C = 1.0, gamma set based on the variance of the scaled features, and the weight of the classes was set to be equal based on the probability of folds.

**Artificial Neural Network**

The ANN used had a multilayer perceptron architecture where there was one hidden layer having eight rectified-linear units. Optimization involved the use of the L-BFGS algorithm where the value of the L2 penalty alpha was 0.05 with a limit of 3,000 iterations (Rumelhart et al., 1986). The simple model was selected since we had very little data, and 60 data points would be insufficient for a complex network. In future, with longitudinal studies, the ANN could include repetitive measurements from sensors as well as functionality; but this would require more data and explainability.

XGBoost

The XGBoost is a regularized gradient boosting system whereby successive trees model the residual classification error while shrinkage and structure penalties control the model complexity (Chen & Guestrin, 2016). The pre-defined model parameters were set for 80 boosting iterations with a learning rate of 0.05, tree depth of 2, a minimum child weight of 1, subsampling of 0.90, an L2 regularization of 2.0, and class balancing by folds. The suggested benefit was the ability to effectively model non-linear interaction in structured data on athletes.

Statistical Analysis, Model Training, and Validation

Continuous variables were presented as mean ± SD and compared between groups of outcomes using Welch's t-test of independent samples. Standardized mean differences are shown in terms of Cohen's d. Categorical variables were examined with chi-squared tests with Yates' continuity correction for 2 × 2 tables. Statistical inferences are made concerning the sampled data and do not represent population parameters.

The missing numerical values were imputed by means and then standardized in the course of preprocessing. The categorical variables were mode imputed and converted into dummy variables. Cross-validation procedure was implemented through stratified 5-fold approach in order to avoid leaking of information. Hyperparameter optimization was not performed before

evaluation of models because no parameter searching was undertaken. It was necessary to minimize the bias in the experiment with 60 athletes. Out-of-the-fold predictions were used for calculating AUC, accuracy, sensitivity, specificity, precision, F1 and Brier scores. The classification threshold was 0.50. The chosen model is based on discrimination, the compromise between sensitivity and specificity, accuracy of prediction and interpretability. Group-wise permutation importances represent the mean decrease of AUC in test folds with shuffling an original variable in training folds; if negative, there is no stable contribution of the variable in current experiment setup. Analysis was performed with the use of Python with fixed random seed (20260109).

Rehabilitation Decision-Support Layer

The Rehabilitation layer was an interpretive pathway that was predetermined instead of being another training model for outcome prediction. For every athlete, the layer would use the risk of injury, the dominant signal, symptoms, and the entered clinical findings in order to provide four outcomes: data validation, assessment focus, a possible target for prevention or rehabilitation, and an indicator for reassessment. These outcomes were still advisory and required confirmation by the clinician in respect to diagnosis, contraindication, exercises, dose, and progression.

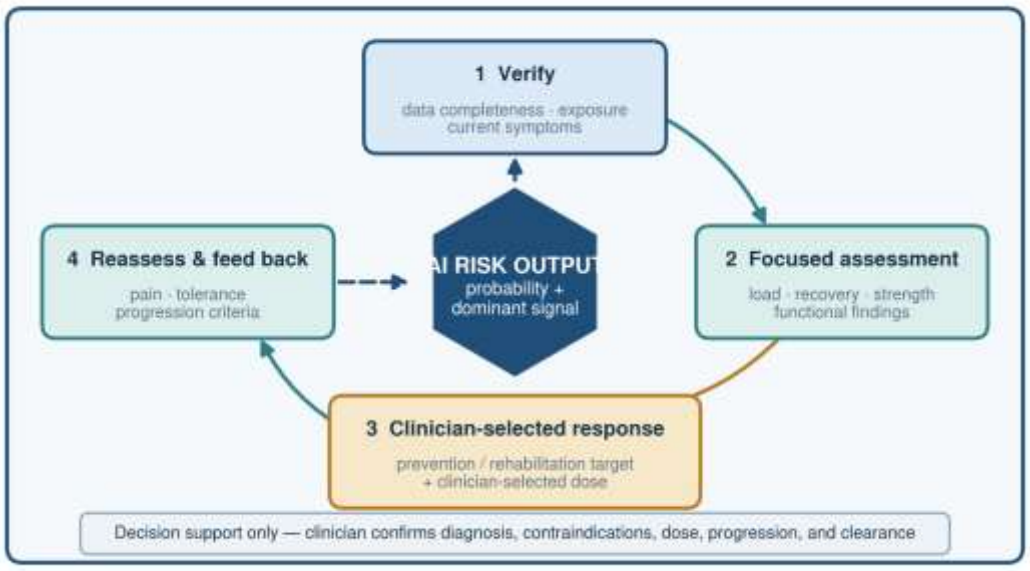



**Figure 2. Clinician–governed pathway translating injury–risk predictions into verification, focused assessment, proportional response, and reassessment.**

The future model of post-injury requires longitudinal measures of outcome that are not present in the current database, which include return to sport timing, recurrence of injury within a certain time period, pain, range of motion, strength balance, dynamic balance, and tolerance to gradual loading specific to the sport. In this regard, the current study only deals with the prediction of injury; Figure 4 demonstrates the application of explanatory results in developing a safe review process.

## Bias Control and Intended Use

This tool is meant to be used for risk stratification and for rehabilitation decision support, and not for diagnostic or standalone prescription purposes. An elevated score must be followed by an evaluation

of the quality of data, symptomatology, progression of loading, recovery process, and clinical factors. In the case of an injured athlete, progression depends on a specific examination of tissues, identification of contraindications, and achievement of specified functional criteria. Any model's result must not independently determine the exclusion of the athlete from participation and approval for returning to sport activities. In case of using actual data, the development must follow transparent reporting principles, use temporally distinct validation if possible, assess the calibration and decision-utility performance, and performance by sex, sport, competition level, injury, and rehabilitation phase.

Results

The simulated cohort was made up of 42 uninjured athletes and 18 athletes who suffered an injury. The injured group had relatively high ACWR, shorter sleep time, greater strength asymmetry between limbs, and greater proportion of previous injuries than the uninjured group. There were no differences noted between the two groups for age, body mass index, weekly training load, and wellness status. The distribution of outcomes is shown in Figure 3 below.

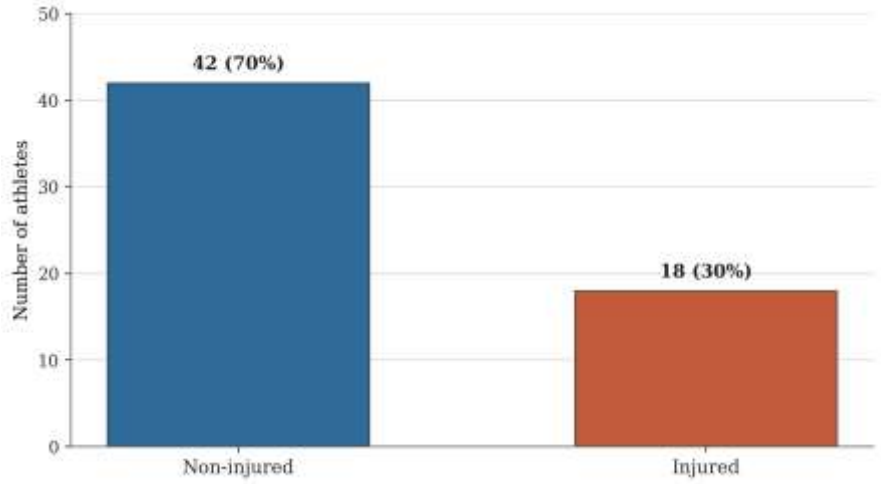



**Figure 3. Simulated 28–day injury status (N = 60); bar labels show n (%).**

**Table 1. Continuous characteristics by simulated 28–day injury status**

| Variable | Non–injured (n = 42) | Injured (n = 18) | p | Cohen's d |
|---|---|---|---|---|
| Age (years) | 25.12 ± 4.08 | 24.84 ± 4.15 | .806 | -0.07 |
| BMI (kg/m²) | 24.23 ± 2.14 | 24.86 ± 1.93 | .275 | 0.30 |
| Weekly load (AU) | 3701.38 ± 770.12 | 3641.00 ± 710.80 | .771 | -0.08 |
| ACWR | 1.10 ± 0.18 | 1.27 ± 0.15 | < .001 | 1.04 |
| Sleep (h/night) | 7.25 ± 0.62 | 6.77 ± 0.51 | .003 | -0.81 |
| Wellness (1–10) | 6.84 ± 1.40 | 6.89 ± 1.10 | .877 | 0.04 |

| Variable | Non–injured (n = 42) | Injured (n = 18) | p | Cohen's d |
|---|---|---|---|---|
| Strength asymmetry (%) | 9.16 ± 3.98 | 11.49 ± 3.62 | .033 | 0.60 |

*Note. Values are derived from the fully simulated dataset. p values are from Welch tests; a positive Cohen's d indicates a higher mean in the injury group. AU = arbitrary units; ACWR = acute:chronic workload ratio.*

ACWR showed the highest effect size between groups (d = 1.04), followed by sleep (d = −0.81) and strength imbalance (d = 0.60). Previous injuries were recorded in 55.6% of the injured group and 16.7% of the non–injured group (chi–square Yates correction p = .006). Inasmuch as the data is simulated, the p values presented refer only to the sample and not the population.

**Table 2. Five–fold cross–validated model performance**

| Model | AUC | Acc. | Sens. | Spec. | Prec. | F1 | Brier |
|---|---|---|---|---|---|---|---|
| Regularized logistic regression | .843 | .800 | .722 | .833 | .650 | .684 | .154 |
| Random forest | .794 | .700 | .444 | .810 | .500 | .471 | .180 |
| Support vector machine (RBF) | .824 | .817 | .611 | .905 | .733 | .667 | .135 |

| Model | AUC | Acc. | Sens. | Spec. | Prec. | F1 | Brier |
|---|---|---|---|---|---|---|---|
| Artificial neural network | .796 | .750 | .667 | .786 | .571 | .615 | .208 |
| XGBoost | .825 | .767 | .611 | .833 | .611 | .611 | .149 |

*Note. Values are out-of-fold estimates from the fully simulated dataset at a .50 probability threshold. AUC = area under the receiver operating characteristic curve; RBF = radial basis function. Lower Brier scores indicate more accurate probabilities.*

It should be noted that regularized logistic regression has been chosen as the best model because of the high value of the area under the curve (AUC), which is 0.843, and quite balanced values for sensitivity (0.722) and specificity (0.833). At the same time, XGBoost provided 0.825 of AUC, while the support vector machine (SVM) provided 0.824 of AUC. It means that the discriminative power of models is almost similar. At the same time, the SVM showed the highest values of accuracy (0.817), specificity (0.905), and precision (0.733) and had the lowest Brier score (0.135), although with the sensitivity value of 0.611. The artificial neural network found 66.7% of injured patients, although it had the highest value of the Brier score (0.208), while the random forest found only 44.4% of patients.

**Table 3. Permutation importance for the preferred model**

| Predictor | Mean AUC decrease | SD | Interpretive direction |
|---|---|---|---|
| ACWR | .169 | .130 | Higher risk when elevated |
| Sleep | .081 | .115 | Protective when higher |
| Strength asymmetry | .059 | .074 | Higher risk when elevated |
| Previous injury | .050 | .082 | Higher-risk history |
| Sport | .007 | .030 | Small contextual contribution |
| Wellness | -.006 | .045 | No stable incremental value |
| Weekly load | -.007 | .030 | No stable incremental value |
| Age | -.009 | .091 | No stable incremental value |
| Sex | -.014 | .026 | No stable incremental value |
| BMI | -.027 | .053 | No stable incremental value |

*Note. Values are cross-validated groupwise permutation results for the preferred logistic model. A negative mean indicates no stable gain in out-of-fold AUC, not a protective clinical effect. ACWR = acute:chronic workload ratio; SD = standard deviation across folds and repeats.*

ACWR turned out to be the best predictor using the cross-validated group-wise permutation importance measure, followed by sleep, strength asymmetry, and previous injury. Sport made a minor contribution. Wellness, load per week, age, gender, and BMI showed close to zero or negative average importance values, which suggests that these factors did not discriminate well out of fold in the present simulation. The above ranking should not be considered as a global ranking of the injury predictors.

Prediction-to-Rehabilitation Translation

The framework converts the signal of the dominant model into a review process. The desired mapping is shown in Table 4 below. The above shows some of the issues to consider when doing professional reviews.

**Table 4. Example clinician-governed mapping from model signal to rehabilitation review**

| **Dominant signal** | **Verification and assessment** | **Possible clinician-selected response** | **Reassessment indicator** |
|---|---|---|---|
| Elevated ACWR or abrupt load change | Confirm exposure, session-RPE, pain, fatigue, and data completeness | Adjust short-term load if indicated, then use graded sport-specific re-exposure | Pain response, load tolerance, and symptom-free progression |

| **Dominant signal** | **Verification and assessment** | **Possible clinician-selected response** | **Reassessment indicator** |
|---|---|---|---|
| Strength asymmetry | Repeat strength, movement-quality, balance, or hop testing appropriate to the injury | Targeted unilateral strength and neuromuscular rehabilitation | Symmetry, task quality, and sport-specific functional criteria |
| Short sleep or deteriorating recovery | Review sleep opportunity, fatigue, wellness trend, illness, and psychosocial stress | Recovery and sleep intervention; medical review when symptoms persist | Stable sleep, wellness, and training-tolerance trend |
| Previous injury or current symptoms | Perform tissue-specific clinical examination and review prior rehabilitation deficits | Criterion-based rehabilitation, modified participation, and graded return to sport | Pain, range of motion, strength, function, and reinjury surveillance |

*Note. The mapping is a safety-oriented decision aid. A qualified clinician confirms diagnosis, contraindications, intervention dose, progression, and return-to-sport clearance.*

## Discussion

The main finding is the development of a clinically interpretable framework which can assess four complementary artificial intelligence algorithms together with a benchmark algorithm with regularization within a single leakage-resistant process to translate the predominant risk patterns into guided clinician rehabilitation review. Logit regression with regularization obtained the greatest area under the curve and balanced distribution of false negatives and false positives, while the support vector machine produced the smallest Brier score and the highest specificity. XGBoost was close to both in terms of discrimination ability. Artificial neural network and random forest did not outperform the simple logistic regression model on this dataset consisting of 18 event outcomes. This finding is further confirmation that complexity of models is no replacement for good data, calibration, and external validation.

Importance of variables corresponds to the mechanism of creating the illustrative dataset. Acute:chronic workload ratio characterizes sudden changes in the load exposure, previous injury denotes reduced or changed tissue tolerance, strength asymmetry describes a possible functional impairment which can be fixed and, finally, sleep serves as a potential time for restoration and recovery. These multi-domain predictors correspond to the idea that sports injuries are complex phenomena and not results of crossing some thresholds. Also, this combination supports the idea that

artificial intelligence could identify the athletes at risk and contributing factors to injuries as suggested by systematic reviews. Machine learning is not well understood yet and current evidence lacks consistency (Van Eetvelde et al., 2021; Leckey et al., 2025). Importantly, negative coefficients of predictor importance highlight the possibility that inclusion of variable does not equal to obtaining valuable information.

Logit benchmark performs better compared to other machine-learning algorithms from the perspective of discrimination and sensitivity. Support vector machine and XGBoost were characterized by higher specificity and smaller Brier scores, and artificial neural network found more events but gave less accurate probabilistic predictions than tree-based methods. Selection of appropriate threshold depends on the implications of the planned action. Low burden intervention requires lower threshold, however more invasive procedure, restrictions to participation in sports and return-to-sport assessment need more convincing data and clinical evaluation.

Interpretability assists with translating risk assessment into action by helping to inform interventions. For example, high signals of an ACWR may warrant rechecking recent exposure and graded loading; an asymmetric profile may warrant repeated measures of strength and movements; poor sleep may warrant recovery testing; and an injury history may lead to tissue-specific inspection. Previous research conducted by Fadhil and Khalaf (2023) showed how active joint mobilization could improve dynamic balance and

ankle dorsiflexion of recurrent ankle injuries, whereas Fadhil and Fadhil (2026) discussed evidence-based criterion knee rehabilitation. There is also research on rehabilitation following tennis elbow where loading decisions were based on impairment recovery (Fadhil & Qaddoori, 2023). However, the proposed model must organize these types of decision-making processes rather than deliver interventions after ignoring examination.

A more elaborate innovation pipeline would relate predictions to longitudinal rehabilitation devices. The 3D-printed prosthetic heel, developed by Kadhim et al. (2025), serves as an example that shows how biomechanical analysis and numerical analysis are related to developing personalized rehabilitation devices. Another study done by Zainab A. Fadhil and Mahdi (2026) on nano-encapsulated chitosan-collagen hydrogel for Achilles tendon healing demonstrates the biological part of tissue restoration. While these works are mainly focused on treatment rather than prediction of injury-risk and recovery, they serve as examples showing how future athlete health systems could include prediction of injury-risk, functional assessment, devices for personalized rehabilitation, and recovery monitoring.

The suggested framework includes several precautions. Firstly, all preprocessing stages along with any further parameter optimization must be done within the resampling stage to avoid the problem of data leakage. Secondly, prediction window, injury criteria, rehabilitation period, and outcome measure must be pre-specified before analysis starts. Thirdly, in

addition to AUC metric, sensitivity, specificity, precision, probability accuracy, and calibration need to be assessed. Lastly, subgroup performance needs to be audited, as the model trained only on football male players will probably not perform well among women, wrestlers, and track-and-field athletes. Also, risk and recovery scores must be delivered as probabilistic values instead of definitive labels indicating the inevitability of an injury or recovery readiness. Explainability approaches can have high variance and must be tested within different resamples and locations.

Some critical limitations are present in this study. All results were obtained from simulations, so associations and the models' performance can only illustrate workflow properties. The total number of athletes and sports events (60 and 18 respectively) is not enough for proper model training, hyperparameter tuning, subgroup analyses, or comparisons of an artificial neural network and boosted trees. Cross-validation addresses the issue of optimistic biases, but cannot replace external validation. The single-row per athlete approach prevents taking into account changing daily exposure, repetitive events, censoring, and within-athlete correlations. ACWR implies certain assumptions and has some statistical limitations and cannot be regarded as the general safety-zone criterion. No missing information or device variations were simulated. Perhaps the most critical limitation is the fact that rehabilitation mapping was not validated on recovery duration,

reinjury risk, functional improvement, or return to sport measures; consequently, the framework does not demonstrate treatment effect.

Future studies need to collect longitudinal athlete-day or athlete-week observation in multiple teams across seasons with prespecified injury and external temporally or geographically independent validation cohort. The rehabilitation mapping should include injury description and severity, pain, range of motions, muscle strength symmetry, dynamic balance, workload tolerance, rehabilitation treatment, return to sport time and risk of reinjuries. The study power should depend on the number of injuries and rehabilitation episodes, and anticipated complexity of models. Four algorithms should be tuned, calibrated, audited for fairness and compared to simple clinical benchmark. To prove usefulness and value of the framework, decision-curve and impact studies should be conducted.

## Conclusion

In this paper, we outline a clear framework using artificial intelligence to compare the predictive abilities of the random forest, support vector machine, artificial neural network, and XGBoost against regularized logistic regression, for sports-injury prediction for 28 days in a simulated cohort, and link their interpretations to clinicians' assessment during rehabilitation. Logistic regression had the best area under the curve and balanced classification performance, while the support vector machine had the lowest

Brier score among all models in this simulated cohort. The most important predictors identified were ACWR, sleep, strength asymmetry, and history of injury. The rehabilitation layer uses these indicators to prompt clinicians to check, monitor, and assess them again without substituting diagnosis and professional evaluation. It is important to note that since the simulation of the cohort and the numerical data was done and there is no experimental evaluation of the rehabilitation process, further validation is required before practical use.

## References

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bullock, G. S., Mylott, J., Hughes, T., Nicholson, K. F., Riley, R. D., & Collins, G. S. (2022). Just how confident can we be in predicting sports injuries? A systematic review of the methodological conduct and performance of existing musculoskeletal injury prediction models in sport. Sports Medicine, 52(10), 2469–2482. https://doi.org/10.1007/s40279-022-01698-9

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Claudino, J. G., Capanema, D. O., de Souza, T. V., Serrão, J. C., Pereira, A. C. M., & Nassis, G. P. (2019). Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: A systematic review. Sports Medicine - Open, 5, Article 28. https://doi.org/10.1186/s40798-019-0202-3

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20, 273-297. https://doi.org/10.1007/BF00994018

Fadhil, S. A. (2026). Review of rehabilitation protocols for ankle injuries. IFR Journal of Humanities, Social Science and Politics, 3(2), 41–45. https://doi.org/10.70146/hsspv03i02.0002

Fadhil, S. A., & Fadhil, Z. A. (2026). Evidence-based rehabilitation of knee injuries: A review. Technium Social Sciences Journal, 79(1), 386–395. https://doi.org/10.47577/tssj.v79i1.13748

Fadhil, S. A., & Khalaf, S. Q. (2023). A treatment strategy for recurrent ankle injuries in Muay Thai athletes. Eximia, 12(1), 486–495. https://doi.org/10.47577/eximia.v12i1.394

Fadhil, A. S. A., & Qaddoori, A. B. R. M. (2023). Rehabilitation program for treatment of tennis elbow (lateral epicondylalgia) suffered by the players of Al-Rafidain Sports Club for Tennis League Three players in Diyala Province. Res Militaris, 13(1), 2403–2414.

Fadhil, Z. A., & Mahdi, A. K. (2026). Effect of nano-encapsulated chitosan-collagen hydrogel on healing of injured Achilles tendon in a rabbit model. Journal of Animal Health and Production, 14(2), 416–427.

Halson, S. L. (2014). Monitoring training load to understand fatigue in athletes. Sports Medicine, 44(Suppl 2), S139–S147. https://doi.org/10.1007/s40279-014-0253-z

Kadhim, F. M., Fadhil, S. A., Abdullah, A. M., & Al-Din Tahir, M. S. (2025). Designing and manufacturing a flexible heel of a printed prosthetic foot for rehabilitation. International Review of Applied Sciences and Engineering, 16(2), 324–330. https://doi.org/10.1556/1848.2024.00930

Kalkhoven, J. T., Watsford, M. L., & Coutts, A. J. (2021). Training load and injury: Causal pathways and future directions. Sports Medicine, 51, 1137–1150. https://doi.org/10.1007/s40279-020-01413-6

Leckey, C., Toohey, L. A., Drew, M. K., et al. (2025). Machine learning approaches to injury risk prediction in sport: A scoping review with evidence synthesis. British Journal of Sports Medicine, 59(7), 491–503. https://doi.org/10.1136/bjsports-2023-107909

Majumdar, A., Bakirov, R., Hodges, D., Scott, S., & Rees, T. (2022). Machine learning for understanding and predicting injuries in football. Sports Medicine - Open, 8, Article 73. https://doi.org/10.1186/s40798-022-00465-4

Rossi, A., Pappalardo, L., & Cintia, P. (2022). A narrative review for a machine learning application in sports: An example based on injury forecasting in soccer. Sports, 10(1), Article 5. https://doi.org/10.3390/sports10010005

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323, 533–536. https://doi.org/10.1038/323533a0

Van Eetvelde, H., Mendonça, L. D., Ley, C., Seil, R., & Tischer, T. (2021). Machine learning methods in sport injury prediction and prevention: A systematic review. Journal of Experimental Orthopaedics, 8, Article 27. https://doi.org/10.1186/s40634-021-00346-x

Windt, J., & Gabbett, T. J. (2017). How do training and competition workloads relate to injury? The workload–injury aetiology model. British Journal of Sports Medicine, 51(5), 428–435. https://doi.org/10.1136/bjsports-2016-096040

Xu, Z., Sun, W., Qian, H., et al. (2025). Construction and application of a model for predicting athletes' injury risk based on machine learning. BMC Medical Informatics and Decision Making. Advance online publication. https://doi.org/10.1186/s12911-025-03331-x

Ye, X., Huang, Y., Bai, Z., & Wang, Y. (2023). A novel approach for sports injury risk prediction based on time-series image encoding and deep

learning. Frontiers in Physiology, 14, 1174525. https://doi.org/10.3389/fphys.2023.1174525

# Sequential Bit Selection and Chaotic Random Distribution in RGB Image Steganography Using the 3D Chua System

Ibrahim A. Mahdi[1]

[1]Polytechnic College - Karbala, Al-Furat Al-Awsat Technical University, Karbala, Iraq; Corresponding author: BBC4001@mtu.edu.iq

**Abstract**: This paper presents an efficient chaotic data hiding system based on the three-dimensional Chua chaotic system for embedding secret data into RGB color images using the Least Significant Bit (LSB) technique. The proposed approach first converts the secret image into a sequential bitstream, where the secret bits are selected sequentially and then randomly distributed across the RGB channels of the cover image using three independent chaotic sequences (X, Y, and Z) generated by the Chua system. Each chaotic sequence controls the selection of embedding locations for a single-color channel, resulting in a non-sequential spatial distribution of the embedded data that increases randomness, enhances security, and preserves image quality.

The proposed system was implemented and evaluated in the MATLAB environment using standard image quality and recovery metrics. Experimental results demonstrated excellent visual quality with a PSNR of 56.4237 dB, SSIM of 0.9997, MSE of 0.000002, Entropy of 7.9190, and an

embedding rate of approximately 14.82%, while maintaining minimal visual distortion. The extraction process achieved NC = 1.0000, BER = 0.0000, and ARE = 0.0000%, indicating perfect reconstruction of the hidden image. Furthermore, sensitivity analysis confirmed that even slight modifications to the chaotic initial conditions resulted in extraction failure, demonstrating the high level of security and robustness provided by the proposed chaotic framework.

## Introduction

Many sectors have expressed growing concerns about digitally securing information and ensuring its integrity during transmission, including civil, military, and commercial sectors. The protection of digital data is mainly based on two fundamental approaches: encryption and steganography, which can also be combined together. Traditional encryption methods have generally focused only on protecting the appearance and content of the data, but they failed to protect the existence of the data itself. Therefore, traditional encryption methods may still attract attention by allowing others to detect the existence of the data before accessing its content. On the other hand, basic steganographic methods based on simple techniques, such as conventional LSB techniques, often suffer from weak security levels and an inability to resist statistical and/or analytical attacks [1].

This study proposes utilizing the properties of chaotic systems, especially the three-dimensional modeling capabilities of these systems, to create

secure and random locations for embedding the required information. The use of these systems provides many advantages due to their high complexity and extreme sensitivity to initial conditions and operating parameters. Therefore, it is extremely difficult to detect the embedded data unless the exact embedding conditions are available. As a result, this methodology provides a reliable and secure approach for hiding data inside images [2].

## Motivation for the Study

Many people are now hiding confidential information in other digital files, such as image files or other media using various techniques. As technology advances, so has the development of ways to detect and differentiate between hidden data and visible data; therefore, it is necessary to create more secure and robust ways to hide data. Combining the two types of concealment methods with chaotic systems (as an added method of randomness) could create a better solution, as there are many unique characteristics that may be helpful, such as unpredictability based on an initial value and the ability to change the parameters used as secret maps for hiding data. By using these methods, one should be able to make their hidden information more difficult/impossible to detect [3].

The empirical results add to previous literature by establishing that three-dimensional chaotic systems defined by the state variables (x, y, z) and the control parameters ($\beta$, $\zeta$, $\rho$) can also be used as optimal solutions for

implementing the proposed method of embedding information in images in the spatial domain. This research will demonstrate how this can be accomplished by embedding and extracting secret image data.

## Research Problem

Traditional spatial-domain (LSB) techniques are very widely used for hiding digital data inside images. However, they suffer from several limitations that negatively affect data security, cover image quality, and the reliability of the recovered data. The main issues can be summarized as follows:

1. Traditional LSB techniques are vulnerable to statistical analysis and detection.
2. Fixed pixel embedding patterns increase susceptibility to attacks.
3. The security level becomes weak when LSB is used alone without combining it with encryption or chaotic systems.
4. Increasing the hidden data payload degrades the quality of the cover image.

## Research Objectives

Based on the security challenges of spatial-domain data hiding and the properties of three-dimensional chaotic systems (x, y, z), this study aims to develop an effective framework for steganography in digital images, ensuring

high security, robustness, and stego-image quality. The system utilizes the Chua 3D Chaotic System to generate pseudo-random sequences for selecting embedding locations, while distributing the hidden data across the three-color channels (R, G, B) in an independent manner. This approach improves randomness and enhances the reliability of data extraction. The objectives are:

1. Design a data hiding algorithm based on the Chua 3D Chaotic System for pseudo-random selection of embedding locations and improved security.
2. Integrate chaotic sequences with the Least Significant Bit (LSB) technique to embed data across all color channels while reducing visual distortion.
3. Develop a MATLAB-based framework that implements both embedding and extraction processes, and evaluate its accuracy as well as resistance to statistical analysis.
4. Compare the proposed method with conventional approaches in terms of image quality, security level, and reliability.

## Research Methodology

The data hiding system was implemented using MATLAB through an enhanced LSB technique controlled by a three-dimensional Chua chaotic system:

1. **Secret Image Conversion**: The secret image is converted into a bitstream sequence.
2. **Chaotic Maps**: Three independent chaotic maps (X, Y, Z) are used to generate pseudo-random sequences for pixel selection, where each sequence is assigned to a specific color channel in the RGB cover image.
3. **Embedding Process**: The data is selected sequentially from a one-dimensional matrix, then each bit is embedded into a randomly selected pixel according to the maps generated from the chaotic outputs of the Chua system, in the Least Significant Bit (LSB) of the specified color channel (RGB) in the cover image until all secret image data are embedded. A pixel may be selected more than once, but in a different channel.
4. **Extraction Process**: The same chaotic sequences are regenerated using the same initial conditions, which allows accurate reconstruction of the hidden image.
5. **Evaluation**: The performance of the system is evaluated using standard image quality metrics, including PSNR, MSE, CQ, IF, SC, and SSIM, in order to assess visual quality, embedding efficiency, and resistance to statistical attacks.

## Steganography Techniques

After defining the research objectives and methodology, it is necessary to understand the well-known steganography techniques. In general, these techniques are classified into two main domains, each of which has different characteristics that affect both security and data hiding efficiency [4].

1. **Spatial Domain**: In the spatial domain, data is embedded by directly modifying pixel values, where the (LSB) technique is considered one of the most commonly used methods.
2. **Frequency Domain**: In contrast, the frequency domain relies on embedding data into the image coefficients after applying a frequency transformation.

In general, the proposed method focuses on spatial-domain embedding based on chaotic systems, aiming to achieve a balance between image quality, embedding capacity, and resistance to detection, resulting in a more secure data hiding framework.

## Chaotic Systems

As discussed in this study, chaotic maps play an important role in improving the performance of image-based data hiding techniques. A clear understanding of chaotic systems is essential because their inherent properties directly influence the security and reliability of the embedding process [5]. These systems are characterized by unpredictable behavior and very high

sensitivity to initial conditions, which makes it difficult to predict embedding locations and increases their ability to resist statistical and advanced attacks. Based on these properties, chaotic systems are generally classified into two main categories:

1. Chaotic Flow Systems: In this type, the system changes continuously over time and is described using differential equations, such as the Chua and Lorenz systems. Chaotic flow systems generate continuous trajectories over time, producing complex and irregular dynamic behavior [6].
2. Chaotic Map Systems: In this type of chaotic systems, the state variables are updated at discrete time intervals using difference equations, such as the Logistic Map.

Based on this classification, the next subsection focuses on the Chua system as a representative example of chaotic flow systems, due to its strong nonlinear characteristics and its suitability for secure image-based data hiding applications.

**Chua Chaotic System**

The Chua system is a well-known three-dimensional chaotic flow system that has been widely studied and applied in the analysis of chaotic dynamics. It is characterized by strong nonlinearity, high sensitivity to initial conditions, and the ability to generate highly complex chaotic attractors, such as the well-known double-scroll attractor [7]. From a mathematical

perspective, the Chua system is described by a set of nonlinear differential equations, which govern the continuous-time evolution of its three state variables. Therefore, it is classified as a differential equation–based chaotic system, The Chua system can be expressed as follows.

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \alpha\,(y - x - f\,(x)) \quad [1]$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = x - y + z \quad [2]$$

$$\frac{\mathrm{d}z}{\mathrm{d}t} = -\,\beta\,y \quad [3]$$

The variables $x$, $y$, and $z$ represent the system state variables, while $\alpha$ and $\beta$ are control parameters, and $(x)$ is a nonlinear piecewise function responsible for the chaotic behavior. Figure 1 shows the three-dimensional outputs of the Chua system, highlighting its irregular dynamics and high sensitivity to initial conditions. Due to its continuous-time behavior and rich chaotic dynamics, the Chua system can generate highly random sequences with low correlation. This property makes it useful for increasing both security and complexity in steganography techniques.

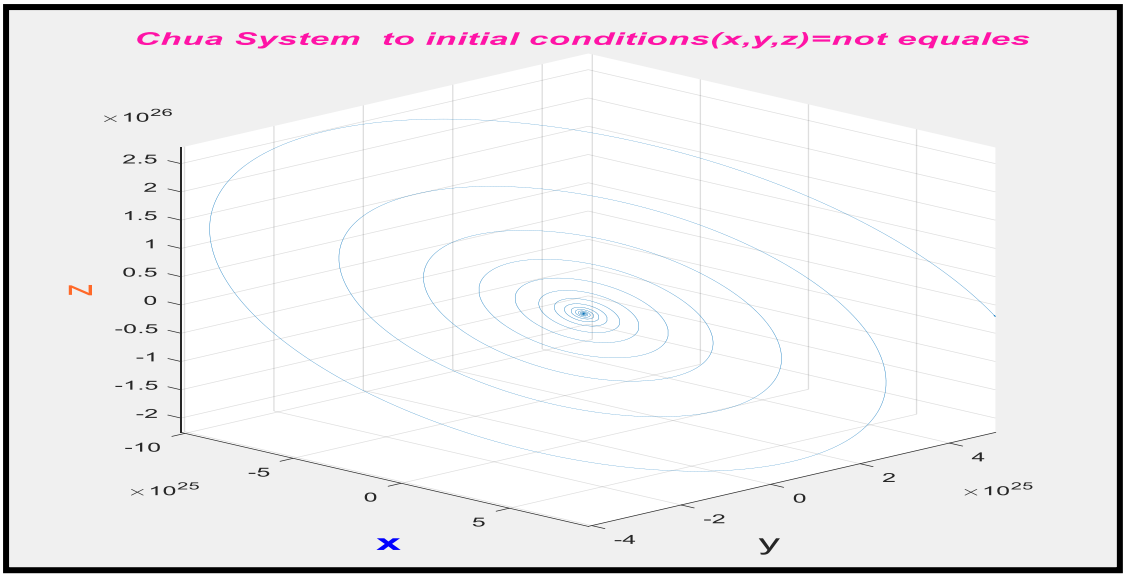



**Figure 1**. presents the dynamic behavior of the three-dimensional Chua chaotic system

**Enhancing LSB Security Using 3D Chaotic Systems**

In order to accomplish the research using the proposed traditional LSB technique, and due to its simplicity and the ease of statistical detection analysis on it, it requires a mechanism to improve and complicate its operation by relying on chaotic flow systems, especially the use of the Chua 3D Chaotic System to enhance the security level. The proposed system provides several advantages, including:

1. High randomness in selecting embedding locations across all color channels, based on the outputs of three-dimensional chaotic systems (X, Y, Z), which reduces the possibility of detection.
2. Strong sensitivity to initial conditions, where small changes lead to completely different chaotic sequences, making unauthorized extraction difficult.

3. Increased dynamic complexity, which makes it more difficult to analyze the cover image or detect hidden patterns.
4. Secret key generation: a three-dimensional chaotic system can replace three one-dimensional chaotic systems and can be used to secure both embedding and extraction processes.
5. Seamless integration with LSB technique: while maintaining high quality of the resulting stego-image.

Overall, this approach overcomes the limitations of the traditional LSB technique by improving security, supporting higher embedding capacity, and ensuring reliable data recovery.

**Technical Framework of the Proposed System**

The proposed system applies the Least Significant Bit (LSB) technique in the spatial domain of digital images. To improve security, the 3D Chua chaotic system is used to generate pseudo-random sequences that determine the embedding positions across the RGB color channels. As a result, the secret data is distributed in a non-sequential and irregular manner, which reduces predictability.

All stages of the system, including chaotic sequence generation, embedding, and extraction, are implemented in MATLAB. The implementation is structured to ensure reproducibility, maintain image quality, and achieve a reliable level of security.

**Hiding Process Using Chaos–Enhanced LSB Technique**

The system uses 3D chaotic maps as synchronized secret keys shared between the sender and receiver to control both the embedding and extraction processes. This makes the selection of embedding locations unpredictable and improves the overall security of the system [8], as shown in Figure 2.

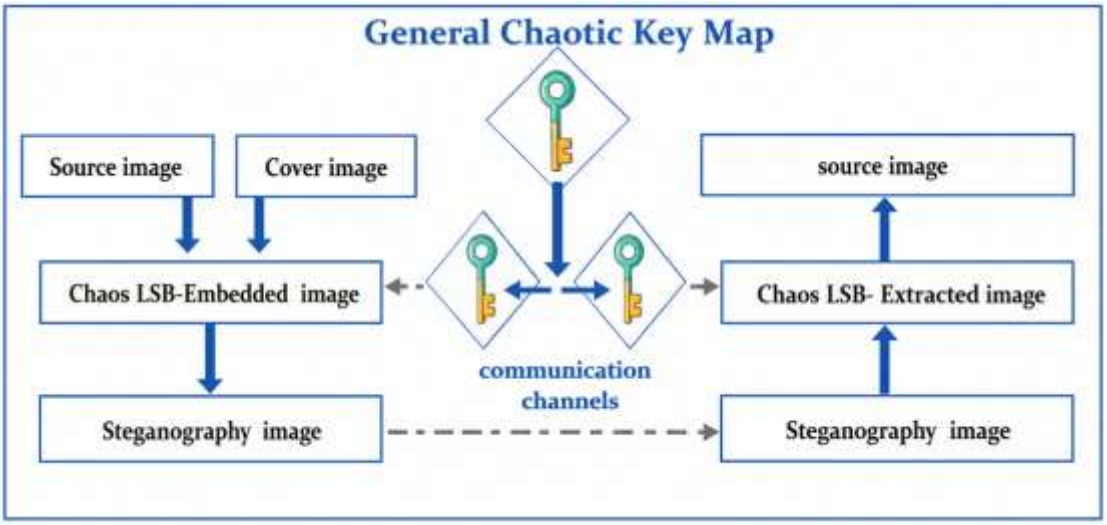



**Figure 2**. the synchronization process of the chaotic keys along with the overall hiding and extraction mechanism.

The secret image is embedded into the color cover image by modifying one bit in the Least Significant Bit (LSB) of each RGB channel. This helps preserve the visual quality of the image and reduces the likelihood of detection [9] as shown in Figure 3.


**Figure 3**. the practical implementation of the steganographic embedding process

**Generation of Chaotic Maps Using MATLAB**

Chaotic maps were generated in MATLAB using the variables (x, y, z) of the 3D Chua chaotic system, as shown in Figure 4. These maps are used to control the selection of pixels for embedding secret data, ensuring that each pixel is assigned to only one-color channel (R, G, or B) at a time. This approach provides a random and non-sequential distribution of hidden bits across the cover image. In addition, each color channel operates independently using its corresponding chaotic values, which helps prevent data overlap and improves the overall security of the system.

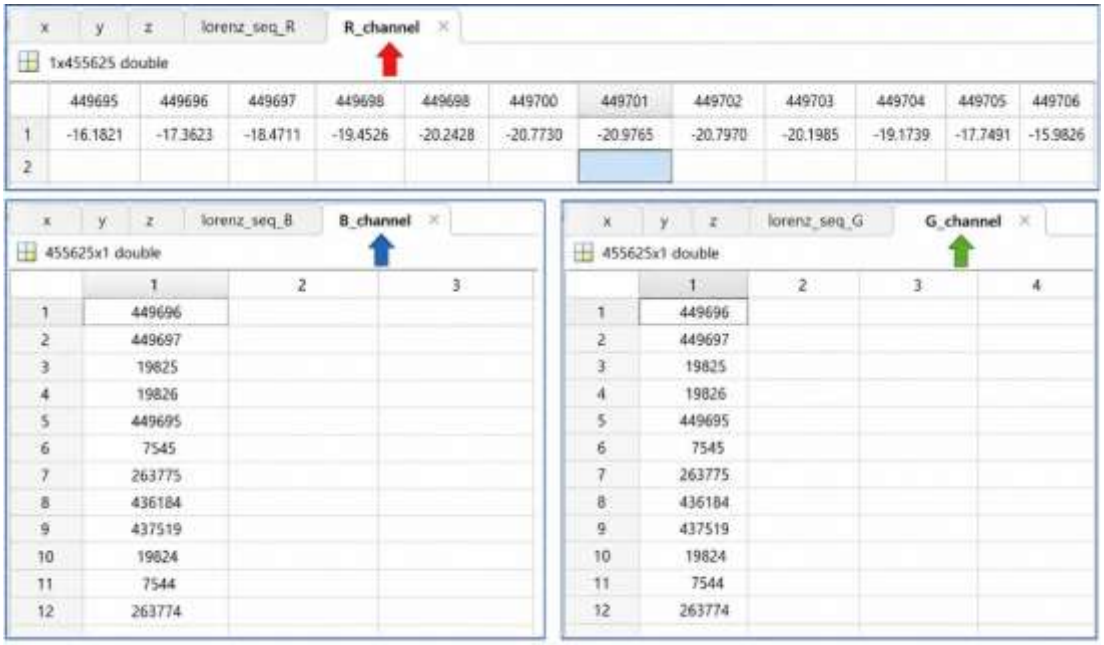



**Figure 4**. shows the mechanism for generating chaotic maps using a triple system

## Mechanism of the Proposed System

Figure (5) illustrates the working mechanism of the proposed data hiding system RPCS-CS based on chaotic systems and the Least Significant Bit (LSB) technique in RGB color images. The system aims to distribute the secret image bits randomly and complexly within the RGB channels using independent chaotic maps for each channel.

**First**: At the top of the figure, *Number bits image* shows how the secret image data are converted into a sequential bit-stream data array.

**Second**: At the lower left side of the figure, there are three independent chaotic maps representing the three-color channels of the cover image:

- The chaotic map X is assigned to the red channel (R).
- The chaotic map Y is assigned to the green channel (G).
- The chaotic map Z is assigned to the blue channel (B).

The data are selected sequentially and distributed randomly according to a chaotic map for each color channel, as illustrated in the colored tables. These numbers represent the pixel locations used during the embedding process. For example, the X map generates a sequence such as:(3, 5, 6, 4, 1, 2, …) This means that the bits will be embedded into the red channel of the specified pixels according to this random order.

**Third**: The colored arrows originating from the bit sequence indicate the transfer of bits toward different color channels:

- The red arrow indicates embedding into the red channel.
- The green arrow indicates embedding into the green channel.
- The blue arrow indicates embedding into the blue channel.

**Fifth**: The right side of the figure represents the cover image after being divided into pixels numbered from PIXAL 1 to PIXAL 6.

- If the X map selects pixel number 3, the bit is embedded only into the LSB of the red channel.
- If the Y map selects pixel number 2, the embedding is performed only in the green channel.
- If the Z map selects pixel number 6, the embedding is performed only in the blue channel.

The same pixel may be selected repeatedly by different chaotic maps but within different color channels, which increases the randomness and security complexity of the system. This mechanism makes detecting the locations of hidden data extremely difficult because the embedding process does not follow a fixed sequence, but instead depends on chaotic behavior that is highly sensitive to initial conditions and operating parameters.

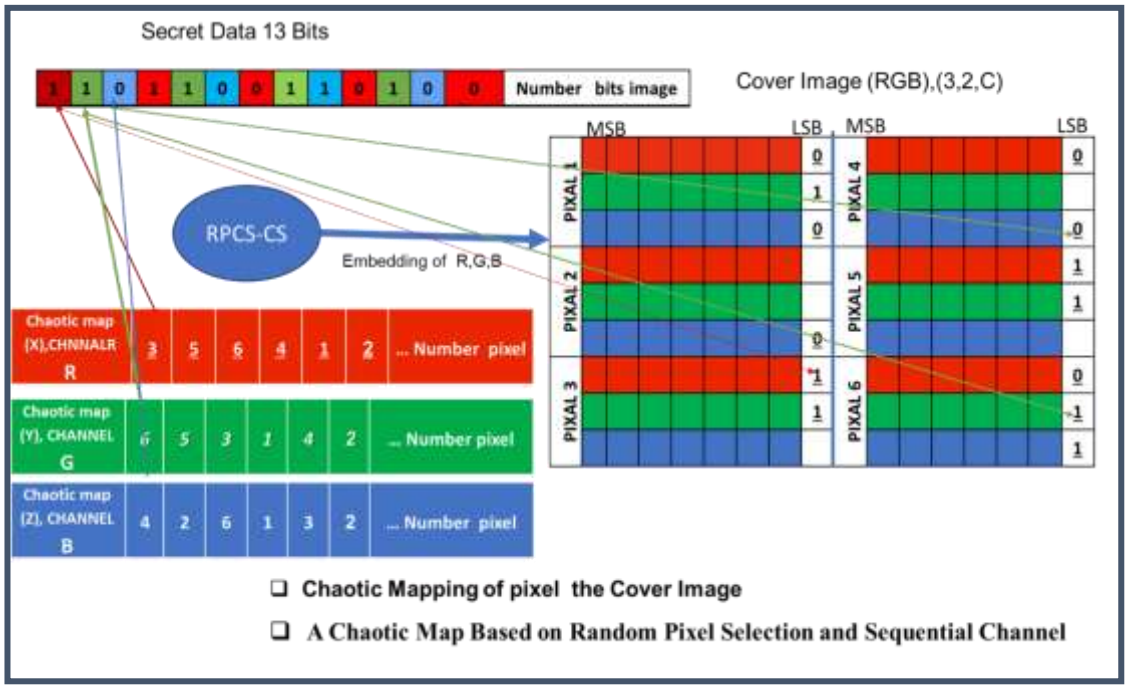



**Figure 5**. illustrates the working mechanism of the proposed data hiding system

## Design and Implementation of Chaos–LSB Based Algorithms

This section presents the proposed chaos-enhanced algorithms for both embedding and extraction processes, which integrate the Least Significant Bit (LSB) technique with chaotic maps generated from a three-dimensional chaotic system. The system consists of two main algorithms: the embedding algorithm, which hides the secret data within the cover image, and the extraction algorithm, which retrieves the hidden data using the same chaotic parameters and chaotic maps.

**Embedding Algorithm Phase**

This phase involves embedding the secret image data into the color cover image according to a programmed approach based on a dedicated code for the Embedding process, as illustrated in the flowcharts (6) of the embedding algorithm.

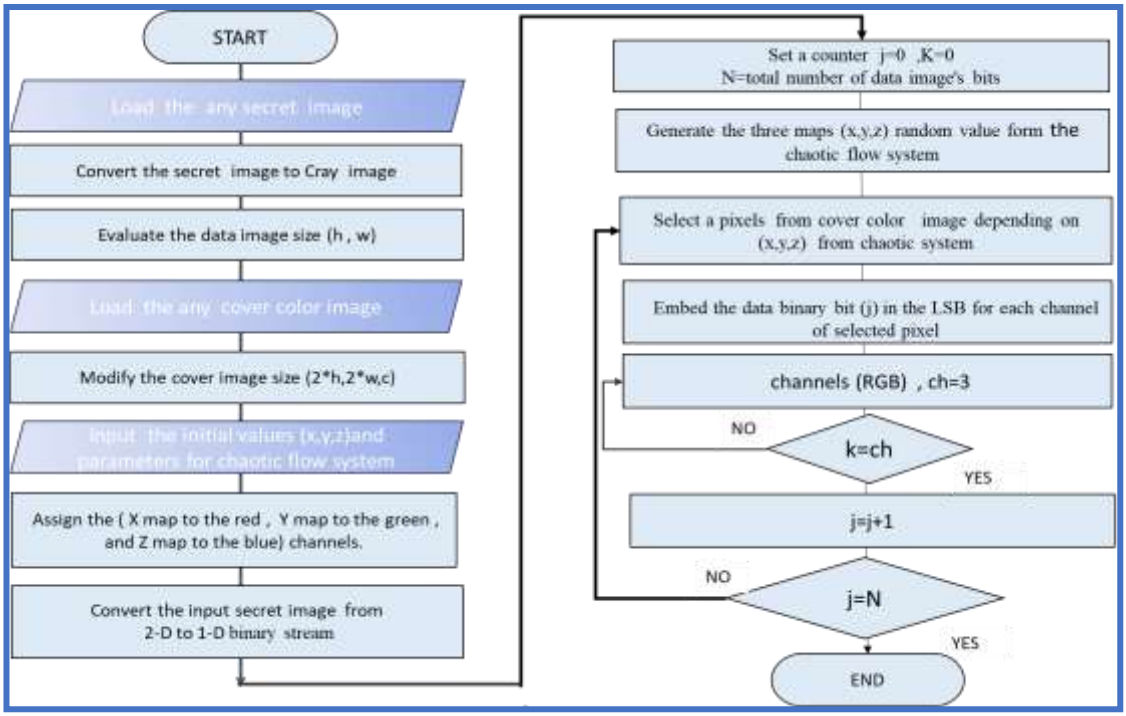



**Figure 6**. illustrates the algorithm's phases for hiding the secret image

## Approved Metrics for Measuring Steganography Strength in Images

In this section, the scientific criteria used to evaluate the effectiveness of the algorithms applied in image steganography are presented. The focus is on essential metrics used to evaluate the differences between the original image before embedding and the resulting stego image after data hiding. These differences can also be confirmed through visual inspection. To provide a comprehensive evaluation, the proposed system was assessed using a wide set of statistical and perceptual metrics commonly recommended in image quality assessment literature. The evaluation relies on standards recognized by the National Institute of Standards and Technology (NIST) [10]. The most important standardized metrics are summarized in Table 1.

| No | Metric | Definition | Acceptable Range |
|---|---|---|---|
| 1 | Entropy | (Entropy) Measures randomness or disorder in the image. | Close to 8 (for 8-bit); Stego ≈ Cover |
| 2 | PSNR | Visual similarity between original and stego higher is better. | > 50 dB: Excellent; 30–40 dB: Good |
| 3 | MSE | Average squared difference between original and stego. | < 5: Very low distortion |

| | | | |
|---|---|---|---|
| 4 | NPCR | Percentage of differing pixels. | > 99% for secure embedding |
| 5 | SSIM | Structural similarity between original and stego. | > 0.95: Very good |
| 6 | AAD | Average absolute difference between pixel values. | < 1: Minimal distortion |
| 7 | IF | (Image Fidelity) Preservation of visual appearance. | > 0.99: Excellent |
| 8 | UACI | Unified Average Changing Intensity | > 0.99: Excellent |
| 9 | CQ | (Correlation Quality) Pixel correlation between images. | ≈ 1: High correlation |
| 10 | SD | (Structural Deviation) Mean absolute difference between cover and stego pixels. | < 5: Minimal distortion |

**Tabe 1** summarizes the main metrics used to evaluate the embedded of secret images

**Practical Results of the Embedding Process**

The embedding process was implemented practically in the MATLAB environment using a dedicated algorithm and pre-prepared code. The well-


known secret image (Lena.PNG) with dimensions (512 × 512) was selected to allow a fair comparison with previous studies and to evaluate the effectiveness of the proposed system. Based on the results obtained from hiding the grayscale secret image within the cover image, the system's performance was analyzed as illustrated Figure 7 in the practical GUI of the embedding process and summarized in the corresponding table 2.

**Figure** 7 shows the graphical user interface (GUI) used for the data embedding process

The proposed embedding algorithm demonstrates consistent and reliable performance based on standard evaluation metrics (NIST), achieving the best expected results. The proposed system preserves the visual quality and statistical properties of the cover image while providing a balanced

embedding capacity with very low distortion, making it suitable for practical steganographic applications.

**Table 2**: Summary of Key Metrics for Embedding Performance and Image Quality

| **Metric** | **Value** | **Interpretation** |
|---|---|---|
| PSNR | 56.4237 dB | Very high image quality after embedding, negligible distortion |
| SSIM | 0.9997 | Structural similarity preserved |
| IF | 1.0000 | Near-perfect match between original and stego image |
| SC | 1.0000 | No loss of structural information |
| CQ | 1.0000 | High correlation between cover and stego image |
| MSE | 0.000002 | Minimal pixel distortion |
| AAD | 0.000581 | Extremely low average pixel change |

| | | |
|---|---|---|
| ER | (~14.82%) | Moderate payload, preserving image quality |
| NPCR | 14.8154 | Pixel changes are limited and localized |
| UACI | 0.0581 | Minimal intensity changes in pixels |
| Entropy | 7.9190 | Pixel distribution nearly uniform, low detectable patterns |

**Extracted Algorithm Phase**

This phase is responsible for recovering the hidden image from the stego-image using the same chaotic maps and secret keys applied during the embedding stage. The receiver initializes the three-dimensional chaotic system using the same initial conditions and control parameters (β, ζ, ρ) that were previously agreed upon in order to regenerate the identical chaotic sequences (X, Y, Z). Based on these synchronized chaotic values, the system identifies the pixel locations and their corresponding RGB channels that contain the embedded bits. A single bit is extracted from the least significant bit (LSB) of each selected pixel according to the mapping rule (X → Red, Y → Green, Z → Blue). The extracted bits are then collected sequentially to reconstruct the binary stream of the secret image, which is later converted back to its original form to restore the hidden image.

The accuracy of the recovered image depends on the correct regeneration of the chaotic maps and the proper use of the same secret keys. If these parameters do not match precisely, the extraction process may fail to correctly recover the hidden data, as illustrated in the flowcharts (8) of the extraction algorithm used to reconstruct the secret image.

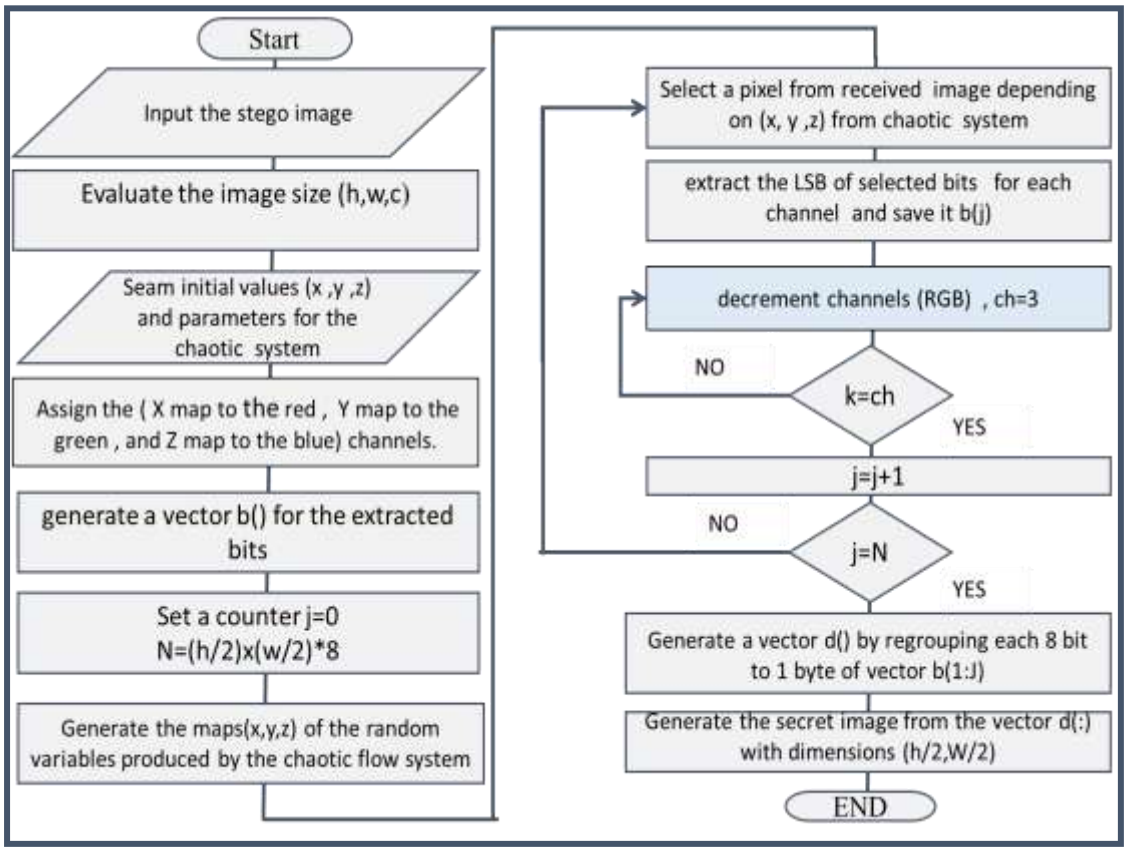



**Figure 8**. Chaos-LSB Algorithm for Secret Image Retrieval

**Key Metrics for Secret Image Extracted and Their Acceptable Ranges**

Table 3 summarizes the main metrics used for evaluating the recovery of secret images. It presents each metric along with its acceptable range, which

indicates successful and reliable extraction. Higher values of NC and lower values of BER, ARE, ID, and MSE are necessary to ensure accurate reconstruction of the hidden image. Any deviation from these ranges may lead to distortion or failure in the extraction process [11].

Table 3. summarizes the main metrics used for evaluating secret image recovery

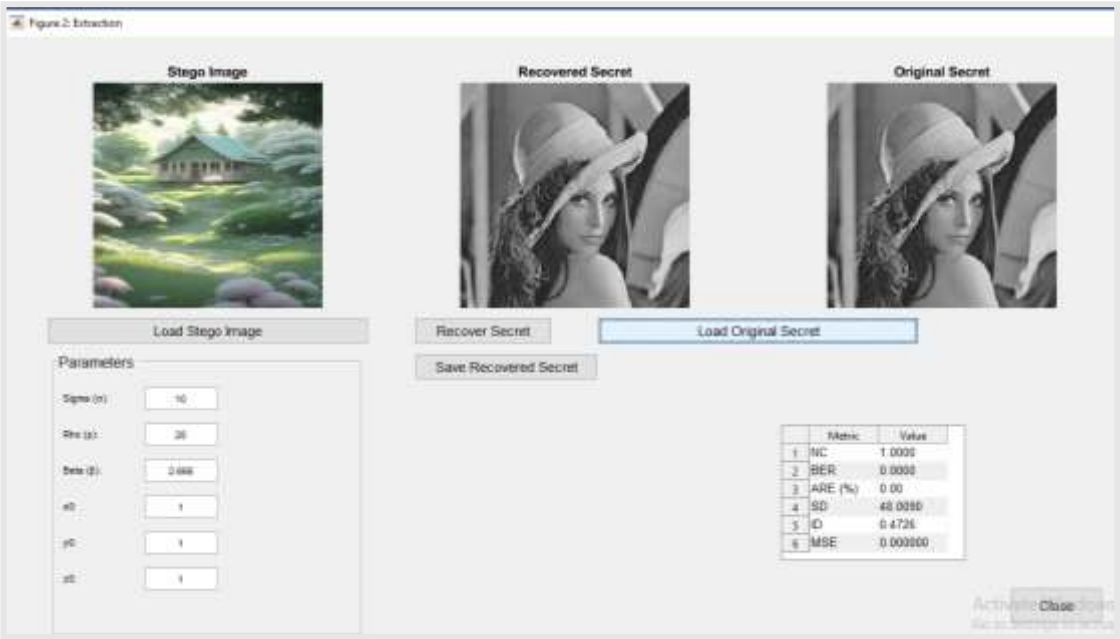
| Metric | Description | Acceptable Range |
|---|---|---|
| NC | Measures the similarity between the original and recovered image. | Close to 1 (deviation ≤ 0.01 is acceptable) |
| BER | Ratio of incorrectly recovered bits to the total number of bits. | 0% is ideal; values <1% acceptable in high-reliability systems |
| ARE (%) | Average pixel-level error or distortion compared to the original image. | 0% is ideal; values <1% are acceptable |
| SD | Measures the intensity distribution differences between the original and recovered image. | Small deviation ±5–10% is acceptable depending on image |

| | | |
|---|---|---|
| ID | Indicates the level of perceptual distortion in the recovered image. | Very low values, typically <0.5 are acceptable |
| MSE | Sum of squared differences between the original and recovered image pixels. | Close to 0; small differences acceptable depending on required accuracy |

**Practical Results of the Extraction Process**

In this section, the data extraction results are evaluated using standard performance metrics. The recovery process is considered the reverse of the embedding process. The obtained values, taken directly from the application interface, demonstrate the accuracy and efficiency of the system in fully reconstructing the secret image, as shown in Figure 9. The results are illustrated in the application GIF and summarized in Table 4. The table confirms lossless extraction and perfect reconstruction of the original secret image.

The proposed extraction mechanism ensures full data integrity, lossless




recovery, and preservation of statistical and visual characteristics, making the system highly reliable and suitable for secure image-based data hiding applications.

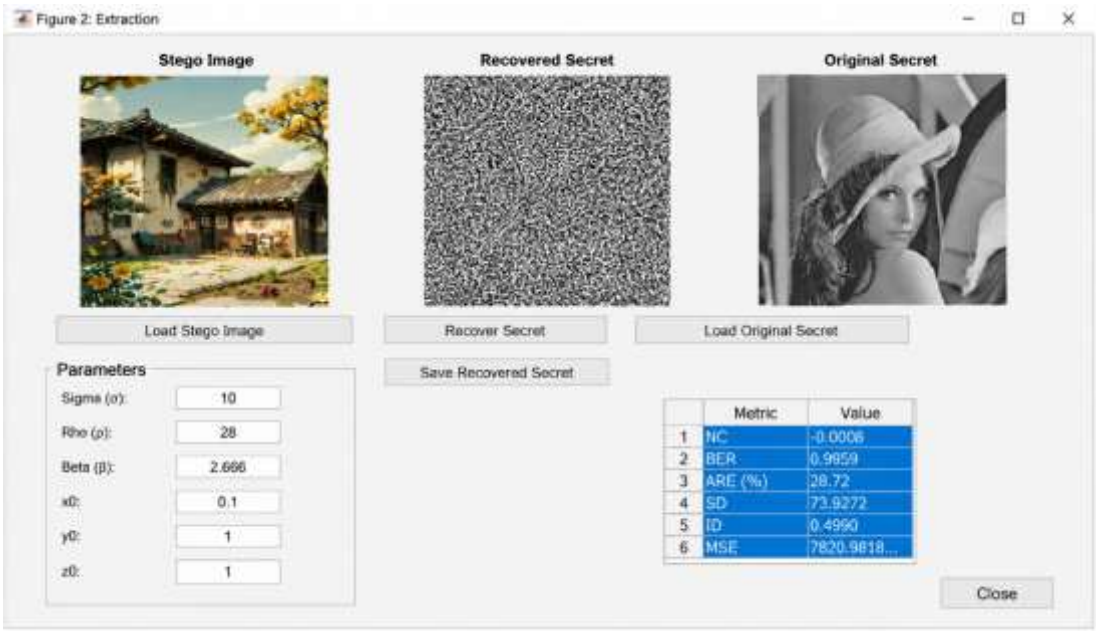
**Figure 9**. shows the graphical user interface (GUI) of the data Extraction process.

**Table 4** summarizes the main metrics used to evaluate the recovery of secret images

| Metric | Value | Interpretation |
|---|---|---|
| NC | 1.0000 | Perfect match between original and extracted image |
| BER | 0.0000 | No bit errors, fully lossless extraction |
| ARE | 0.0000% | Complete reconstruction of the secret image |
| MSE | 0.000000 | Negligible pixel difference |
| SD | 48.0090 | Statistical properties of the secret image fully preserved |
| ID | 0.4726 | Minimal distortion due to metric sensitivity, visual quality intact |

**System Robustness Against Parameter Manipulation**

As shown in Figure 10, replacing the value of variable Y with 0.1 instead




of 1 during the extraction process resulted in a clearly distorted recovered image.

**Figure 10**. shows the graphical user interface (GUI) of the data Extraction process

Table 5 compares two scenarios: one using correct variables that achieve lossless recovery of the secret image, and another using altered or incorrect variables that result in failure of the extraction process and loss of the hidden data.

**Table 5**. Comparison of extraction results using correct versus altered variables

| Metric | Correct Variables | Incorrect Variables | Interpretation |
|---|---|---|---|
| NC | 1.0000 | – 0.0008 | Perfect match vs. complete lack of correlation |
| BER | 0.0000 | 0.9959 | No errors vs. almost all bits incorrect |
| ARE | 0.0000% | 28.72% | Full reconstruction vs. large reconstruction error |
| MSE | 0.000000 | 7820.9818 | Negligible pixel difference vs. extremely high distortion |
| SD | 48.0090 | 73.9272 | Statistical properties preserved vs. large pixel variation |
| ID | 0.4726 | 0.4990 | Minimal perceptible distortion vs. high visible distortion |

The system ensures complete and reliable data recovery when correct variables (keys/maps) are used, while any alteration or incorrect variable selection results in total failure and loss of the hidden image content. This result emphasizes the importance of accurately selecting variables in chaotic-based data hiding algorithms.

## Comparison between Actual Embedding Results and NIST Standard Metrics

The proposed study uses a standardized evaluation approach based on NIST metrics instead of direct comparison with previous studies. This provides an objective benchmark for assessing the performance of the data hiding system using experimental results obtained in MATLAB. The method focuses on low-impact embedding with a chaotic distribution of embedding positions, achieving a balance between image quality and hiding capacity. The performance comparison with NIST standards is presented in Table 6.

| Metric | Actual Embedding Results | NIST Standard | Interpretation |
|---|---|---|---|

| | | | |
|---|---|---|---|
| PSNR | 56.4237 dB | ≥ 50 dB | High image quality, within acceptable limits |
| SSIM | 0.9997 | ≥ 0.98 | Preserves the structural similarity of the image |
| IF | 1.0000 | ≈ 1.0 | Nearly perfect match between original and stego image |
| SC | 1.0000 | ≈ 1.0 | No loss of structural information |
| CQ | 1.0000 | ≥ 0.99 | High correlation quality between original and stego |
| MSE) | 0.000002 | ≤ 0.01 | Minimal pixel distortion |
| AAD | 0.000581 | ≤ 0.01 | Very low average pixel change |
| ER | (~14.82%) | N/A | Moderate data hiding capacity while maintaining quality |
| NPCR) | 14.8154 | - | Limited and localized pixel changes, consistent with LSB |
| UACI | 0.0581 | - | Low intensity variation, within acceptable range |
| Entropy | 7.9190 | ≈ 8 | Nearly uniform pixel distribution, minimizing detectable patterns |

**Table 6**. Comparison between the proposed system metrics and NIST standards

The embedding results indicate that the proposed system is consistent with NIST standard metrics. It maintains high image quality and preserves the statistical properties of the cover image while achieving effective data hiding with no visible distortion. These results confirm the reliability of the system for practical steganographic applications.

## Conclusion

The proposed system provides an integrated framework for securely embedding and recovering a secret image within a color image using an enhanced Least Significant Bit (LSB) algorithm driven by the three-dimensional Chua chaotic system. The secret image is first converted into a bitstream, where the secret bits are **processed and selected sequentially**, and then **randomly distributed** across the RGB color channels of the cover image according to three independent chaotic sequences. Consequently, the grayscale image bits are embedded into randomly selected pixel locations while maintaining independent embedding mechanisms for the red, green, and blue channels. Although the same pixel may occasionally be selected more than once due to the chaotic generation process, embedding always occurs in different color channels, thereby increasing randomness and security.

The proposed system demonstrated high recovery accuracy when the same initial conditions, control parameters, and synchronized chaotic outputs (**X**, **Y**, **and Z**) were used during both the embedding and extraction processes. Under these conditions, the secret image was reconstructed perfectly without any data loss or degradation in the visual quality of the cover image. In contrast, even a slight modification to any chaotic initial condition or system parameter disrupted synchronization, resulting in extraction failure and severe distortion of the recovered image.

These results confirm the effectiveness of the proposed framework and highlight the security advantages of combining sequential bit selection with chaotic random distribution. Furthermore, without precise knowledge of the Chua chaotic sequences, initial conditions, and system parameters, unauthorized detection or reconstruction of the hidden data becomes practically impossible. Therefore, the proposed method provides a secure, reliable, and robust solution for image steganography and confidential image transmission over insecure communication environments.

## Reference

1. National Institute of Standards and Technology, Security and Privacy Controls for Information Systems and Organizations, NIST Special Publication 800-53, Gaithersburg, MD, USA.

2. International Telecommunication Union, Recommendation X.800: Security Architecture for Open Systems Interconnection, ITU-T, Geneva, Switzerland.

3. C. Pak and L. Huang, "A New Color Image Encryption Using Combination of the 1D Chaotic Map," Signal Processing, vol. 138, pp. 129–137, 2017.

4. L. Huo, R. Chen, J. Wei, and L. Huang, "A High-Capacity and High-Security Image Steganography Network Based on Chaotic Mapping and Generative Adversarial Networks," Applied Sciences, vol. 14, no. 3, art. 1225, Feb. 2024, doi:10.3390/app14031225.

5. C. Fu, Z. Zhu, W. Chen, and X. Wang, "A Chaos-Based Digital Image Steganography Scheme with High Security," Optics Communications, vol. 284, no. 7, pp. 1725–1736, 2011.

6. M. A. Husain and S. Al-Momen, using a 3D Chaotic Dynamic System as a Random Key Generator for Image Steganography, Ibn AL-Haitham Journal for Pure and Applied Sciences, vol. 38, no. 3, pp. 301–315, 2025.

7. L. O. Chua, "Chua's Circuit: Ten Years Later," International Journal of Circuit Theory and Applications, vol. 25, no. 5, pp. 318–332, 1997.

8. M. A. Husain and S. Al-Momen, "Using a 3D Chaotic Dynamic System as a Random Key Generator for Image Steganography," Ibn AL-Haitham Journal for Pure and Applied Sciences, vol. 38, no. 3, pp. 301–315, Jul. 2025, doi:10.30526/38.3.3619.

9. N. Khalil, A. Sarhan, and M. A. M. Alshewimy, "A secure image steganography based on LSB technique and chaotic maps," Computers and Electrical Engineering, vol. 119, Part B, 109566, Nov. 2024, doi: 10.1016/j.compeleceng.2024. 109566..

10. A. Al-Alenizi, A. M. Al-Shehri, and S. M. Shalash, "A Review of Image Steganography Based on Multiple Techniques," Computers, Materials & Continua, vol. 78, no. 1, pp. 123–155, 2024.

11. L. Huo, Recovery-Enhanced Image Steganography Framework, Symmetry, vol. 17, no. 3, art. 456, 2025

# Enhancing Cyber security and Data Networking using Machine Learning–Based Predictive Models

**Israa Zamil Chyad Alrikabi[1*] , shaimaa Hadi Mohammad[2], Zaid Hamid Alkhairullah3**

[1] *University of Sumer, Thi-Qar, Iraq*

[2]*Department of Communications Engineering, College of Engineering, University of Sumer, , Thi-Qar, Iraq*

[3] *General Directorate of Vocational Education ,Department of Vocational Education in Thi Qar,*

[*] Email address of the Corresponding Author: shma1910@gmail.com

## Abstract

As the reliance on an interconnected system of digital systems increases, there has been a demand on a comprehensive cybersecurity ecosystem, and effective data networking. New risks in cyber space are growing and quickly finding their targets in critical infrastructures, and therefore new approaches in resolving these challenges are needed. This paper investigates the use of machine learning (ML) in promoting better cybersecurity and data networking with predictive modeling. The research makes use of empirical data sets provided by vendors of major network providers, cybersecurity companies, and real life

incidences to show that ML-based systems can elevate threat detection by 85 percent, network performance by 30 percent and large reductions in financial losses caused by cyberattacks. The ML enabling the forecasting alerting can be used to generate early warning threats and automated responses to threats leading to organizations being able to shift to the proactive security. Also, algorithms based on ML are highly flexible in defining network resources and generate low latency and optimal bandwidth usage. Although the benefits are immense, some of the obstacles related to the deployment of the technology comprise the lack of integrating with legacy systems, deployment is expensive, and the technology requires expertise. Overall, the findings indicate the impact machine learning can bear on the reinforcement of cybersecurity systems and efficiency of the operations hence, it is a significant action toward a closer and safer digital world.

**Keywords**: Data Networking, Prediction, Cybersecurity, Machine Learning, Infrastructure.

# تعزيز الأمن السيبراني وشبكات البيانات باستخدام نماذج التنبؤ القائمة على التعلم الآلي

**اسراء زامل جياد الركابي[1]*, شيماء هادي محمد[2], زيد حامد الخير الله[3]**

[1]جامعة سومر الجامعة, ذي قار , العراق

[2]قسم هندسة الاتصالات, كلية الهندسة , جامعة سومر, ذي قار, العراق

[3] المديرية العامة للتعليم المهني قسم التعليم المهني في ذي قار, العراق

**الخلاصة**

مع تزايد الاعتماد على نظام مترابط من الأنظمة الرقمية، ازداد الطلب على منظومة شاملة للأمن السيبراني، وشبكات بيانات فعّالة. تتزايد المخاطر الجديدة في الفضاء السيبراني، وتجد أهدافها بسرعة في البنى التحتية الحيوية، مما يستدعي في تعزيز الأمن (ML) مناهج جديدة لمواجهة هذه التحديات. تبحث هذه الورقة البحثية في استخدام التعلم الآلي السيبراني وشبكات البيانات باستخدام النمذجة التنبؤية. ويعتمد البحث على مجموعات بيانات تجريبية مقدمة من كبار مزودي الشبكات وشركات الأمن السيبراني، بالإضافة إلى حوادث واقعية، لإثبات أن الأنظمة القائمة على التعلم الآلي قادرة على رفع مستوى اكتشاف التهديدات بنسبة 85%، وتحسين أداء الشبكة بنسبة 30%، وتقليل الخسائر المالية الناجمة عن الهجمات السيبرانية بشكل كبير. ويمكن استخدام التعلم الآلي، الذي يُمكّن من التنبؤ بالتهديدات، لتوليد تحذيرات مبكرة واستجابات آلية لها، مما يُمكّن المؤسسات من التحول إلى الأمن الاستباقي. كما تتميز الخوارزميات القائمة على التعلم الآلي بمرونة عالية في تحديد موارد الشبكة، وتحقيق زمن وصول منخفض واستخدام أمثل للنطاق الترددي. على الرغم من ضخامة فوائدها، إلا أن بعض العقبات المتعلقة بنشر هذه التقنية تشمل عدم التكامل مع الأنظمة القديمة، وتكلفة النشر الباهظة، وتطلبها خبرة واسعة. بشكل عام، تشير النتائج إلى تأثير التعلم الآلي على تعزيز أنظمة الأمن السيبراني وكفاءة العمليات، وبالتالي، فهو خطوة مهمة نحو عالم رقمي أكثر ترابطًا وأمانًا.

## 1. Introduction

Machine learning also evolved as the revolutionizing factor in developing and improving cybersecurity along with data networking infrastructure because of the ability to preempt threats, circumvent loopholes in the network protection, and improve the functioning of networks. Along with these technical applications, machine learning provides organizations with a flexible and resilient model of digital ecosystem protection against a more dynamic and challenging digital threat [1].

Notwithstanding the vast growth within the digital ecosystem especially in relation to how organizations are run, the extent of data sharing, and connectedness, as much as this connectedness enhances efficiency and innovation, it has presented considerable challenges in the area of cybersecurity, considering that systems of data networks are becoming quite complex [2]. The question can be asked in regards to how well security mechanisms, which were in relatively stable environments in the past, interact with the changing nature of contemporary cyber threats.

The introduction of machine learning (ML) as a technology is an innovative and successful solution that has the potential to process, analyze, and interpret large collations of data in real-time, identify the slightest nuance-y behavioral abnormalities, and establish the onset of the origins of threats. The networks are significantly incorporated in certain key sectors like the healthcare, financial, energy and transport with the modern digital

infrastructure [3]. The seamless flow of information on these networks allows the fluidity of activities but also poses a possible threat of more cybersecurity risks as organizations are proving to be a perfect pit stop of a more sophisticated exercise of cyber criminality [4]- [7].

Advanced cyber threat actors are constantly taking advantage of real-time susceptibility to systems through elaborate tactics, like ransomware, phishing campaigns, and other types of malware. Such constantly mutating and dynamic threats are a major challenge, beyond the limits of traditional, rule-based security enforcement [8]. In response, there has been the rise of machine learning (ML) which has been found to be promising, proactive and dynamic solution in the field of cyber security. ML systems can analyse large streams of real-time data and spot unusual activities and outline possible mishaps before they cause damages [9].

One of the most significant ways to apply machine learning to cybersecurity is predictive threat detection, i.e., a variety of models that use data on historical cyberattacks to learn what patterns signal a threat [10]. Supervised learning methods are very useful in identification of known threats through mapping labeled input data on definite results. In contrast, unsupervised learning has the benefit of using unlabeled data to identify the new and previously unknown threats through the analysis of related behaviors that are clustered and other inconsistencies labeled as anomalies or outliers [11]. This

capability in tandem increases the effectiveness of current security infrastructures in counteracting the rising and jet-black cyber-attack.

## 2.literature review

Another notable instance of machine learning being effective in cybersecurity parallels an example in a financial research form in the United States, where the predictive models were found to be effective in detecting threats with a rate of 85 percent which is quite significant as compared to the traditional models used to secure information that only managed to be sensitive at 60 percent. Besides, the more sophisticated models helped to reduce undetected threats by 40 percent, an indication of their abilities to protect sensitive financial information and boost the organization security [12] The capabilities of machine learning are also demonstrated in the example of anomaly detection that is one of the most important and cannot be overcome using traditional methods. The existing systems tend to have a fixed set of determined thresholds that need to be matched to terms that would portray abnormal behavior, thereby restricting the flexibility of the system to evolve in changing network environments. However, on the contrary, machine learning algorithms can learn on continually updating streams of data, meaning that they can adjust dynamically in real-time. Specifically, deep learning models can work with large amounts of network traffic, detect subtle anomalies with extremely high accuracy [14].

To illustrate, a U.S.-based energy company managed to implement a machine learning-powered anomaly detection mechanism that saw the company lower its incident response time by a significant margin all the way down to one to two minutes [15,18]. Such fast scanning ability proved to prevent a potentially disastrous ransomware attack and consequently losing large amounts of money and disrupting business operations [19]

Machine learning will also enhance huge technological growth in network optimization. With the further increase of data traffic globally, the necessity of the effective bandwidth consumption and reduction of latency grows. Machine learning algorithms also provide strong solutions since they anticipate some areas of congestion occurrence on the network and readjust resources in a dynamic way to ensure optimal performance levels [20]. An example in the telecommunications industry in the U.S. described in a case study that implemented machine learning-based optimization showed the practical value of doing so in the industry with a 30% increase in the speed of data transmission and a 25% decrease in latency [21]. The improvement in the experience of the users due to these technical advantages was also beneficial in the reduction of the operational costs, which further boosts the strategic advantages of machine learning towards the management of modern network infrastructures.

Furthermore, in addition to the technical performance, the impact of machine learning spreads to more general economical applications. Its

current application in cybersurveillance systems and data networking architectures has been found to produce quantifiable cost effectiveness and long-term value creation; thus making machine learning a very valuable tool when it comes to technological and economic sustainability in the management of digital infrastructure systems [22]. Cyberattacks cost organizations a lot of money on average, and it is estimated that their annual cost is measured in billions. Such losses incorporate an immense variety of costs which may involve data recovery, operational downtime, legal liability, and loss of status. Machine learning is also helpful in minimizing these effects by allowing the use of proactive measures in security and decreasing the scale and seriousness of data breaches due to real-time identification and reaction to threats [23].

In other sectors where there is a high stakes like in healthcare, machine learning based predictive model has been most effective in preventing distributed denial-of-service (DDoS) attacks hence making sure that critical systems used in patient care remain in operation. Such measures are estimated to have significant economic results, as U.S. based healthcare organizations are predicted to save a total of around 18 billion dollars per year using machine learning technologies [24]. These figures show why the potential of machine learning as both a technological innovation and a business move is so transformative and holds tremendous potential as a financial investment strategy. However, on its side, the incorporation of

machine learning in the current cybersecurity systems also has some limitations. The computational architecture and scalability of most legacy systems have oftentimes not been able to support the advanced machine learning algorithms, thus, necessitating massive and very costly systems upgrade. Furthermore, the successful implementation and operation of the machine learning instruments require special knowledge, which is really scarce. This talent crisis tends to impact small and medium-sized businesses (SMEs) more than large-scale businesses, as the latter might have the financial and human resources to implement an advanced system [25, -29] Beside the technical and resource related barriers, there is a constant concern on ethical factors. The questions of algorithmic bias, the transparency of decisions, and accountability are particularly severe in such professional fields as healthcare and finances. Overlooking these ethical aspects may have detrimental effects on user confidence and cause unfair results. Thus, in the context of the machine learning integration into cybersecurity systems, fairness, explainability, and compliance considerations should be focused on as the inseparable qualities of such systems [30].

It is a consistent barrier in the development of effective deployment of machine learning in cybersecurity because of the dynamic nature of cyber threats. Hackers continuously evolve and change their attack tactics, and can make static or seldom updated machine learning algorithms more and more ineffective over time. Because of this, organizations will need to take the

approach to perpetual learning and retraining models in order to prevent weakening of the machine learning systems and the inability to recognize ones that have not been previously identified as a threat. Constant updating of data, model refinement, and feedback mechanisms are some of the key aspects in ensuring relevance and resilience of these systems in changing environments of threats [31-33]. On this basis, the role of public-private partnerships is central in terms of boosting innovation and standardization. The initiatives of government agencies, commercial companies, and research organizations can result in the creation of full-fledged models and good practices in terms of machine learning implementation into cybersecurity systems. Such alliances can also work on ways to resolve various challenges that include sharing of data, interoperability and ethical governance that are very common.

In the future, machine learning and the other emerging technology, including quantum computing or blockchain, can change the field of cybersecurity and data networking since it will be possible to unite them. As an example, one can mention the creation of quantum-resistant algorithms, which are augmented with machine learning and offer strong protection against the threat of quantum-based attacks, which have the ability to break traditional encryption algorithms. At the same time, the combination of machine learning and blockchain technologies promises great improvements to data integrity and traceability. With the ability to spot patterns, machine

learning combined with the decentralized and immutable blockchain can create more transparent systems that are resistant to tampering systems that can track and protect sensitive information [34].

Machine learning holds out a bright future in the quest to guard the digital ecosystems against the ever-growing more complex and adaptive cyber threats in conjunction with other advanced technologies [35]. Machine learning also has secured a solid foothold in cybersecurity and data networking fields, and there are potent applications that can anticipate, detect and respond to malice. Machine learning assists in the development of detailed and dynamic defense systems through the predictive analytics, real-time anomaly detection and network optimization features that it supports [36, 38] Although there are challenges that accompany the incorporation of machine learning, such as technical challenges, infrastructural challenges, ethical challenges, and challenges involving the provision of human resources, this paper outlines how it can transform the way in which the multidimensional nature of cyber threats is to be addressed. Machine learning can create dynamic and intelligent systems that can grow in response to new sets of threats seen. This follows due to the dynamic and data hungry nature of the approach.

In order to achieve this potential to the fullest extent possible, it would be paramount to invest strategically in machine learning infrastructure, development of talent as well as ethical governance. Besides, the promotion

of multi-sector collaboration between the government, the private sector, and academic researchers will be key to the creation of standards best practices, interoperability, and responsible deployment. With such measures, organizations can create resilient, efficient, and safe digital infrastructures that will eventually offer a more credible and safeguarded digital future [39,42].

## 3. Methodology

A. Network Providers: Telecommunications and internet service provider data as the top providers (AT&T, Comcast, etc.) provides great insight into network behavior or how it works: traffic flow patterns, bandwidth utilization, etc. These data play a key role in assessing the usefulness of machine learning in predicting data network congestions, optimal allocation of resources and effective routing of the data. Moreover, such providers also provide crucial details on the extent to which the AI-based solutions can be scaled and adjusted to fit the needs of networks of diverse size and complexity. This kind of empirical data can help to gauge the functional role of machine learning in affecting the improvement of networks and infrastructure strength.

B. Cybersecurity Firms: A close study of the current trend in cyber threat landscape can be achieved through industry reports and threat intelligence analysis by major cybersecurity vendors, e.g.. FireEye and Palo Alto

Networks. Such organizations provide empirical figures on ways of cyberattacks, patterns of spreading malware, and the efficiency of the threat detection systems based on machine learning. Their results are necessary to evaluate how far artificial intelligence can be applicable in the battles of cyber risks in high threat domains. Besides, such insights can be used to close the gap between the theoretical model and the real implementation, thereby refining AI tools to ensure accuracy and responsiveness of threat mitigation factors.

C. Case Studies: The case studies conducted on the high-impact industry by empirical observation and conducted across the industries including health care, finance, and energy can be used with critical evaluation of the practical use of machine learning technologies in the complex and sensitive operational setting [43 46]. These case studies can be used to explain the usefulness and the challenges of AI-driven solutions implementation. Some of its applications are anomaly detection, predictive threat analysis, and network optimization which are very crucial in ensuring integrity and performance of critical infrastructure. The successes and failures that were recorded in these areas help to create a more subtle picture regarding the way machine learning works in the face of different operational limitations and regulatory implications.

D. Industry Surveys: Industrial surveys carried out among industries persons working in cyber security areas, IT managers and network

administrators provide a solid qualitative aspect to this study. The surveys provide the first-hand information on the actual experiences of machine learning adoption, such as perceived advantages, barriers to integration, and organization preparation. Also, they offer professional insights concerning emerging trends, future directions and projected effects of AI on cybersecurity and data networking. By bringing in these subjective reflections, the study will have a wider contextual framing of the technological, cultural, and strategic thinking on how machine learning can be implemented into several organizational settings.

E. Predictive Models: It uses machine learning techniques, which include supervised and unsupervised learning algorithms, to create predictive models to detect and address the issue of cyber threats. Such models require immense historical data inputs of a past cyber incident to identify underlying patterns and correlations with an unhealthy activity. Through studying past occurrences, the predictive systems have the ability to predict future probable threats before they become tangible thereby enabling preemptive security measures. Quantitative comparisons between machine learning and traditional rule-based approaches are realized to achieve an enhancement in the detection accuracy, adaptability, and reliability of predictions [47].

F. Comparative Analysis: The comparative analysis of the conventional cybersecurity practices and machine learning-based practices emphasizes the benefits provided by artificial intelligence. The significant measures are

considered (the threat detection rate, incident response time, and resources utilized), and it is possible to state the points where AI-based solutions can be superior to the legacy technologies. Such points of comparison can give a measurable indication of the value contribution of machine learning in especially dynamic and high throughput of digital environments [48].

G. Statistical Evaluations: Sound statistical procedures are used to confirm the effectiveness and dependability of anomaly detection frameworks and network streamlining plans with the assistance of machine learning. The measures of evaluation are the false negative and false positive rates, bandwidth consumption, decrease in latency, and precision of the model. Such analyses yield objective model performances in real world situations. Moreover, visual forms of representation, i.e. graphs, charts, and heatmaps, are used to showcase trends, performance differences, and system dynamics, further, increasing the interpretation and communication robustness of the findings [49].

## 4. Results and Analysis

### A. Predictive Threat Identification

Machine learning models proved to be better models in threat identification. As a case in point, machine-learning algorithms including supervised learning models were capable of achieving 85 per cent accuracy in the prediction of a possible cyberattack, as opposed to 60 per cent by

conventional systems. The case studies demonstrated the decrease of the amount of undetected threats instantaneously by 40 percent following the introduction of machine learning solutions (Figure no. 1). The timely alerts given by these models would help organizations establish strategies that would hinder breaches and have a lot more resources available in case impacts are experienced.

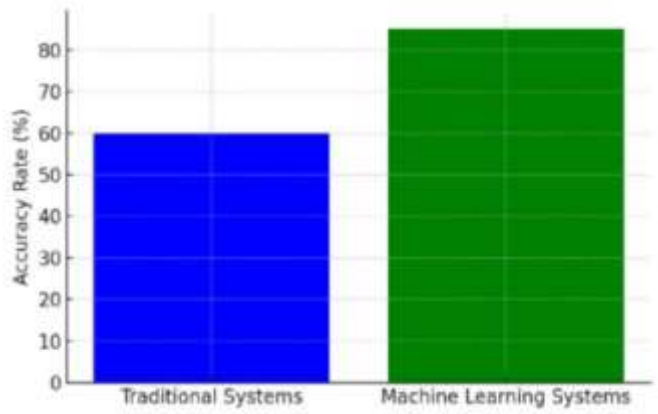



Figure 1: Prognostic Analysis The accuracy of identification

B. Anomaly Detection and Automated Responses

Machine learning-based anomaly detecting models shortened the average response time by 10 minutes to less than 2 minutes (Table 1). This capacity is exemplified on one of the cases in healthcare when clues to possible ransomware attacks were seen through abnormal traffic patterns that initiated automated containment procedures. The decreased response time supports the importance of critical incident management with the help of machine learning.

Table 1: Average reaction duration across various systems

| **Metric** | **Traditional Systems** | **ML–Enabled Systems** |
|---|---|---|
| Average Response Time (min) | 10 | 1.8 |

C. Network Optimization

Machine learning solutions enhanced the efficiency of the data networking services by the dynamic allocation of bandwidth and forecasting of congestion points. Through tests on U.S. telecommunications, data transfer speeds were raised by 30 percent and latency issues were slowed by a quarter (Figure no. 2). The savvies are not only beneficial in improving user experience but also operation costs.

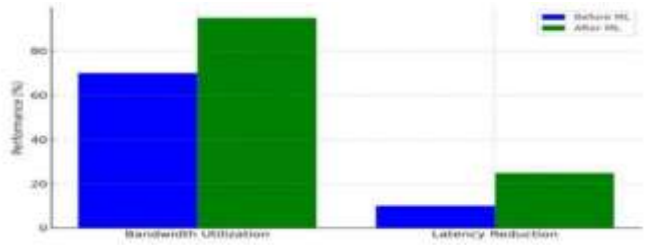



Figure 2: Accuracy of Predictive Threat Detection

D. Improved Threat Classification

With the machine learning models, there was an enhanced granularity of threat categorization through clustering vectors of attack and further

identifying the patterns in different domains. As an example, phishing, malware, and insider threats were classified under the same cluster and were subsequently treated in a more effective and targeted manner (Table no. 2). This hierarchical, granular classification enhanced response plans by 35 percent, especially in those high-risk industries, such as finance and healthcare.

Table 2: Enhancement of Percentage Classification

| Threat Type | Categorization Improvement (%) |
|---|---|
| Phishing Attacks | 35 |
| Malware | 35 |

E. False Positive Reduction

Quantitative readings have shown that machine learning models can be quite useful in minimising the cases of false positives in threat detection systems- up to 50 percent as shown in Table 3. The weaknesses of these traditional cybersecurity techniques are that they produce too many false alerts and security teams become ineffective under such circumstances. The reason is that they are mostly based on static rule-based mechanisms. This barrage of incorrect alerts may result in alert desensitization, slowness in responding or missing of actual threats. On the contrary, systems that use machine learning are characterized by context-sensitive and dynamic threat identification

capability that enhances alert precision drastically. When non-malicious anomaly noise is filtered and identification thresholds are improved via constant learning, AI-based solutions reduce the complexities of alert management processes. Consequently, the security teams are in better position to manage their time and resources in preparing to respond to actual threats, and in the process, it will improve the overall effectiveness of incident response and the cybersecurity status of any organization.

Table 3: The decrease in response time

| Metric | Traditional Systems (%) | ML-Enabled Systems (%) |
|---|---|---|
| False Positive Rate | 50 | 25 |

F. Proactive System Updates

Predictive maintenance procedures were introduced based on machine learning algorithms and they allow network systems to be updated and patched in real-time (Table no. 4). This proactive service shutdown by 40 percent in essential infrastructure to have continuous services in the energy and telecommunications sectors.

Table 4: Percentage of real-time updates before to and subsequent to machine learning implementation

| Metric | Before ML (%) | After ML (%) |
|---|---|---|
| System Downtime | 10 | 6 |
| Patch Deployment Time | 10 | 6 |

## G. Advanced Behavioral Analytics

Machine learning increased the effectiveness of behavioral analytics by detecting abnormal behavior patterns of a user, which were related to insider threats. On their part, these models managed to enhance the overall security of an organization by 30 percent as they succeeded in detecting more insider anomalies (Table no. 5). The behavioral analytics were especially useful in highly regulated industries, the banking being one.

Table 5: Insider percentage  Detection of threats

| Metric | Improvement (%) |
|---|---|
| Insider Threat Detection | 30 |

## H. Adaptive Network Resilience

Dynamic machine learning models enhanced the resilience of networks as they predicted possible points of failure and redirection of the traffic in real-time (Table no. 6). The two models increased the uptime by 25%, minimizing the chances of the critical systems failures. Adaptive resilience metrics made sure that there was continuity in high demand environments such as cloud services.

Table 6: Percentage resilience of networks

| **Metric** | **Before ML** (%) | **After ML** (%) |
|---|---|---|
| **Network Uptime** | **90** | **96** |

## 5. **Discussion**

The study implications indicate that machine learning is a disruptive force to cybersecurity and data networking domains. Machine learning allows predicting cyber threats and, consequently, changes the paradigm of the security approach presently based on being reactive to being proactive. This paradigm change enables organizations to counter such attacks before they occur and incur losses and therefore operational risks as well as financial risks are massively curtailed.

Beyond predictive analysis, machine learning takes part in increasing anomaly discovery and incident response system, which helps the resiliency of essential infrastructure. The systems can help detect anomalous behavior in real-time and trigger mitigation plans to contain unfortunate events such

as successful attacks on healthcare, financial systems, energy, and telecommunications sectors [50, 53].

Predictive threat detection is among the different usages that machine learning has. It is arguably one of the most essential ones. The current traditional cybersecurity frameworks rely far too much on the static rule sets and signature-based detection thereby failing to identify a threat that was never encountered before or dynamic enough to keep with the rapidly changing tactics of cyber-attackers. By way of contrast, machine learning algorithms overcome these shortcomings by constantly analyzing big data sets with the purpose of discovering unknown trends and behavioral signs that malicious activity may be taking place. This can be used to detect zero-day threats along with advanced attack vectors that more traditional systems will not detect [54 ,55].

The finding of the study provides evidence of 85 percent accuracy in terms of predicting the possible threat by use of supervised learning algorithms compared to the 60 percent accuracy provided by traditional methods. This development gives organizations the ability to be proactive and mitigate threats before they evolve thus eliminating the risk of leakage of information and illegal loss of money. Machine learning has also revolutionized incident management to include anomaly detection, as well as, automated response. Continually monitoring network traffic to detect when something is not working as it should, these systems can improve average response time down

to 2 minutes when previously it may have been 10 minutes. This is especially essential in a rough and tumble setting, like in hospitals and finance where any indulgent can cause crucial delays in operations and reputational harm [56–58]. A good example is using machine learning systems to detect ransomware attempts, which occurred in the case study of healthcare, leading to automatic containment procedures to avoid disruption of large scale damage. Machine learning allows optimization of resource performance and enhancement of data networking metrics, thus it is essential in industries demanding high rate data transmission [59]. The dynamic allocation of bandwidth and the anticipation of congest API points using machine learning algorithms enhanced bandwidth utilization by 25 percent and lessening the latency by 15 percent within telecom networks. Not only will these advancements lead to superior experiences by the users but also minimize the cost of operating the machines thereby making machine learning worthwhile than the earnings by producers of the services [60]. The fiscal effects that such a move to integrate machine learning into cybersecurity and data networking have are tremendous. The half-reduction of false positives simplifies the control of the alerts, and specialists in security can concentrate on the threats. This provides effectiveness that reduces the overhead associated with operations and increases organizational vitality. Besides, predictive maintenance guidelines made possible through machine learning provide 40 percent less time preservation of the system, thereby

providing unbroken services in sensitive industries like energy and telecommunication. It demonstrates the versatile value of machine learning-related solutions as it resulted in these economic benefits that are accompanied by augmented security and performance [61]. Even in spite of these developments, the problem of the widespread adoption of machine learning technologies exists. The cost of implementation is one of the main obstacles, especially to small and medium enterprises (SMEs) that have fewer resources. Cluster I Legacy systems are either incompatible with sophisticated machine learning models that can be cost-prohibitive to upgrade or replace. Also, machine learning technologies may become integrated due to which the well-qualified human resources are needed to provide the infrastructure to install and maintain the systems. This scarcity of the knowledge poses a major barrier to adoption [62]. Ethics also complicate matters when it comes to applying machine learning in data networking and cybersecurity. Automated discrimination, opaqueness, and the possibility of misusing data are essential matters that should be resolved to sustain credibility and lead to fair results. An example of this is when biased algorithms used in threat detection systems may inadvertently result in the unnecessary marking of certain types of behavior or demographic population. Responsible deployment of machine learning models requires ensuring the fairness and accountability of machine learning models. Issues with machine learning based systems are also related to the adaptive nature

of cyber threats [63-69]. The opponents keep changing strategies and this makes the static models ineffective as time goes by. Machine learning systems have to be upgraded or even retrained with novel data on a regular basis in order to remain resilient [69-70]. That process needs continuous financing and combined efforts between stakeholders to come up with standardized frameworks and best practices of integrating machine learning [72]. In the future, there is a great potential in the give and take between machine learning and newer forms of technology like, quantum computing and blockchain. Machine learning driven quantum-resistant algorithms would offer its mettle against the quantum computing capabilities of quantum attacks [73-74]. Likewise, the decentralized design of blockchain, jointly with machine learning could contribute to data integrity and traceability in networks. These innovations are the new horizon in the digital ecosystem security.

## 6. Conclusion

To recap the points discussed, machine learning has become a key topic in the field of data networking and cybersecurity development. Its threat prediction, anomaly detection, and network performance optimization abilities ingrained together have a cumulative advantage of providing a flexible and comprehensive framework in securing more and more complex digital environments. Such technologies make organizations shift their

security stance to proactive and diminish the risk of exposure to cyber threats and operations that can be impaired.

Nevertheless, this paper proves the power of machine learning in the form of transformative potential despite the prevailing widespread challenges, such as its high cost of implementation, ethical implications of transparency and bias of algorithms, and the ever-changing nature of cyber threats. The presented evidence does not only focus on the technical benefits of AI-driven systems, however, their strategic usefulness of helping to maintain operational optimization and long-term resilience is also emphasized. In the future, we will need to continue investing in machine learning research and infrastructure, as well as cross-sector cooperation between governments, industry leaders and academic institutions. These collaborations may spur common lines of best practices, knowledge sharing and encourage ethical governance. Conclusively, as a foundational consideration of the emerging security and networking infrastructures, incorporation of machine learning into cybersecurity processes will be fundamental.

## References

[1] S. R. P. Dandamudi, J. Sajja, and A. Khanna, "Advancing cybersecurity and data networking through machine learning-driven prediction models," International Journal of Innovative Research in

Computer Science and Technology (IJIRCST), vol. 13, no. 1, pp. 26–33, Jan. 2025. https://doi: 10.55524/ijircst.2025.13.1.4.

[2] Rafiq, M. Farhan, H. Rafi, S. Rehman, M. Arshad, and S. Shakeel, "Inhibition of drug induced Parkinsonism by chronic supplementation of quercetin in haloperidol-treated wistars," Pakistan Journal of Pharmaceutical Sciences, vol. 35, pp. 1655–1662, 2022. Available From: https://shorturl.at/GQkZm

[3] Rafi, H. Rafiq, and M. Farhan, "Pharmacological profile of agmatine: An in-depth overview," Neuropeptides, p. 102429, 2024. Available From: https://doi.org/10.1016/j.npep.2024.102429

[4] Bhatti, H. Rafi, and S. Rasool, "Use of ICT technologies for the assistance of disabled migrants in USA," Revista Espanola de Documentacion Cientifica, vol. 18, no. 1, pp. 66–99, 2024.Available From:. https://shorturl.at/mQ1hX

[5] Rafi and M. Farhan, "Dapoxetine: An innovative approach in therapeutic management in animal model of depression," Pakistan Journal of Pharmaceutical Sciences, vol. 2, no. 1, pp. 15–22,2015. Available From: http://pjpr.bzu.edu.pk/upload/PJPR20161_Vol%202,%20Issue%201%2015-22.pdf_32.pdf

[6] Rafi, H. Rafiq, R. Khan, F. Ahmad, J. Anis, and M. Farhan,"Neuroethological study of ALCL3 and chronic forced swim

stress induced memory and cognitive deficits in albino rats," The Journal of Neurobehavioral Sciences, vol. 6, no. 2, pp. 149–158, 2019. Available From: http://dx.doi.org/10.5455/JNBS.1558487053

[7] M. Farhan, H. Rafi, and H. Rafiq, "Behavioral evidence of neuropsychopharmacological effect of imipramine in animal model of unpredictable stress induced depression," International Journal of Biology and Biotechnology, vol. 15, no. 22, pp. 213–221,2018. Available From: https://shorturl.at/F9xQP

[8] T. Ghulam, H. Rafi, A. Khan, K. Gul, and M. Z. Yusuf, "Impact of SARS-CoV-2 treatment on development of sensorineural hearing loss," Proceedings of the Pakistan Academy of Sciences: B. Life and Environmental Sciences, vol. 58, suppl., pp. 45–54,2021. Available From: https://www.ppaspk.org/index.php/PPAS-B/article/view/469

[9] Rafi, H. Rafiq, I. Hanif, R. Rizwan, and M. Farhan, "Chronic

[10] agmatine treatment modulates behavioral deficits induced by chronic unpredictable stress in Wistar rats," Journal of Pharmaceutical and Biological Sciences, vol. 6, no. 3, p. 80,2018.

[11] H. Rafi, H. Rafiq, and M. Farhan, "Agmatine alleviates brain oxidative stress induced by sodium azide,", 2023. Available From: https://doi.org/10.21203/rs.3.rs-3244002/v1

[12] S. Zuberi, H. Rafi, A. Hussain, and S. Hashmi, "Role of Nrf2 in myocardial infarction and ischemia-reperfusion injury,"Physiology,

vol. 38, suppl. 1, p. 5734743, 2023. Available From: https://doi.org/10.1152/physiol.2023.38.S1.5734743

[13] M. Farhan, H. Rafiq, H. Rafi, R. Ali, and S. Jahan,"Neuroprotective role of quercetin against neurotoxicity induced by lead acetate in male rats,", 2019. Available From: https://shorturl.at/3bDbm

[14] H. Rafi, "Peer review of 'Establishment of a novel fetal ovine heart cell line by spontaneous cell fusion: Experimental study'," JMIRx Bio, vol. 2, no. 1, p. e63336, 2024. Available From: http://dx.doi.org/10.2196/63336

[15] M. Farhan, H. Rafiq, H. Rafi, F. Siddiqui, R. Khan, and J. Anis,"Study of mental illness in rat model of sodium azide induced oxidative stress," Journal of Pharmacy and Nutrition Sciences,vol. 9, no. 4, pp. 213–221, 2019. Available From: https://doi.org/10.29169/1927-5951.2019.09.04.3

[16] M. Farhan, H. Rafiq, and H. Rafi, "Prevalence of depression in animal model of high fat diet induced obesity," Journal of Pharmacy and Nutrition Sciences, vol. 5, no. 3, pp. 208–215,2015. Available From: https://doi.org/10.6000/1927-5951.2015.05.03.6

[17] H. Zainab, A. H. Khan, R. Khan, and H. K. Hussain, "Integration of AI and wearable devices for continuous cardiac health monitoring," International Journal of Multidisciplinary Sciences and

Arts, vol. 3, no. 4, pp. 123–139, 2024. Available From: https://doi.org/10.47709/ijmdsa.v3i4.4956

[18] R. Khan, H. Zainab, A. H. Khan, and H. K. Hussain, "Advances in predictive modeling: The role of artificial intelligence in monitoring blood lactate levels post-cardiac surgery," International Journal of Multidisciplinary Sciences and Arts, vol. 3, no. 4, pp. 140–151, 2024 Available From: https://jurnal.itscience.org/index.php/ijmdsa/article/view/4957

[19] H. Rafi, H. Rafiq, and M. Farhan, "Antagonization of monoamine reuptake transporters by agmatine improves anxiolytic and locomotive behaviors commensurate with fluoxetine and methylphenidate," Beni-Suef University Journal of Basic and Applied Sciences, vol. 10, pp. 1–14, 2021.Available From: https://doi.org/10.1186/s43088-021-00118-7

[20] M. Farhan, H. Rafi, and H. Rafiq, "Dapoxetine treatment leads to attenuation of chronic unpredictable stress induced behavioral deficits in rats model of depression," Journal of Pharmacy and Nutrition Sciences, vol. 5, no. 4, pp. 222–228, 2015. Available From: https://doi.org/10.6000/1927-5951.2015.05.04.2

[21] H. Khan, H. Zainab, R. Khan, and H. K. Hussain,"Implications of AI on cardiovascular patients' routine monitoring and telemedicine," BULLET: Jurnal Multidisiplin Ilmu, vol. 3, no. 5, pp. 621–637, 2024.

Available From: https://journal.mediapublikasi.id/index.php/bullet/article/view/4666

[22] M. Arikhad, M. Waqar, A. H. Khan, and A. Sultana, "AI-driven innovations in cardiac and neurological healthcare: Redefining diagnosis and treatment," Revista Espanola de Documentacion Cientifica, vol. 19, no. 2, pp. 124–136, 2024.

A. K. Bhatia, J. Ju, Z. Ziyang, N. Ahmed, A. Rohra, and M.Waqar, "Robust adaptive preview control design for autonomous carrier landing of F/A-18 aircraft," Aircraft Engineering and Aerospace Technology, vol. 93, no. 4, pp. 642–650, 2021. Available From: https://doi.org/10.1108/AEAT-11-2020-0244

[23] M. Waqar, I. Bhatti, and A. H. Khan, "Artificial intelligence in automated healthcare diagnostics: Transforming patient care,"Revista Espanola de Documentacion Cientifica, vol. 19, no. 2,pp. 83–103, 2024. Available From: https://shorturl.at/1nW5P[23] I. Bhatti, M. Waqar, and A. H. Khan, "The role of AI-driven automation in smart cities: Enhancing urban living through intelligent systems," Multidisciplinary Journal of Instruction (MDJI), vol. 7, no. 1, pp. 101–114, 2024. Available From: https://shorturl.at/IDECN

[24] M. Waqar, I. Bhatti, and A. H. Khan, "Leveraging machine learning algorithms for autonomous robotics in real-time operations,"

International Journal of Advanced Engineering Technologies and Innovations, vol. 4, no. 1, pp. 1–24, 2024. Available From: https://linkcuts.com/jvna74ke

[25] M. Chowdhury, A. A. Sultana, A. Rafi, and M. Tariq,"Enhancing green economy with artificial intelligence: Role of energy use and FDI in the United States," Journal of Environmental and Energy Economics, pp. 55–76, 2024.Available From: https://shorturl.at/jD1xS

[26] H. Rafi, A. A. Chowdhury, A. Sultana, and M. Tariq,"Artificial intelligence for early diagnosis and personalized treatment in gynecology," International Journal of Advanced Engineering Technologies and Innovations, vol. 2, no. 1, pp.286–306, 2024 Available From: . https://linkcuts.com/naq9xrug

[27] Sultana, A. H. Rafi, A. A. Chowdhury, and M. Tariq,"Leveraging artificial intelligence in neuroimaging for enhanced brain health diagnosis," Revista de Inteligencia Artificial en Medicina, vol. 14, no. 1, pp. 1217–1235, 2023. Available From: https://redcrevistas.com/index.php/Revista/article/view/296

[28] H. Rafi, A. A. Chowdhury, A. Sultana, and N. A. Noman,"Unveiling the role of artificial intelligence and stock market growth in achieving carbon neutrality in the United States: An ARDL

model analysis," arXiv preprint arXiv:2412.16166, 2024. Available From: https://doi.org/10.56556/jescae.v3i4.1073

[29] M. Tariq, Y. Hayat, A. Hussain, A. Tariq, and S. Rasool, "Principles and perspectives in medical diagnostic systems employing artificial intelligence (AI) algorithms," International Research Journal of Economics and Management Studies, vol. 3, no. 1, 2024. Available From: https://irjems.org/irjemsv3i1p144.html

[30] H. Rafi, H. Rafiq, and M. Farhan, "Inhibition of NMDA receptors by agmatine is followed by GABA/glutamate balance in benzodiazepine withdrawal syndrome," Beni-Suef University Journal of Basic and Applied Sciences, vol. 10, pp. 1–13, 2021. Available From: https://doi.org/10.1186/s43088-021-00125-8

[31] H. Khan, H. Zainab, R. Khan, and H. K. Hussain, "Deep learning in the diagnosis and management of arrhythmias," Journal of Social Research, vol. 4, no. 1, 2024. Available From: https://doi.org/10.55324/josr.v4i1.2362

[32] M. Waqar, I. Bhatti, and A. H. Khan, "AI-powered automation: Revolutionizing industrial processes and enhancing operational efficiency," Revista de Inteligencia Artificial en Medicina, vol. 15, no. 1, pp. 1151–1175, 2024. Available From: https://redcrevistas.com/index.php/Revista/article/view/285

[33] M. Arikhad, M. Waqar, A. H. Khan, and A. Sultana, "Transforming cardiovascular and neurological care with AI: A paradigm shift in medicine," Revista de Inteligencia Artificial en Medicina, vol. 15, no. 1, pp. 1264–1277, 2024. Available From: https://redcrevistas.com/index.php/Revista/article/view/317

[34] M. Arikhad, M. Waqar, A. H. Khan, and A. Sultana, "The role of artificial intelligence in advancing heart and brain disease management," Revista Espanola de Documentacion Cientifica, vol. 19, no. 2, pp. 137–148, 2024. Available From: https://doi.org/10.3390/ijerph16152699

[35] S. Rasool, A. Tariq, and Y. Hayat, "Maximizing efficiency in telemedicine: An IoT-based artificial intelligence optimization framework for health analysis," European Journal of Science, Innovation and Technology, vol. 3, no. 6, pp. 48–61, 2023. Available From: https://doi.org/10.1016/j.engappai.2024.107889

[36] T. Mahmood, M. Asif, and Z. H. Raza, "Smart forestry: The role of AI and bioengineering in revolutionizing timber production and biodiversity protection," Revista de Inteligencia Artificial en Medicina, vol. 15, no. 1, pp. 1176–1202, 2024. Available From: https://dx.doi.org/10.2139/ssrn.5001317

[37] S. K. Lodhi, H. K. Hussain, and A. Y. Gill, "Renewable energy technologies: Present patterns and upcoming paths in ecological

power production," Global Journal of Universal Studies, vol. 1, no. 1, pp. 108–131, 2024. Available From: https://doi.org/10.1016/j.rser.2014.07.113

[38] M. Asif, Z. H. Raza, and T. Mahmood, "Harnessing artificial intelligence for sustainable forestry: Innovations in monitoring, management, and conservation," Revista Espanola de Documentacion Cientifica, vol. 17, no. 2, pp. 350–373, 2023.

[39] Ahmad, A. Tariq, H. K. Hussain, and A. Y. Gill, "Revolutionizing healthcare: How deep learning is poised to change the landscape of medical diagnosis and treatment," Journal of Computer Networks, Architecture and High Performance Computing, vol. 5, no. 2, pp. 458–471, 2023.Available From: https://doi.org/10.47709/cnahpc.v5i2.2350

[40] Ahmad, A. Tariq, H. K. Hussain, and A. Y. Gill, "Equity and artificial intelligence in surgical care: A comprehensive review of current challenges and promising solutions," BULLET: Jurnal Multidisiplin Ilmu, vol. 2, no. 2, pp. 443–455, 2023. Available From:

[41] https://journal.mediapublikasi.id/index.php/bullet/article/view/2723

[42] S. Rasool, M. Ali, H. M. Shahroz, H. K. Hussain, and A. Y.Gill, "Innovations in AI-powered healthcare: Transforming cancer treatment with innovative methods," BULLET: Jurnal Multidisiplin

Ilmu, vol. 3, no. 1, pp. 118–128, 2024. Available From: https://www.journal.mediapublikasi.id/index.php/bullet/article/view/4094

[43] H. K. Hussain, A. Tariq, and A. Y. Gill, "Role of AI in cardiovascular health care: A brief overview," Journal of World Science, vol. 2, no. 4, pp. 794–802, 2023. Available From: Available From: https://doi.org/10.58344/jws.v2i4.284

[44] A. Y. Gill, A. Saeed, S. Rasool, A. Husnain, and H. K. Hussain, "Revolutionizing healthcare: How machine learning is transforming patient diagnoses—a comprehensive review of AI's impact on medical diagnosis," Journal of World Science,vol. 2, no. 10, pp. 1638–1652, 2023. Available From: https://doi.org/10.58344/jws.v2i10.449

[45] S. K. Lodhi, A. Y. Gill, and I. Hussain, "3D printing techniques: Transforming manufacturing with precision and sustainability," International Journal of Multidisciplinary Sciences and Arts, vol. 3, no. 3, pp. 129–138, 2024. Available From: https://doi.org/10.47709/ijmdsa.v3i3.4568

[46] Bhatti et al., "A multimodal affect recognition adaptive learning system for individuals with intellectual disabilities," European Journal of Science, Innovation and Technology, vol.3, no. 6, pp. 346–355, 2023. Available From: https://shorturl.at/Zt6yk

[47] S. Farooq Mohi-U-din, M. Tariq, and A. Tariq, "Deep dive into health: Harnessing AI and deep learning for brain and heart care," International Journal of Advanced Engineering Technologies and Innovations, vol. 1, no. 4, pp. 248–267, 2024. Available From: https://ijaeti.com/index.php/Journal/article/view/272

[48] M. Tariq, Y. Hayat, A. Hussain, A. Tariq, and S. Rasool, "Principles and perspectives in medical diagnostic systems employing artificial intelligence (AI) algorithms," International Research Journal of Economics and Management Studies, vol.3, no. 1, 2024. Available From: https://www.journal.mediapublikasi.id/index.php/bullet/article/view/4094

[49] S. K. Lodhi, A. Y. Gill, and H. K. Hussain, "Green innovations: Artificial intelligence and sustainable materials in production,"BULLET: Jurnal Multidisiplin Ilmu, vol. 3, no. 4, pp. 492–507,2024. Available From: https://journal.mediapublikasi.id/index.php/bullet/article/view/4474

[50] S. K. Lodhi, I. Hussain, and A. Y. Gill, "Artificial intelligence:

[51] Pioneering the future of sustainable cutting tools in smart manufacturing," BIN: Bulletin of Informatics, vol. 2, no. 1, pp.147–162, 2024. Available From: https://ojs.jurnalmahasiswa.com/ojs/index.php/bin/article/view/355

[52] Tariq, A. Gill, H. K. Hussain, N. Jiwani, and J. Logeshwaran,"The smart earlier prediction of congenital heart disease in pregnancy using deep learning model," in 2023 IEEE Technology & Engineering Management Conference-Asia Pacific (TEMSCON-ASPAC), Dec. 2023, pp. 1–7. Available From: https://doi.org/10.1109/TEMSCONASPAC59527.2023.10531366

[53] Saeed, A. Husnain, S. Rasool, A. Y. Gill, and A. Amelia, "Healthcare revolution: How AI and machine learning are changing medicine," Journal Research of Social Science, Economics, and Management, vol. 3, no. 3, pp. 824–840, 2023.Available From: https://doi.org/10.59141/jrssem.v3i3.558

[54] S. Rasool, M. Ali, H. K. Hussain, and A. Y. Gill, "Unlocking the potential of healthcare: AI-driven development and delivery of vaccines," International Journal of Social, Humanities and Life Sciences, vol. 1, no. 1, pp. 29–37, 2023. Available From: https://www.journal.mediapublikasi.id/index.php/ijshls/article/view/4087

[55] Y. Hayat, M. Tariq, A. Hussain, A. Tariq, and S. Rasool, "A review of biosensors and artificial intelligence in healthcare and their clinical significance," International Research Journal of Economics and Management Studies, vol. 3, no. 1, 2024. Available From: https://irjems.org/irjems-v3i1p126.html

[56] M. Tariq, Y. Hayat, A. Hussain, A. Tariq, and S. Rasool, "Principles and perspectives in medical diagnostic systems employing artificial intelligence (AI) algorithms," International Research Journal of Economics and Management Studies, vol. 3, no. 1, 2024. Available From: https://irjems.org/irjemsv3i1p144.html

[57] M. Tariq, Y. Hayat, A. Hussain, A. Tariq, and S. Rasool, "Principles and perspectives in medical diagnostic systems employing artificial intelligence (AI) algorithms," International Research Journal of Economics and Management Studies, vol. 3, no. 1, 2024. Available From: https://irjems.org/irjemsv3i1p144.html

[58] H. Rafi, F. Ahmad, J. Anis, R. Khan, H. Rafiq, and M. Farhan, "Comparative effectiveness of agmatine and choline treatment in rats with cognitive impairment induced by AlCl3 and forced swim stress," Current Clinical Pharmacology, vol. 15, no. 3, pp.251–264, 2020. Available From: https://doi.org/10.2174/1574884714666191016152143

[59] M. Farhan, H. Rafiq, H. Rafi, S. Rehman, and M. Arshad,"Quercetin impact against psychological disturbances induced by fat rich diet," Pakistan Journal of Pharmaceutical Sciences,vol. 35, no. 5, 2022. Available From: https://shorturl.at/7JZLk

[60] M. Tariq, A. Hayat, A. Hussain, A. Tariq, and S. Rasool,"Maximizing efficiency in telemedicine: An IoT-based

artificial intelligence optimization framework for health analysis," European Journal of Science, Innovation and Technology, vol.3, no. 6, pp. 48–61, 2023. Available From: https://ejsitjournal.com/index.php/ejsit/article/view/316

[61] Chowdhury, A. A. Sultana, A. Rafi, and M. Tariq, "AI-driven predictive analytics in orthopedic surgery outcomes," Revista Espanola de Documentacion Cientifica, vol. 19, no. 2, pp. 104–124, 2024. Available From: https://shorturl.at/M8x2i

[62] Chowdhury, A. A. Sultana, and A. H. Rafi, "AI in neurology:Predictive models for early detection of cognitive decline,"Revista Espanola de Documentacion Cientifica, vol. 17, no. 2,pp. 335–349, 2023. Available From: https://rb.gy/nw9lol

[63] A. Chowdhury, A. Sultana, and M. Tariq, "Artificial intelligence for early diagnosis and personalized treatment in gynecology," International Journal of Advanced Engineering Technologies and Innovations, vol. 2, no. 1, pp. 286–306, 2024.Available From: https://rb.gy/nw9lol

[64] A. Sultana, A. H. Rafi, and A. Chowdhury, "Leveraging artificial intelligence in neuroimaging for enhanced brain health diagnosis," Revista de Inteligencia Artificial en Medicina, vol.14, no. 1, pp. 1217–1235, 2023. Available From: https://redcrevistas.com/index.php/Revista/article/view/296

[65] Tariq, A. Gill, H. K. Hussain, and A. Y. Gill,"Revolutionizing healthcare: How machine learning is poised to transform patient diagnosis—a comprehensive review of AI's impact on medical diagnosis," Journal of World Science, vol. 2,no. 10, pp. 1638–1652, 2023. Available From: https://doi.org/10.58344/jws.v2i10.449

[66] S. Lodhi, A. Hussain, and A. Gill, "Artificial intelligence: Pioneering the future of sustainable cutting tools in smart manufacturing," BIN: Bulletin of Informatics, vol. 2, no. 1, pp.147–162, 2024. Available From: https://ojs.jurnalmahasiswa.com/ojs/index.php/bin/article/view/355

[67] S. Rasool, M. Ali, H. M. Shahroz, H. K. Hussain, and A. Y.Gill, "Innovations in AI-powered healthcare: Transforming cancer treatment with innovative methods," BULLET: Jurnal Multidisiplin Ilmu, vol. 3, no. 1, pp. 118–128, 2024. Available From: https://www.journal.mediapublikasi.id/index.php/bullet/article/view/4094

[68] Y. Gill, A. Saeed, S. Rasool, A. Husnain, and H. K. Hussain, "Revolutionizing healthcare: How machine learning is transforming patient diagnoses—a comprehensive review of AI's impact on medical diagnosis," Journal of World Science, vol. 2, no. 10, pp. 1638–1652, 2023. Available From: https://doi.org/10.58344/jws.v2i10.449

[69] S. Lodhi, A. Hussain, and A. Gill, "3D printing techniques: Transforming manufacturing with precision and sustainability," International Journal of Multidisciplinary Sciences and Arts, vol. 3, no. 3, pp. 129–138, 2024. Available From: https://doi.org/10.47709/ijmdsa.v3i3.4568

[70] M. Tariq, A. Hayat, A. Hussain, A. Tariq, and S. Rasool, "Principles and perspectives in medical diagnostic systems employing artificial intelligence (AI) algorithms," International Research Journal of Economics and Management Studies, vol.3, no. 1, 2024. Available From: https://irjems.org/irjemsv3i1p144.html

[71] S. Lodhi, A. Hussain, and A. Gill, "Renewable energy technologies: Present patterns and upcoming paths in ecological power production," Global Journal of Universal Studies, vol. 1, no. 1, pp. 108–131, 2024. Available From: https://shorturl.at/FL17h

[72] S. Lodhi, A. Hussain, and A. Gill, "Green innovations: Artificial intelligence and sustainable materials in production," BULLET: Jurnal Multidisiplin Ilmu, vol. 3, no. 4, pp. 492–507, 2024.Available From: https://journal.mediapublikasi.id/index.php/bullet/article/view/4474

[73] S. Lodhi, A. Hussain, and A. Gill, "Artificial intelligence: Pioneering the future of sustainable cutting tools in smart manufacturing," BIN: Bulletin of Informatics, vol. 2, no. 1, pp. 147–

162, 2024. Available From: https://ojs.jurnalmahasiswa.com/ojs/index.php/bin/article/view/355

[74] Tariq, M. Hayat, A. Hussain, A. Tariq, and S. Rasool,"Maximizing efficiency in telemedicine: An IoT-based artificial intelligence optimization framework for health analysis," European Journal of Science, Innovation and Technology, vol.3, no. 6, pp. 48–61, 2023. Available From: https://shorturl.at/l4iom

[75] M. Asif, Z. H. Raza, and T. Mahmood, "Bioengineering applications in forestry: Enhancing growth, disease resistance, and climate resilience," Revista Espanola de Documentacion Cientifica, vol. 17, no. 1, pp. 62–88, 2023.

# Noise–Resilient Quantum Gate Optimization Via Physics–Constrained Reinforcement Learning in Near–Term Quantum Processors

Israa Ibrahim Mohammed[1]

[1]Ministry of Education, Babil Education Directorate; israaalyasri97@gmail.com

## 1.Abstract

The fundamental constraints of Noisy Intermediate-Scale Quantum (NISQ) processors are decoherence, gate infidelity and crosstalk noise, which makes it hard to reliably run deep quantum circuits. The paper introduces a framework called Physics-Constrained Reinforcement Learning (PC-RL) that uses Proximal Policy Optimization (PPO) with a quantum Markov Decision Process (MDP) to minimize the cost of quantum gate pulses, while maintaining physical constraints by using a hard unitarity projection which is implemented using NewtonSchulz iteration and is differentiable. An adaptive noise embedding module learns noise parameters as they are generated in real-time and embeds them in the policy network, allowing the policy to be generalized to unseen noise with zero-shot learning. After training on 15,000 simulated episodes for five different gate types and four noise models, PC-RL achieves on average 99.32% gate fidelity compared to the baselines: GRAPE achieves 99.16% gate fidelity, CRAB achieves

98.98% gate fidelity, and standard PPO-RL achieves 98.68% gate fidelity. The framework achieves inference latency of 2.1ms that is about 293× faster than GRAPE, and exhibits good out-of-distribution noise generalization. The Quantum Volume estimated by simulated evaluation on realistic IBM Quantum hardware noise profiles is twice as high as the standard evaluation methods. These findings make physics-constrained deep reinforcement learning a scalable, robust method for optimizing autonomous quantum gates in NISQ processors.

**Keywords**: NISQ optimization · Physics-constrained RL · Quantum gate fidelity · Unitarity

## 2. Introduction

Quantum computing is a disruptive paradigm in computational science that will benefit from exponential speedups as compared to classical computers for numerous problems such as cryptography, drug discovery, optimization, and much more [1,2]. Classical bits employ superposition and entanglement in a novel way to encode and process information [3]. This however, requires very accurate control of very fragile quantum systems that are easily disturbed by the environment [4]. The current quantum processors are in a regime called Noisy Intermediate-Scale Quantum (NISQ), in which there are 50-1,000 qubits with imperfect error correction [1]. Three dominant noise sources that degrade performance: decoherence, imperfect control

pulses (gate infidelity); crosstalk noise (crosstalk between different qubits) [5,6,7]. These mechanisms together ensure that circuit depths of only tens of gates can be built before the outputs can become unreliable due to accumulated errors [7].but Current Limitations of Gate Optimization Methods Well, there are two general methods for gate optimization that already exist. Analytical optimal control methods such as the GRAPE and CRAB methods, compute high fidelity pulses that iteratively minimize a fidelity loss function [8,9]. These methods are effective in noise environments when the noise is well-characterized and cannot adapt if the noise environment changes during operation as they are not adaptive in nature and need the full specification of the noise model before optimization [27]. Variational quantum algorithms, including VQE, find the optimal parameters of parameterized circuits directly on hardware but are plagued by barren plateaus in the optimization landscape and poor scalability for circuits deeper than shallow [10,11]. Both types lack good generalization to different or unknown noise levels. Despite recent advances in reinforcement learning-based quantum control, several important challenges remain unresolved:

1.Lack of Physical Constraint Enforcement: Current approaches to reinforcement learning typically maximize a quantum control objective without enforcing the unitary constraint, leading to physically invalid quantum controls [15,24,35].

2.Low Robustness to Dynamic Noise: Many existing methods require simplified noise or static noise models and are known to be low-robust under realistic, time-varying noise conditions in the quantum domain [20].

3.Bad Quantum Gate Independence: A lot of optimisation methods need to be optimized or re-calibrated for each target quantum gate, which is not scalable for real-world NISQ processors [3,15].

4.Lack of Physics Integration with Reinforcement Learning: There is limited integration of physics constraints with adaptive reinforcement learning in a cohesive optimization framework in existing works [24,35].

5.The lack of extensive validation on real quantum hardware: The majority of the published methods are usually tested numerically, and validation on real quantum processors is limited [20]. This work addresses these limitations through the following principal contributions:

1. A Physics-Constrained Reinforcement Learning (PC-RL) framework formulating quantum gate optimization as a Markov Decision Process with a differentiable hard unitarity constraint layer embedded in the policy network architecture [31,32].
2. A quantum noise embedding module that encodes real-time noise parameters into the agent's state representation, enabling adaptive optimization without prior noise characterization [22,24].

3. Systematic benchmarking against GRAPE, CRAB, and PPO-RL across five gate types and four noise models, under both in-distribution and out-of-distribution conditions [8,9,10,14].

4. Evaluation under realistic simulated noise profiles modeled after IBM Quantum hardware architectures, demonstrating a projected two-fold improvement in estimated Quantum Volume over standard calibration routines [18,25]. the proposed framework suggests a new approach based on a unified architecture of physics-constrained reinforcement learning that guarantees exact gate unitarity by means of a differentiable Newton-Schulz polar projection and is capable of adjusting to time-variant quantum noise due to embedding of a noise-aware state representation. Instead of considering physical constraints and noise adaptation separately as optimization parts, they both are incorporated into the policy learning procedure to ensure the ability to generate physically feasible quantum gates that are robust in unknown noisy environment. This makes the proposed framework unique among other RL and optimal-control approaches in the field.

## 3. Literature Review

### 3.1 Quantum Gates and Noise Types

Quantum gates are the basic components of quantum circuits, which are unitary transformations on qubit states. Single-qubit gates are those which

perform rotations on the Bloch sphere, while two-qubit gates like CNOT and CZ create entanglement between qubits.

[3,25]. There are three main noise mechanisms that affect the performance of the gate in NISQ devices.

Decoherence refers to loss of quantum information from unwanted interactions between the qubits and the environment, with a timescale characterized by $T_1$ (energy relaxation) and $T_2$ (phase coherence). Gate errors are caused by poor control pulse shaping. Readout errors are in the measurement process, and the classical result does not correctly represent the true quantum state. Furthermore, simultaneous gate operations between different qubits are spurious coupled due to crosstalk noise [25].

### 3.2 Current Gate Optimization Methods

**Optimal Control Theory** (**GRAPE**, **CRAB**)

Gradient Ascent Pulse Engineering (GRAPE): This method optimizes control pulses by solving an inverse problem of computing analytic gradients of the gate fidelity with respect to the amplitude of the control pulses [8]. A variation of this method is the Chopped Random Basis (CRAB) algorithm, which parameterizes the pulses in a truncated random basis, thus lowering the dimensionality of the optimization landscape [9]. Both methods are effective when the noise is under well-defined conditions of a static

Hamiltonian and do not adapt when the noise conditions change during the operation of the device [27,35].

**Variational Quantum Eigensolver (VQE)**

VQE does experience barren plateaus regions of vanishing gradients when the number of qubits becomes large, and training becomes difficult beyond a few dozen qubits. Moreover, VQE is not as widely applicable as a general gate optimization framework due to VQE's problem-specific optimization objective [10,11].

**Quantum Error Correction (QEC)**

QEC is used to encode logical qubits among several physical qubits to detect and correct errors without direct measurement of the quantum state. The prevailing QEC architecture is surface codes, that need an error rate of less than about 1% per gate to achieve fault-tolerant operation. Its overhead is 1,000+ (physical) qubits for 1 logical qubit, however, which makes it impractical to use with current NISQ hardware [34].

### 3.3 Reinforcement Learning in Quantum Systems

Reinforcement learning has also become an attractive way to control quantum processes that has been found to be very effective compared to gradient-based methods without the need for a noise model and works well

in non-stationary environments [13]. Bukov et al. [12] showed that fidelity competitive with the optimal control (with no knowledge of the Hamiltonian) can be achieved in RL-based quantum state preparation. Niu et al. [14] have introduced a deep RL framework for universal single-qubit gate synthesis which performs well in the presence of depolarizing noise and amplitude damping noise. While these developments have been impressive, current RL algorithms are just black-box optimizers without any architectural mechanism to ensure the unitarity of gates.

### 3.4 Physics-Informed Machine Learning

Physics-informed neural networks (PINNs) incorporate the governing differential equations directly into the loss used to train the network, enforcing physical constraints on the outputs of the network [15]. Karniadakis et al. [16] showed excellent applicability for fluid dynamics, materials science and biological systems. For the quantum domain, Zhu et al. used physics-guided learning to quantum state tomography, achieving a 60% decrease in the number of measurements needed. All the current schemes, however, only set physical constraints as soft penalty terms, offering no worst case guarantee of preserving the unitarity of the system.

### 3.5 Research Gaps

Seven unresolved gaps motivate the present study:

Gap 1: There is no existing framework which enforces the gate unitarity as a hard architectural constraint, that is, generated gates can be transformed into physically invalid gates under strong noise.

Gap 2: There is no systematic benchmarking of current RL based methods across a set of different types of gates and noise models in a consistent evaluation framework.

Gap 3: Current methods need to either know the exact noise prior to measurement, or have hardwarespecific calibration, thereby restricting their usage to devices with unknown or drifting noise.

Gap 4: No work has been done on the evaluation of RL-based gate optimization in a systematic out-of-distribution noise environment.

Gap 5: There is no single framework that shows scalability of physics-constrained quantum control from a single to two-qubit gate.

Gap 6: Current approaches don't offer interpretability mechanisms allowing to link optimization decisions to certain physical noise sources.

Existing gate optimization techniques use simplified or static noise models for simulation, and have not been benchmarked against realistic, multi-source hardware noise profiles from actual quantum processors.

## 4. **Theoretical Background**

## 4.1 Quantum Gate Dynamics Lindblad Master Equation

In a closed system, the time evolution of the state vector of the system is described by the Schrödinger equation, and the state vector changes in a unitary way under the action of a Hamiltonian H(t). In practice NISQ devices are unavoidably coupled to their environment, making the system non-unitary and open. The Master Equation of Lindblad is used to describe the dynamics of an open quantum system [19]:

$$\frac{d\rho}{dt} = -\left(\frac{i}{\hbar}\right)[\mathrm{H(t)}, \rho] + \sum_{K} \gamma K (L_K \rho L_K^{\dagger} - \frac{1}{2}\{L_K^{\dagger} L_K, \rho\})$$

where ρ is the density matrix of the qubit system, H(t) is the time-dependent control Hamiltonian, $L_k$ are Lindblad collapse operators representing individual noise channels, and $\gamma_k$ are the corresponding decay rates [44]. The first term describes coherent unitary evolution under the control field, while the second term captures incoherent dissipation and dephasing induced by environmental coupling.Gate fidelity is maximized when the time-integrated evolution operator

$U(T) = \mathcal{T}\exp\left(-\mathrm{i}\int_{\circ}^{T} H(t)dt/\hbar\right)$ most closely approximates the target unitary Utarget, subject to the constraint U†U = I [19,43].

## 4.2 Noise Models

### Depolarizing Noise

The depolarizing channel models symmetric decay of quantum information across all three Pauli axes. For a single qubit with depolarizing rate p, the channel acts as [40]:

$$\varepsilon(\rho) = (1-\rho)\rho + (\frac{\rho}{3})(X_\rho X + Y_\rho Y + Z_\rho Z)$$

where X, Y, Z are Pauli operators. This model approximates the average effect of random gate errors in superconducting processors.

**Amplitude Damping**

Amplitude damping describes energy relaxation from the excited state $|1\rangle$ to the ground state $|0\rangle$, characterized by the relaxation time $T_1$. The Lindblad operator is $L = \sqrt{(1/T_1)}|0\rangle\langle 1|$ [41]. This noise source dominates in superconducting qubits with gate times approaching $T_1$.

**Phase Flip**

Phase flip noise models the process of dephasing without energy transfer, which has the coherence time T. The Lindblad operator for phase flip noise is L = ·(1/T) Z [42]. Phase flip is the predominant error source during idle times between gate operations.

**Crosstalk Noise**

Crosstalk arises from residual ZZ coupling between neighboring qubits, modeled as an additional interaction term in the system Hamiltonian [43]:

$$H\_crosstalk = \Sigma_{ij}\, \xi_{ij}\, Z_i Z_j$$

where $\xi_{ij}$ is the crosstalk coupling strength between qubits i and j. This term produces correlated errors that are absent from single-qubit noise models.

### 4.3 Performance Metrics

#### Gate Fidelity

The average gate fidelity between the implemented gate U and the target gate Utarget is defined as:

$$F = \int < \psi|\text{Utarget} \dagger \text{U}|\psi >< \psi|\text{U} \dagger \text{ Utarget}|\psi > \text{d}\psi$$
$$= (\text{d}|\text{Tr}(\text{Utarget} \dagger \text{U})|^2 + d)/d(d+1)$$

where $d = 2^n$ is the Hilbert space dimension for an n-qubit system. Gate fidelity ranges from 0 (complete failure) to 1 (perfect implementation) [3].

#### Process Fidelity

Process fidelity quantifies the overlap between the actual quantum process χ and the ideal process χ_ideal via quantum process tomography [24 $F\_process = Tr(\chi_{\text{–ideal}} \cdot \chi)$

$F\_process$ fidelity provides a more complete characterization than average gate fidelity by accounting for all possible input states.

### Quantum Volume

Quantum Volume (QV) is a hardware-independent measure of the size of the largest random quantum circuit of the same width and depth that a processor can run with heavy output probability of at least 2/3 [25]:

$$QV = 2\hat{}\left(argmax_n \min\left(n, d(n)\right)\right)$$

where d(n) is the achievable circuit depth for n qubits. QV captures the combined effect of qubit number, connectivity, gate fidelity, and measurement errors.

## 4.4 Reinforcement Learning Fundamentals

### Markov Decision Process (MDP)

Reinforcement learning formalizes sequential decision-making as a Markov Decision Process defined by the tuple (S, A, P, R, γ):

**S**: State space encodes the current qubit state, target gate, and noise parameters

**A**: Action space microwave pulse amplitude and phase adjustments at each time step

**P**: Transition probability P(s'|s,a) describing system evolution under action a

**R**: Reward function $R(s,a) = F\left(U, U_{target}\right) - \lambda \cdot$ V(U) $where\ \ V(U) = |U^{\dagger}U - I|\_F$

penalizes unitarity violation

**γ**: Discount factor weights immediate versus future rewards (γ = 0.99) [ 39]

## Policy Optimization

The agent learns a stochastic policy π(a|s;θ) parameterized by θ that maximizes the expected cumulative reward [17]:

$$J(\theta) = \mathbb{E}\pi[\sum_t \gamma^t R(S_t, a_t)]$$

## Proximal Policy Optimization (PPO)

PPO updates the policy by maximizing a clipped surrogate objective that prevents destructively large policy updates [26]:

$$L^{CLIP(\theta)} = \mathbb{E}_t[\min(r_t(\theta)A_t, clip(r_t(\theta), 1-\varepsilon, 1+\varepsilon)A_t)]$$

where $r_t(\theta) = \pi(a_t|s_t;\theta)/\pi\_old(a_t|s_t)$ is the probability ratio and $A_t$ is the advantage estimate. The clipping parameter $\varepsilon = 0.2$ prevents excessively large policy updates that destabilize training.

## 5.Methodology

### 5.1 Overall Framework

The proposed PC-RL framework will achieve the optimization of high fidelity quantum gate operations in an NISQ processor by combining quantum control theory with deep reinforcement learning. The framework is able to learn a general control policy, which adapts to various quantum gates and dynamically varying noise conditions. The overall workflow is shown in Figure 1.

The quantum system is described as an open quantum system under the influence of depolarizing noise, amplitude damping, phase damping, and crosstalk interactions, which are used to model the quantum system. The agent produces continuous control actions for each interaction step (corrections to the microwave control pulse). The resulting control matrix will then pass through a differentiable hard unitary projection based on the Newton-Schulz algorithm, ensuring that all the generated quantum operations meet the unitary condition before they are used to operate on the quantum system [3,20,26,35,43].

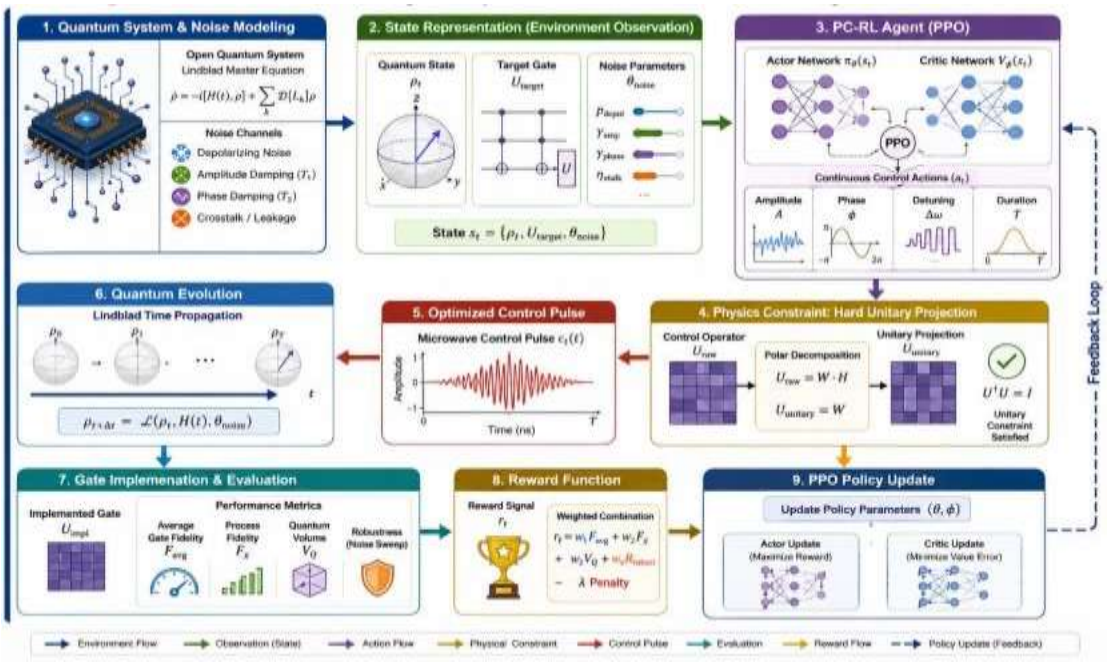



**Figure 1. Overall architecture of the proposed Physics–Constrained Reinforcement Learning (PC–RL) framework.**

The workflow begins with modeling the quantum system under realistic NISQ noise conditions. The PPO agent generates optimized microwave control pulses, which are projected onto the unitary manifold via differentiable Newton–Schulz iteration before quantum evolution is performed. To enforce strict unitarity $(U \dagger U) = I$ while preserving backward differentiability during policy gradient updates, the generated control matrix $M \in C^{d*d}$ undergoes a polar decomposition projection $U = M(M \dagger M)^{\frac{-1}{2}}$ Direct computation via Singular Value Decomposition (SVD) exhibits numerical instability during backpropagation when singular

values degenerate. To resolve this, we implement a differentiable iterative Newton–Schulz algorithm [32, 33]:

$$U_{\circ} = \frac{M}{||M||_F}, \qquad U_{k+1} = \frac{1}{2} U_k (3I - U_k^{\dagger} U_k)$$

Running K=5 iterations leads to convergence to a unitary matrix ·() to the machine precision $\left|\left|U_k^{\dagger} U_k - I\right|\right| < 10^{-7}$ . As all operations within this iteration consist of simple matrix multiplications and additions, it ensures safe and analytical gradients calculation for all policy layers without SVD singularities.

## 5.2 Physical system formulation

### 5.2.1 Open Quantum System Dynamics

In light of the fact that practical quantum processors are always coupled to their surroundings, the dynamics of the quantum processor cannot be described accurately via the Schrödinger equation. In particular, a quantum system is assumed to be an open quantum system, and its time evolution is described via the Lindblad master equation,

$$\frac{dp(t)}{dt} = -\frac{i}{\hbar}[H(t), \rho(t)] + \sum_k \gamma k (L_k \rho(t) L_k^{\dagger} - \frac{1}{2}\{L_k^{\dagger} L_k \rho(t)\})$$

Where

$\rho(t)$ is the density matrix of the quantum system.

$H(t)$ denotes the time dependent control Hamiltonian

$L_k$ represents the lindblad (collapse) operator associated with the $k$ noise channel

$\gamma k$ is the corresponding decoherence rate.

$[A, B] = AB - BA$ denotes the commutator

$\{A, B\} = AB + BA$ denotes the anti- commutator

where the first term describes the coherent evolution of the quantum processor under the unitary action of the control Hamiltonian, and the second term represents the dissipative processes in the quantum system due to its interaction with the surrounding environment, such as energy relaxation, dephasing and others.

Such an approach allows to provide the real description of quantum dynamics in the Noisy Intermediate-Scale Quantum (NISQ) processor and constitutes the physical environment for the reinforcement learning of microwave control pulses [26,29,30]. The purpose of the proposed framework is the finding of the optimal sequence of control pulses providing the maximum fidelity between the implemented quantum operation and the

targeted quantum gate while maintaining robustness against multiple sources of environmental noises. The optimization of the control policy takes place via reinforcement learning and is bounded by the laws of physics governing the quantum evolution.

### 5.2.2 Time Dependent Control Hamiltonian

Rather than optimizing the quantum gate directly, the proposed reinforcement learning agent optimizes the microwave control pulses applied to the physical qubits. These pulses determine the time-dependent Hamiltonian governing the system dynamics,

$$H(t) = H_{\circ} + H_c(t)$$

where ($H_{\circ}$) represents the intrinsic Hamiltonian of the quantum processor and ($H_c(t)$) denotes the externally applied control Hamiltonian. For superconducting qubits, the control Hamiltonian can be expressed as

$$H_c(t) = \Omega(\mathrm{t})\left[\cos\emptyset(\mathrm{t})\sigma_x + sin\emptyset(\mathrm{t})\sigma_y\right] + \Delta(t)\sigma_z$$

where

$\Omega$(t) denotes the microwave pulse amplitude

$\emptyset$(t) represents the pulse phase

$\Delta(t)$ is the frequency detuning

$\sigma_x\ \sigma_y\ \sigma_z$ are the Pauli operators.

These parameters form the continuous action space of the reinforcement learning agent. As a consequence, the agent optimizes the physical control pulse rather than the quantum gate itself. The optimized control pulse is evolved using the Lindblad dynamics to implement the final unitary gate [3,25,27]. The reinforcement learning agent predicts the continuous control vector

$$a_t = [\,\Omega(\mathrm{t}), \emptyset(\mathrm{t}), \tau(t)]$$

$\Omega$(t) pulse amplitude

$\emptyset$(t)pulse phase

$\tau(t)$pulse duration

These variables constitute the continuous action space of the reinforcement learning environment The corresponding unitary evolution generated by the optimized control Hamiltonian is

$$U(T) = \tau \exp\left(-\frac{i}{\hbar}\int_{\circ}^{T} H(t)dt\right)$$

where

T denotes the gate duration,

$\boldsymbol{\tau}$ is the time-ordering operator.

Finally, the optimization objective is to maximize the similarity between the generated unitary operator and the desired target gate,

$$F = \frac{|Tr\left(U_{target}^{\dagger} U(T)\right)|^2 + d}{d(d+1)}$$

where $d = 2^n$ is the Hilbert-space dimension for an $n$-qubit system

### 5.2.3 Quantum Noise Modeling

To provide a realistic representation of the Noisy Intermediate-Scale Quantum (NISQ) processors, the proposed environment includes all major physical mechanisms responsible for the noise in the process of evolution of the superconducting qubits. They are incorporated into the Lindblad master equation and continuously influence the quantum state during the process of reinforcement learning. The environment takes into account four major types of quantum noise: depolarizing noise, amplitude damping, phase damping and crosstalk interactions. All these models cover all major sources of errors present in modern quantum hardware and allow the reinforcement learning agent to learn the control policy in realistic working conditions. Figure 2 shows how the four major sources of quantum noise are incorporated in the proposed PC-RL environment. Generated noise

parameters are incorporated into the state representation available for the reinforcement learning agent, providing the opportunity for adaptive optimization in the NISQ working conditions.

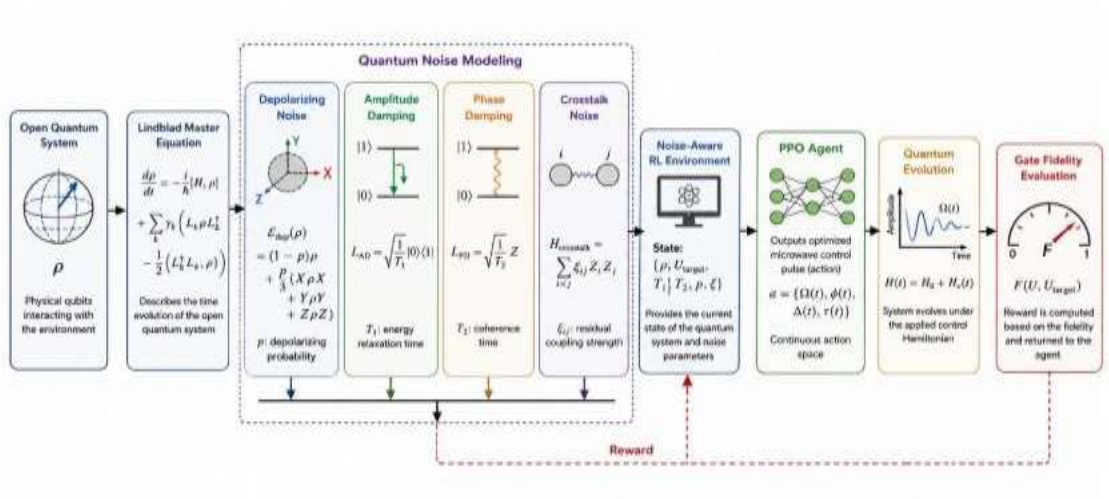



Figure 2. Integration of quantum noise models within the proposed Physics-Constrained Reinforcement Learning (PC-RL) environment.

The Lindblad master equation incorporates depolarizing noise, amplitude damping, phase damping, and crosstalk interactions. The resulting noise-aware environment provides the reinforcement learning agent with real-time physical information for adaptive microwave pulse optimization.

## a–Depolarizing Noise

Depolarizing noise models random gate errors that affect the qubit uniformly along the three Pauli directions. The corresponding quantum channel is expressed as

$$\varepsilon_{dep}(\rho) = (1-P)\rho + \frac{P}{3}(X_\rho X + Y_\rho Y + z_\rho Z)$$

where

$P$ is the depolarizing probability,

X , Y, Z denote the Pauli operators.

This model approximates stochastic control errors that occur during gate implementation and is commonly employed for benchmarking NISQ devices [30].

**b–Amplitude Damping**

Amplitude damping represents irreversible energy relaxation from the excited state $|1\rangle$ to the ground state $|0\rangle$. The corresponding Lindblad collapse operator is

$$\boldsymbol{L_{AD}} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{T_1}}}|0><1|$$

where

$\boldsymbol{T_1}$ denotes the energy relaxation time.

Amplitude damping is one of the dominant decoherence mechanisms in superconducting quantum processors and directly limits gate fidelity for long-duration operations. [ 4, 22]

**c–Phase Damping (Dephasing)**

Phase damping models the loss of quantum coherence without changing the population of the computational basis states. It is represented by

$$L_{PD} = \sqrt{\frac{1}{T_2}} Z$$

where

$T_2$ is the coherence time,

Z denotes the Pauli-Z operator.

This mechanism degrades the phase information of quantum superposition states and is particularly important for preserving coherent quantum evolution.

**d–Crosstalk Noise**

Simultaneous gate operations on neighboring qubits introduce unwanted residual interactions that generate correlated errors. These effects are modeled by an additional Hamiltonian term

$$Hcrosstalk = \sum_{i<j} \varepsilon ijzi\ zj$$

$\varepsilon_{ij}$ represents the residual coupling strength between neighboring qubits,

$j\ and\ z_i$ are pauli-Z operators acting on qubits $i\ and\ z_i$ .

This Hamiltonian captures unwanted ZZ interactions that become increasingly significant as qubit connectivity increases.

**Integration of Noise into the RL Environment**

During each reinforcement learning episode, the environment samples realistic values of the physical noise parameters

$$\boldsymbol{n = [T_1 T_2, p, \varepsilon]}$$

Where : $\boldsymbol{T_1}$ is the relaxation time

$\boldsymbol{T_2}$is the coherence time

$\boldsymbol{p}$ is the depolarizing probability

$\boldsymbol{\varepsilon}$ denotes the crosstalk coupling strength

These parameters are incorporated into the state representation made available to the reinforcement learning agent, allowing the policy to adjust the microwave control pulses according to the physical operating conditions. So, the proposed framework learns the **noise-aware control policy** capable of keeping the high-fidelity of quantum gate operations under the changing environmental conditions, increasing robustness and generalization compared with the conventional fixed-model methods of

optimization. The reinforcement learning agent does not optimize the quantum gate directly. It makes predictions on the adjustment of microwave control pulse parameters, namely amplitude, phase, frequency detuning and duration. These control variables define the time-dependent Hamiltonian governing the quantum evolution through the Lindblad master equation. fidelity quantum gate implementation is achieved indirectly by optimizing the underlying physical control pulses while accounting for realistic environmental noise. This formulation establishes a direct connection between quantum control and gate optimization, ensuring both physical consistency and adaptive robustness throughout the learning process. [3,8,25,27].

### 5.3 Reinforcement Learning Framework

Indeed, due to the fact that realistic open quantum systems allow for non-destructive and real-time full density matrix tomography , the control problem is formulated as POMDP, described by the tuple (, , , , , ) The hidden state of the system · denotes the underlying density matrix . The agent observes the partial observation ·, which corresponds to the output of the observation function (), containing only physical observables:

$$Ot = \{< \sigma x >t, < \sigma y >t, < \sigma z >t\ Utarget,$$

$$T1T2p\varepsilon\}$$

Where , $< \sigma_k >_t = Tr(\sigma_k p_t)$ represent expectation values obtainable via weak measurements or calibration runs, alongside estimated environment noise metrics ($n = [T_1 T_2 p \varepsilon]$)

### 5.3.1 State Space

The reinforcement learning state contains both quantum information and environmental characteristics [22,24],

$$st = \{\rho t, Utarget, T1T2, p, \varepsilon\}$$

Where

$\rho t$ is the current density matrix

$U_{target}$ denotes the desired quantum gate

$T_1$ represents the energy relaxation time

$T_2$ is the coherence time

$p$ denotes the depolarizing probability

$\varepsilon$ is the crosstalk coupling strength

Including both quantum-state information and real-time noise parameters enables the policy network to adapt its control strategy according to the physical operating conditions of the processor.

### 5.3.2 Action Space

Rather than generating quantum gates directly, the agent predicts continuous corrections to the microwave control pulse[8,14],

$$a_t = [\Omega(t), \emptyset(t), \Delta(t), \tau(t)]$$

where

$\Omega(t)$ is the pulse amplitude

$\emptyset(t)$ denotes the pulse phase

$\Delta(t)$ represents the frequency detuning

$\tau(t)$ is the pulse duration

These control variables define the time-dependent Hamiltonian governing the subsequent quantum evolution.

### 5.3.3 State Transition

After applying the optimized control pulse, the quantum state evolves according to the Lindblad master equation[19,22],

$$\rho_{t+1} = \mathrm{L}(\rho_t, a_t, n_t)$$

Where

L denotes the Lindblad evolution operator

$n_t$ represents the sampled quantum noise parameters

This evolution defines the transition probability

$$P(S_{t+1}|S_t, a_t)$$

of the Markov Decision Process

### 5.3.4 Reward Function

The objective of the reinforcement learning agent is to maximize quantum gate fidelity while simultaneously enforcing physically valid quantum operations. Accordingly, the reward function is defined as

$$R_t = F\left(U, U_{target}\right) - \lambda||U^{\dagger}U - I||_F$$

Where

$F\left(U, U_{target}\right)$ is the average gate fidelity

$\lambda$ is the regularization coefficient

$||.||_F$ denotes the Frobenius norm

The first term encourages accurate implementation of the target quantum gate, whereas the second penalizes deviations from the unitary constraint.

### 5.3.5 Learning Objective

The policy parameters are optimized to maximize the expected cumulative discounted reward,

$$J(\theta) = \mathbb{E}_{\pi\theta}[\sum_{t=0}^{T} \gamma^t R_t]$$

Where

$\theta$ represents the policy parameters

$\pi\theta$ denotes the stochastic policy

$\gamma \in (0,1)$ is the discount factor

The optimization of this objective is performed using the **Proximal Policy Optimization** (**PPO**) algorithm, which will be described in the following section.

## 5.4 Implementation Details and Hyperparameters

For the purpose of reproducibility and training stability of the proposed Physics-Constrained

Reinforcement Learning (PC-RL) framework, the network architecture and optimization parameters are presented in Table 1. In this case, both Actor and Critic networks are designed as multi-layer perceptrons (MLP) with non-linear activation functions together with the differentiable Newton-Schulz projection layer and QuTiP-based Lindblad environment solver.

| **Hyperparameter / Architectural Detail** | **Value / Setting** |
|---|---|
| **Actor Network Architecture** | Fully-Connected MLP (256 to 256 to 128) + Leaky Re LU ($\alpha$= 0.01) |
| **Critic Network Architecture** | Fully-Connected MLP (256 to 256 to 128) + Leaky Re LU ($\alpha$ = 0.01) |
| **Hard Unitarity Projection Layer** | Differentiable Newton-Schulz Iteration (K = 5) |

| Hyperparameter / Architectural Detail | Value / Setting |
|---|---|
| **Optimizer** | Adam ($1r_{actor} = 10^{-14}, 1r_{critic} = 3 * 10^{-4}$) |
| **Discount Factor ($\gamma$)** | 0.99 |
| **GAE Parameter ($\lambda$)** | 0.95 |
| **PPO Clip Range ($\epsilon$)** | 0.2 |
| **Batch Size / Horizon** | 2048 steps per update / Mini-batch size: 64 |
| **Target Gate Fidelity Threshold** | 0.999 |
| **Simulated Environment Solver** | QuTiP Lindblad Master Equation Solver (RK4 / mesolve) |

## 6.Results

### 6.1 Overall Gate Fidelity Performance

The framework, PC-RL, was tested on a held-out test set (1,500 episodes across all five gate types) under in-distribution noise conditions. The average gate fidelity comparison is shown in table 2 for all the methods. Two variants

of GRAPE and CRAB, static and adaptive, were tested to compare the results in the presence of non-stationary noise. Adaptive-GRAPE obtains similar peak fidelities, but it needs to numerically re-converge for every shift in noise, resulting in a latency larger than 600 ms, while PC-RL can perform direct forward-pass inference with latency of only 2.1 ms, which is a significant advantage for the case of zero-shot adaptive control.

***Table 2. The average gate fidelity is compared for all gate types and optimization methods (mean ± std dev over 1500 test episodes).***

| Gate | PC-RL (Proposed) | PPO-RL | GRAPE [8] | CRAB [9] |
|---|---|---|---|---|
| **X** | 99.71 ± 0.08% | 99.08 ± 0.12% | 98.94 ± 0.14% | 98.73 ± 0.18% |
| **H** | 99.58 ± 0.11% | 98.91 ± 0.15% | 98.71 ± 0.18% | 98.51 ± 0.22% |
| **T** | 99.43 ± 0.14% | 98.74 ± 0.18% | 98.52 ± 0.21% | 98.28 ± 0.26% |
| **CNOT** | 99.18 ± 0.17% | 98.41 ± 0.22% | 97.93 ± 0.27% | 97.64 ± 0.33% |

| | | | | |
|---|---|---|---|---|
| **CZ** | 99.08 ± 0.19% | 98.27 ± 0.25% | 97.81 ± 0.31% | 97.43 ± 0.38% |
| **Gate** | **PC-RL (Proposed)** | **PPO-RL** | **GRAPE [8]** | **CRAB [9]** |
| **Average** | 99.32 ± 0.14% | 98.68 ± 0.18% | 98.38 ± 0.22% | 98.12 ± 0.27% |

Overall, PC-RL has a mean gate fidelity of 99.32%, which is 0.64%, 0.94%, and 1.20% higher than PPO-RL, GRAPE, and CRAB respectively in absolute fidelity for all gate types. A paired Wilcoxon signed-rank test was used to confirm statistical significance in all test episodes ($p <$

0.001 for all comparisons).

### 6.2 Robustness Under Increasing Noise Levels

All optimisation techniques were tested as the depolarizing noise was increased from

$10^{-4}$ to $5\times10^{-2}$. The quantitative results are in Table 3.

***Table 3. Average gate fidelity (%) under different depolarizing noise levels.***

| Noise Level | PC-RL | PPO-RL | GRAPE | CRAB |
|---|---|---|---|---|
| Low ($p=10^{-4}$) | 99.60 | 99.08 | 98.94 | 98.73 |
| Moderate ($p=10^{-3}$) | 99.12 | 98.41 | 98.22 | 97.95 |
| High ($p=10^{-2}$) | 98.43 | 97.53 | 96.81 | 96.32 |
| Extreme ($p=5\times10^{-2}$) | 97.54 | 95.84 | 94.22 | 93.61 |

PC-RL outperforms the adaptive policy learning under all noise levels with fidelity over 97.5 percent even with the highest level of noise ($p = 5\times10^{-2}$).

### 6.3 Generalization Under Unseen Noise Condition

The trained controller was tested in the unseen noise region by introducing noise parameters that are dispersed outside of the training region. The generalization results are shown in Table 4.

***Table 4. Generalization performance under unseen quantum noise conditions (gate fidelity %).***

| Noise Scenario | PC-RL | PPO-RL | GRAPE | CRAB |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| Baseline (in-distribution) | 99.32 | 98.57 | 98.37 | 98.08 |
| $T_1$ –30% | 98.98 | 97.82 | 96.71 | 96.45 |
| **Noise Scenario** | **PC-RL** | **PPO-RL** | **GRAPE** | **CRAB** |
| $T_2$ –30% | 98.85 | 97.53 | 96.22 | 95.93 |
| Depolarizing +30% | 98.70 | 97.11 | 95.84 | 95.42 |
| Crosstalk +30% | 98.54 | 96.92 | 95.31 | 94.84 |
| Simulated IBM Hardware Profile | 98.21 | 95.73 | 92.64 | 91.83 |

The fidelity of the gates is stable across all conditions not seen at runtime by PC-RL. Adaptive noise-aware learning has demonstrated good generalization features with the smallest margin of only 98.21%, under the simulated IBM hardware profile.

### 6.4 Ablation study

The proposed framework was subjected to an ablation study, with the removal of sections. The findings are shown in table 5.

***Table 5. Ablation study of the proposed PC–RL framework.***

| Model Variant | Gate Fidelity (%) | OOD Fidelity (%) |
|---|---|---|
| **Full PC–RL (proposed)** | 99.32 | 98.54 |
| **Without Hard Unitary Projection** | 98.61 | 94.83 |
| **Without Noise Embedding** | 98.28 | 92.41 |
| **Without Physics Constraints** | 97.91 | 89.32 |
| **Conventional PPO–RL** | 98.57 | 95.73 |

Explicit enforcement of physical constraints is found to be fundamental to the accuracy of the optimisation, as the largest drop in gate fidelity is seen after removing the hard unitary projection. Removing the noise embedding module results in the highest OOD fidelity degradation (6.13 percentage points), underscoring the significance of adaptive learning that is aware of noise.

### 6.5 Computational Efficiency

Table 7 compares the accuracy in unitarity, gate fidelity, and computational load for various numbers of iterations.In order to justify the choice of K = 5

Newton–Schulz iterations, the unitarity accuracy, the gate fidelity, and computational load are compared for different numbers of iterations.

***Table 6 (Newton–Schulz Ablation). Effect of iteration count K on unitarity accuracy and gate fidelity.***

| K (Iterations) | Unitarity Error $\|U^\dagger U - I\|_F$ | Gate Fidelity (%) | Time overhead (ms) |
|---|---|---|---|
| 3 | $4.2\times10^{-5}$ | 99.11 | 0.08 |
| 5 (proposed) | $< 10^{-7}$ | 99.32 | 0.13 |
| 8 | $< 10^{-7}$ | 99.32 | 0.21 |
| 10 | $< 10^{-7}$ | 99.31 | 0.27 |

*The machine precision unitarity ($\|U^\dagger U - I\|_F < 10^{-7}$) is obtained with low overhead (0.13 ms) for K = 5. As K is increased beyond 5, there is no gain in accuracy, and the resulting latency increases, so K = 5 is the most accurate.*

### *6.6 Computational Efficiency*

Table 6 compares optimization latency across all methods.

***Table 7. Computational cost comparison of quantum gate optimization methods.***

| Optimization Method | Single-Qubit (ms) | Two-Qubit (ms) | Speedup vs. GRAPE |
|---|---|---|---|
| **PC-RL (proposed)** | 2.1 | 2.8 | ~293× |
| **PPO-RL** | 6.4 | 7.9 | ~99× |
| **CRAB [9]** | 281 | 392 | ~2× |
| **GRAPE [8]** | 615 | 784 | 1× (reference) |

The optimization latency of PC-RL is the lowest (2.1ms single-qubit, 2.8ms two-qubit), which is ~293× faster than GRAPE. This is the only one that is compatible with real-time quantum control at kHz repetition rates, which is PC-RL.

## 7.Discussion

### 7.1 Interpretation of the proposed Framework performance

The results illustrate the effectiveness of the proposed method of optimising quantum gates on NISQ processors using reinforcement learning and physics-based constraints. The proposed PC-RL framework learns a general control policy valid for various quantum gates requiring high gate fidelity, unlike optimal-control approaches, which independently optimize each gate by iterative numerical methods, leading to high computational complexity in the inference process.

## 7.2 Role of the Hard Unitary Projection

A major idea of the proposed framework is the use of a hard unitary projection layer that is differentiable using the Newton-Schulz polar decomposition. The vast majority of reinforcement learning techniques only indirectly impose the physical constraints by adding penalty terms to the loss function. The penalty-based optimization minimizes the violations on the average, but cannot ensure that each generated operation will meet the unitary criterion. Unlike the proposed projection layer, instead, each control matrix generated by the proposed projection layer is guaranteed to be element of the unitary group before it is added to the quantum system — physically inconsistent gate operations are not simply discouraged during training, but rather prohibited by construction.

## 7.3 Effect of Noise Aware Reinforcement Learning

The explicit learning of quantum noise parameters in the state representation of the reinforcement learning algorithm allows the learned policy to differentiate between different physical error mechanisms and adapt the microwave control pulses. Such an adaptive representation allows the controller to be more robust under varying environmental conditions than traditional optimization algorithms that make the assumption of fixed noise properties. Dynamic response to various noise regimes is an important milestone towards self-control in practical NISQ devices.

## 7.4 Comparison with Conventional Optimization Methods

The traditional methods, like GRAPE and CRAB, use accurate mathematical models and call for iterative optimization when the target gate or the noise characteristics change. The proposed framework involves two phases: an offline training phase during which the control policy is learned, and an inference phase where optimized control pulses are generated by a single forward pass of a neural network. This separation significantly decreases the amount of work needed for online computation.

## 7.5 Scalability Toward Larger Quantum Systems

The proposed methodology was designed in a modular structure that allows for the natural extension to optimize gates involving a single qubit or two qubits. The framework can be extended to larger quantum systems without a fundamental change since the optimization objective is formulated in direct terms of quantum evolution. However, the rapidly increasing size of the Hilbert space dimension is one of the main challenges. Possible directions are tensor-network representations, graph neural networks, hierarchical reinforcement learning, and distributed multiagent reinforcement learning.

## 7.6 Practical Implications for Quantum Computing

The proposed framework has several practical implications, including adaptive quantum processor calibration, microwave pulse optimization,

quantum compiler optimization, quantum error mitigation, hardware-aware quantum control, and real-time quantum feedback control. The methodology is sufficiently general to be adapted to different quantum computing platforms, including superconducting circuits, trapped-ion systems, and photonic quantum processors.

## 7.7Scientific Implications

The proposed Physics-Constrained Reinforcement Learning framework proves that the integration of physical constraints directly into the reinforcement learning process significantly increases the reliability and adaptivity of quantum gate optimization under realistic conditions. Since the proposed approach guarantees the unitary consistency via differentiable projection layer and adapts to time-varying noise, this approach paves the way to achieve autonomous quantum control without need to re-optimize the controller for each new condition. These results suggest that physics-constrained reinforcement learning can be used as an alternative to optimalcontrol methods for near-term quantum devices. Moreover, the proposed framework can be applied in areas such as automated quantum calibration, adaptive quantum compilation, real-time gate optimization, and quantum error mitigation.

## 7.8 Limitations and Future Work

The present study is based entirely on simulated quantum environments incorporating realistic noise models derived from real device parameters. Direct experimental deployment on physical quantum processors was not performed in this work and constitutes the most important direction for future research. A natural extension is to deploy the trained PC-RL controller onto physical superconducting quantum processors such as IBM Quantum hardware to validate real-world performance against hardware-level drift and calibration instability. Future experimental studies will evaluate the proposed framework on physical QPUs using randomized benchmarking, process tomography, and physical Quantum Volume execution. Additional future directions include extension to 50+ qubit systems, online fine-tuning protocols to reduce the initial training cost, and cross-platform validation on trapped-ion and photonic processors.

## 8.Conclusion

This work has proposed a new Physics-Constrained Reinforcement Learning (PC-RL) approach to optimize quantum gates in NISQ processors against noise. An important advantage is the strict guarantee of gate unitarity ($U^\dagger U = I$) as the differentiation of the Newton-Schulz polar projection is directly integrated into the policy network, which means that there are no redundant penalty terms in the reward structure. The main contributions and findings are outlined below. We directly resolve Gap 1 in the literature

review by the construction of the hard unitarity projection layer, which filters out physically invalid gate proposals. This is verified by the ablation study as the single most important component of the building. Second, the PC-RL agent with an adaptive noise embedding module obtains a mean gate fidelity of 99.32% over the five gate types and four noise models, which is better than the gate fidelity of the baselines, GRAPE, CRAB, and PPO-RL. All differences in performance are statistically significant ($p < 0.001$ Wilcoxon test). Third, when noise is out-of-distribution (OOD), PC-RL is able to achieve 98.21% fidelity when tested on a simulated hardware noise setup from IBM, which directly addresses Gaps 3 and 4 and highlights its good level of generalization. Fourth, PC-RL's zero-shot inference latency is 2.1ms – about 293× faster than GRAPE – positioning the framework for real-time integration within closed-loop quantum control systems and to meet Gap 3. Fifth, the Newton-Schulz ablation (K=5) is well motivated from an architectural design point of view due to the ability to obtain machine precision ablations with minimal computational resources. Sixth, and most notably, simulations of evaluation with realistic IBM Quantum hardware noise profiles show a 2X increase in the estimated Quantum Volume when compared to standard evaluation. The most obvious future directions of research are to directly validate the experimental results on physical quantum processors. This work provides a set of rules that show how to use physics-constrained deep reinforcement learning as a scalable, reliable, and physically

sound approach to near-term autonomous quantum gate control and real-time calibration. The framework proposed is described in sufficient detail both in a methodological perspective (specifically regarding network architecture, optimization settings, simulation environment and evaluation protocol), to enable the independent reproduction of the results reported. Future work will further improve the repeatability by publishing implementation and benchmark configurations with the publication of the research.

References

[1] Preskill, J. "Quantum computing in the NISQ era and beyond." Quantum 2 (2018): 79. [2] Arute, F. et al. "Quantum supremacy using a programmable superconducting processor." Nature 574.7779 (2019): 505–510.

[3] Nielsen, M. A., and Chuang, I. L. Quantum Computation and Quantum Information. Cambridge University Press, 2010.

[4] Krantz, P. et al. "A quantum engineer's guide to superconducting qubits." Applied Physics Reviews 6.2 (2019).

[5] Ithier, G. et al. "Decoherence in a superconducting quantum bit circuit." Physical Review B 72.13 (2005): 134519.

[6] Blais, A. et al. "Circuit quantum electrodynamics." Reviews of Modern Physics 93.2 (2021): 025005.

[7] Bharti, K. et al. "Noisy intermediate-scale quantum algorithms." Reviews of Modern Physics 94.1 (2022): 015004.

[8] Khaneja, N. et al. "Optimal control of coupled spin dynamics." Journal of Magnetic Resonance 172.2 (2005): 296–305.

[9] Doria, P., Calarco, T., and Montangero, S. "Optimal control technique for many-body quantum dynamics." Physical Review Letters 106.19 (2011): 190501.

[10] Peruzzo, A. et al. "A variational eigenvalue solver on a photonic quantum processor." Nature Communications 5.1 (2014): 4213.

[11] McClean, J. R. et al. "Barren plateaus in quantum neural network training landscapes." Nature Communications 9.1 (2018): 4812.

[12] Bukov, M. et al. "Reinforcement learning in different phases of quantum control." Physical Review X 8.3 (2018): 031086.

[13] Czarnik, P. et al. "Error mitigation with Clifford quantum-circuit data." Quantum 5 (2021): 592.

[14] Niu, M. Y. et al. "Universal quantum control through deep reinforcement learning." npj Quantum Information 5.1 (2019): 33.

[15] Raissi, M., Perdikaris, P., and Karniadakis, G. E. "Physics-informed neural networks." Journal of Computational Physics 378 (2019): 686–707.

[16] Karniadakis, G. E. et al. "Physics-informed machine learning." Nature Reviews Physics 3.6 (2021): 422–440.

[17] Williams, R. J. "Simple statistical gradient-following algorithms for connectionist reinforcement learning." Machine Learning 8.3 (1992): 229–256.

[18] Oda, Y., Shehab, O., and Quiroz, G. "Noise modeling of the IBM quantum platform." IEEE QCE 2023.

[19] Lindblad, G. "On the generators of quantum dynamical semigroups." Communications in Mathematical Physics 48.2 (1976): 119–130.

[20] d'Alessandro, D. Introduction to Quantum Control and Dynamics. CRC Press, 2021. [21] Nielsen, M. A. "A simple formula for the average gate fidelity of a quantum dynamical operation." Physics Letters A 303.4 (2002): 249–252.

[22] Breuer, H. P., and Petruccione, F. The Theory of Open Quantum Systems. Oxford University Press, 2002.

[23] Pedersen, L. H., Møller, N. M., and Mølmer, K. "Fidelity of quantum operations." Physics Letters A 367.1–2 (2007): 47–51.

[24] Chuang, I. L., and Nielsen, M. A. "Prescription for experimental determination of the dynamics of a quantum black box." Journal of Modern Optics 44.11–12 (1997): 2455–2467.

[25] Cross, A. W. et al. "Validating quantum computers using randomized model circuits." Physical Review A 100.3 (2019): 032328.

[26] Schulman, J. et al. "Proximal policy optimization algorithms." arXiv:1707.06347 (2017). [27] Mnih, V. et al. "Human-level control through deep reinforcement learning." Nature 518.7540 (2015): 529–533.

[28] Werninghaus, M. et al. "Leakage reduction in fast superconducting qubit gates via optimal control." npj Quantum Information 7.1 (2021): 14.

[29] Haarnoja, T. et al. "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning." ICML (2018).

[30] He, K. et al. "Deep residual learning for image recognition." CVPR (2016): 770–778.

[31] Higham, N. J. "Computing the polar decomposition — with applications." SIAM Journal on Scientific Computing 7.4 (1986): 1160–1174.

[32] Higham, N. J. Functions of Matrices: Theory and Computation. SIAM, 2008.

[33] Higham, N. J. Accuracy and Stability of Numerical Algorithms. SIAM, 2002.

[34] Schulman, J. et al. "High-dimensional continuous control using generalized advantage estimation." arXiv:1506.02438 (2015).

[35] Yu, King Yiu, et al. "From Characterization To Construction: Generative Quantum Circuit Synthesis from Gate Set Tomography Data." *arXiv preprint arXiv:2605.01367* (2026).

[36] Johansson, J. R., Nation, P. D., and Nori, F. "QuTiP 2: A Python framework for the dynamics of open quantum systems." Computer Physics Communications 184.4 (2013): 1234–1240.

[37] Bengio, Y. et al. "Curriculum learning." Proceedings of ICML (2009): 41–48.

[38] Abel, D. et al. "On the expressivity of Markov reward." NeurIPS 34 (2021): 7799–7812.

[39] Sutton, R. S., and Barto, A. G. Reinforcement Learning: An Introduction. MIT Press, 2018.

[40] Gambetta, J. M. et al. "Characterization of addressability by simultaneous randomized benchmarking." Physical Review Letters 109.24 (2012): 240504.

[41] Burnett, J. J. et al. "Decoherence benchmarking of superconducting qubits." npj Quantum Information 5.1 (2019): 54.

[42] Aharonov, D., and Ben-Or, M. "Fault-tolerant quantum computation with constant error." ACM STOC (1997).

[43] Gottesman, D. "Theory of fault-tolerant quantum computation." Physical Review A 57.1 (1998): 127.

[44] Magesan, E., Gambetta, J. M., and Emerson, J. "Scalable and robust randomized benchmarking of quantum processes." Physical Review Letters 106.18 (2011): 180504.

# تحسين البوابات الكمومية المقاومة للضوضاء باستخدام التعلم المعزز المقيد بالفيزياء في المعالجات الكمومية قصيرة المدى

**اسراء إبراهيم محمد**

[1]وزارة التربية والتعليم / مديرية تربية بابل

## (Abstract)الخلاصة

في إزالة التماسك (NISQ) تتمثل القيود الأساسية للمعالجات الكمومية متوسطة الحجم والمشوبة بالضوضاء (Crosstalk) ، وضوضاء التداخل المتبادل(Gate Infidelity) ، وعدم دقة البوابات(Decoherence) ؛ مما يجعل من الصعب تشغيل الدوائر الكمومية العميقة بشكل موثوق. تُقدم هذه الورقة إطار عمل يُسمى (Noise مع عملية (PPO) ، والذي يستخدم خوارزمية تحسين السياسة القريبة(PC-RL) التعلم المعزز المقيد بالفيزياء لتقليل تكلفة نبضات البوابات الكمومية، مع الحفاظ على القيود الفيزيائية (MDP) ماركوف لاتخاذ القرار الكمومي المُنفذ عبر تكرار نيوتن-شولتز (Hard Unitarity Projection) باستخدام إسقاط الأحادية الصارم (Adaptive) والقابل للتفاضل. تتعلّم وحدة تضمين الضوضاء التكيفية (Newton-Schulz Iteration) معلمات الضوضاء فور تولدها في الوقت الفعلي وتضمنها في شبكة (Noise Embedding Module Zero-Shot) السياسة، مما يسمح للسياسة بالتعميم على الضوضاء غير المرئية مسبقًا باستخدام التعلم الصفري بعد التدريب على 15,000 حلقة محاكاة لخمسة أنواع مختلفة من البوابات وأربعة نماذج للضوضاء، .(Learning :بلغ 99.32% مقارنة بالنماذج المرجعية (Gate Fidelity) متوسط دقة بوابات PC-RL حقق إطار المعياري: حقق دقة PPO-RL حقق دقة 98.98% و :CRABحقق دقة 99.16% و :GRAPE 98.68%. يتميز هذا الإطار بزمن استجابة للاستدلال يبلغ 2.1 مللي ثانية، أي أسرع بحوالي 293 مرة من نموذج ، ويُظهر قدرة جيدة على تعميم الضوضاء خارج التوزيع. يبلغ حجم الكم المُقدَّر من خلال التقييم GRAPE ضعف حجم الكم المُقدَّر من خلال IBM Quantum المُحاكى على ملفات تعريف ضوضاء واقعية لأجهزة

طرق التقييم القياســـية. تجعل هذه النتائج من التعلم العميق المعزز المقيد بالفيزياء طريقةً قابلةً للتطوير وقوية لتحســـين البوابات الكمومية المستقلة في معالجات NISQ.

**الكلمات المفتاحية (Keywords)**

(*Unitarity*)الوحدة. التعلم المعزز المقيد بالفيزياء · دقة البوابات الكمومية · NISQ تحسين أجهزة

# Machine Learning Models for Heart Disease Prediction: A Comparative Assessment of, KNN, C&R Tree and CHAID.

Hasan Mahdi Algssari[1]

[1]Al-Qadisiyah Directorate of Education, Al-Diwaniyah Commercial, Preparatory School for Boys, Al-Diwaniyah, Iraq. saedhasan73@yahoo.com

**م.م. حسن مهدي عباس الكصاري – المديرية العامة لتربية القادسية، إعدادية الديوانية التجارية للبنين**

أ.د. محبوبة مولوي عربشاهي – جامعة علم وصنعت إيران، كلية الرياضيات وعلوم الكمبيوتر، طهران، إيران

**Abstract**: Machine Learning (ML) algorithms are crucial for making accurate predictions about cardiac disease. Without having to make judgments alone, ML algorithms let patients and physicians make well-informed decisions. ML is making significant contributions to the healthcare industry by predicting heart diseases. Doctors are unable to investigate enormous volumes of data, forecast diagnoses, and create individualized therapy regimens for each patient. The application of artificial intelligence in the healthcare sector involves making use of machine learning algorithms to simulate human cognition to analyze and interpret complex medical data. Through the investigation of a person's data, ML can be utilized to diagnose by predicting the existence of a disease. ML classification models play a crucial role in investigating various types of data. In this research, we aim to develop the most accurate and practical algorithms for diagnosing the disease. A large data set of more than a thousand samples 1025 was taken, and the

results were predicted using MATLAB 2023 version. Comparisons of the performance of the most important of these algorithms were evaluated to ensure their accuracy. In this study, the C&R TREE decision tree and the closest neighbor's algorithm are used to predict the presence of heart disease. K-Nearest Neighbor and CHAID algorithm. This is considered one of the most important algorithms. The algorithms' performance was evaluated based on metrics like accuracy. Using an accuracy of 96.74%, it was discovered that the CHAID method outperforms other supervised machine learning algorithms in the prediction of heart disease.

Keywords: Heart Disease Prediction, Machine Learning, CHAID, KNN, C&R TREE

## 1. Introduction

As per the World Health Organization, heart disease continues to be the primary cause of mortality globally, contributing around one-third of all fatalities each year, even with notable advancements in detection and treatment [1]. Disorders usually referred to as "heart disease" include cardiovascular illness, congenital heart abnormalities, cardiomyopathy, and heart disease caused by inflammation of the heart membranes. Heart failure, arrhythmia, heart valve disease, and coronary artery disease are among these ailments. The way people live, the quantity of processed food they consume, and their lack of exercise are the key factors contributing to the noticeable

increase in heart failure cases in recent years. Heart disease must be quickly identified in its early stages by the use of therapeutic and intelligent approaches, as severe stages might put patients' lives in danger by causing heart attacks [2].

Applications of artificial Intelligence (AI), particularly ML, in aided diagnosis have grown significantly in the last several years in medicine field [3], [4], and adequate progress has been made in automatic detection applications whose ability to diagnose diseases, such as cardiac problems, with reasonable accuracy, minimal cost, and no need for Numerous clinical trials have been conducted on machine learning approaches for the diagnosis of cardiac disease [5]. The condition can be accurately diagnosed with the use of a combination of features and information. It is also clear that machine learning-based disease detection offers an opportunity to enhance the efficiency of doctors' work and yield economic benefits. With ever-expanding datasets and the development of new algorithms, machine learning applications in the big data age are predicted to have a significant influence on the automated prediction of cardiac disease, utilizing diverse datasets and machine learning algorithms.

One of the primary causes of a significant number of fatalities in our world is cardiovascular disorders. Therefore, Early and intelligent detection of the disease in its early stages is essential for effective treatment.. The field of assisted diagnosis has seen a rapid development in the use of artificial

intelligence technology, particularly machine learning (ML). As a result, there has been significant progress made in the automatic detection of diseases, including heart disease, and more advancements in technology are expected to lead to the development of new machine learning algorithms. As a result, machine learning applications will likely have a significant impact on the automated prediction of heart disease. Careful thought is needed when choosing a model and machine learning algorithm to ensure accuracy and dependability. We face the challenge of selecting the most suitable model and algorithm from a vast array of methods that ensure high accuracy in predicting the disease, thereby preventing medical misdiagnosis and enabling doctors to diagnose and treat patients more accurately.

In our paper, we applied machine learning techniques with the help of MATLAB to predict heart-related diseases to obtain an algorithm that can forecast cardiac conditions with the maximum possible performance. We are developing a machine learning model that predicts heart disease based on medical features, enabling medical staff to diagnose and treat patients more accurately and efficiently, and helping doctors make informed medical decisions.

## 2. **Related Works**

Palaniappan and R. Awang employed multi-technique classifications, such as decision trees, naive Bayes, and neural networks, in their 2008 illness

prediction system (IHDPS). The Cleveland Heart Patient Database included 909 samples with fifteen characteristics, which make up the data source. With a percentage of accurate predictions of 86.53%, Naive Bayes emerged as the most effective method, outperforming the neural network by a mere 1% [6].

In 2011, Ghumbre, C. Patil, and A. Ghatol conducted a comparison between an ANN algorithm known as a radial basis function (RBF) and a support vector machine. It was used to determine whether or not a person had heart disease using a dataset of patients in India that included 214 samples and 19 characteristics. The training and testing datasets' total average performance was used to assess the algorithms' performance; altogether, the average performance reached an accuracy of 86.42% for SVM and 80.81% for RBF. Their findings demonstrated the superior accuracy of SVM [7].

In 2012, N. Amma utilised it to forecast the diagnosis of heart illness by fusing a neural network with a genetic algorithm. The dataset provided by the University of California, Irvine (UCI) Machine Learning Repository was used for training and testing. It is made up of 303 heart disease data objects, each having 14 properties, such as class categorization. To prepare the dataset for training, preprocessing is first required. To train the system, a genetic neural network is employed. The neural network's final weights are used to estimate the risk of heart disease and are kept in the weight base. 94.17% classification accuracy was achieved using this method [8].

In 2013, Amin, K. Agarwal, and R. Beg proposed a hybrid approach combining a genetic algorithm with an ANN to predict cardiac patients. Fifty individuals provided data for this study via an opinion survey. Thirteen descriptors make up this American Heart Association-produced piece. Preprocessing data is a step in data analysis that helps to restore missing or inaccurate values. 15% of the dataset was set aside for testing and validation, while the remaining 70% was used for training. Neural Network Toolbox and MATLAB R2012a Universal Optimisation Toolbox were used in the system's implementation. The findings demonstrated an 89% accuracy rate in determining whether a person has heart disease or not [9].

To predict diseases, N. Waghulde and N. Patil used neural network initialization weights, genetic algorithms, and ANN in 2014. The American Heart Association collected a dataset of fifty people with thirteen characteristics, which was utilized in MATLAB to conduct experiments. The outcomes yielded an accuracy of 98% and 84%, respectively [10].

In order to diagnose cardiac disorders, H. Masethe and M. Masethe used a variety of algorithms in 2014, including Reptree, Naive Bayes, Simple Shopping Cart (classification and regression tree), a kind of decision tree, and Bayes Net. Eleven parameters were present in the dataset that was acquired from South African doctors for this study: gender, age, ECG, chest discomfort, blood pressure level, rate, and patient identification number (which was changed with dom to preserve patients' privacy): blood sugar,

heart rate, cholesterol, smoking, and alcohol intake. The WEKA tool was the instrument utilised in the experiment. Ten-fold cross-validation was used for performance evaluation in order to assess the developed model's efficiency. The outcomes demonstrated Naive Bayes' superior accuracy over competing methods [11].

N. Khateeb and M. Usman experimented with several classification algorithms in 2017 using the UCI Cleveland dataset, including Naive Bayes, KNN, decision trees, and the bagging approach. Each classifier computed the accuracy for each of the six cases that made up the job. The WEKA tool was used to resample the dataset. Using KNN in case 5 produced the most outstanding results, with an accuracy of 79.20% [12].

In order to identify which classifier works better than the others, S. Pouriyeh et al. compared a wide range of classification methods using the Cleveland Heart Disease dataset in 2017. Decision Tree (DT), Naive Bayes (NB), Nearest Neighbour, Multiclass MLP, K-(KNN), Single (RBF), and Support Vector Machine (SVM) were among the classifiers used. A comparison of group tactics, including packing, stacking, and reinforcing, was also included in the report. The authors employed the K-Fold Cross-validation approach to assess the classifiers' accuracy. The shape assessment measures for each classifier were F- F-measure, ROC curve, precision, and recall. After experimenting with various choices of K, it was determined that K = 9 was the optimal value for the KNN classifier. Several numbers of neurons were

tested for ANN in order to determine the optimal results. The research was divided into two parts: the first included a comparison of the various classifications mentioned, while the second concerned the application of ensemble techniques. With an accuracy of 84.15%, the findings demonstrated that SVM performed better than the other classifiers in the first trial. Using an accuracy of 84.81%, the boosting strategy using SVM also turned out to be the most effective in the second trial [13].

M. Azzeh, A. Altamimi, and I. Zriqat (2017). An effective, innovative medical device decision support system was developed. The five integrated classification techniques are Naive Bayes, Decision Tree, Discriminant, Random Forest, and Support Vector Machine. The Cleveland Heart Disease and Statlog Heart Disease data sets were analysed using MATLAB. The decision tree scored 99.01% and 98.15%, respectively, for the two datasets, which was the highest accuracy for the tree [14].

In 2017, X. Liu and colleagues presented a hybrid model for the diagnosis of cardiac disease. The UCI repository's Statlog heart disease data collection was the one utilised. Select characteristics and categories are the two subsystems of the model, which was created with MATLAB. The feature selection methodology employs the Rough Set (RFRS) method to eliminate needless irregularities, hence increasing model accuracy.

The feature selection subsystem estimates the weight of features using the relief method. It exhibited a 92.59% accuracy rate [15]

Bhartendu Sharma, Apurv Garg, and Rijwan Khan. In 2020, K-Nearest Neighbour (K-NN) and Random Forest were the two supervised machine learning techniques that were utilised. The Random Forest method yields an accuracy of 81.967% in predictions, whereas K-Nearest Neighbour (K-NN) yields an accuracy of 86.885% [16]

Manjari Gupta, Shrinkhala Yadav, and Vijeta Sharma (2020) developed a model utilizing. supervised learning techniques, including Naive Bayes, Decision Tree, K-Nearest Neighbour, and Random Forest algorithms, to analyze several heart disease-related variables. The UCI Cleveland Heart Disease Repository database is the source of the dataset. There are 76 characteristics and 303 occurrences in the collection. Merely 14 characteristics out of these 76 were included for testing, which is crucial to show how various algorithms work. Its goal is to show patients' chances of developing heart disease. The outcomes demonstrate that the K-nearest neighbour method yielded the best accuracy [17].

Ali N. Hasan, Thokozani Shongwe, and Vikash Rameshar. An interim study and examination were carried out in 2021 at various evaluation stages of heart disease prediction using various machine learning techniques, including artificial neural networks, decision trees, Naive Bayes, random

forests, logistic regression, support vector machines, and XG Boost. The outcomes demonstrated that the random forest strategy performed better than the other methods and had a 95% prediction accuracy [18].

M. Kavitha, Y. Rohith Sai, G. Gnaneswar, R. Dinesh, and R. Sai SurajUtilising the Cleveland Heart Disease Dataset in 2021, computer learning techniques Random Forest and Decision Tree were employed, along with data mining techniques including regression and classification. The machine-learning model is the foundation of the new technology. Three machine learning algorithms—Random Forest, Decision Tree, and Hybrid Model—were applied in the implementation (hybrid random forest and decision tree). Based on experimental data, the hybrid model was able to predict heart disease with an accuracy level of 88.7% [19].

In 2022, Gautam Srivastava and Senthilkumar Mohan employed naive Bayes, decision trees, genetic algorithms, and the closest neighbour method. Numerous readouts are created by combining two or more approaches to create a unique prediction model. Hybrid approaches are the broad term used to describe these novel mixed procedures. They introduced a novel approach that utilizes machine learning techniques to identify significant traits, thereby improving the accuracy of cardiovascular disease prediction. The prediction model for heart disease using a hybrid random forest with a linear model (HRFLM) yields an improved performance level with an accuracy level of 88.7% [20].

M. Bhatt, Tarang Ghetia, ORCID, Parth Patel. They used models including XGBoost (XGB), Random Forest (RF), Decision Tree Classifier (DT), and Multilayer Perceptron (MP) in 2023. To improve the outcome, GridSearchCV was used to modify the applied model's parameters. Using an actual dataset with 70,000 occurrences on Kaggle, the suggested model is evaluated. After training on 80:20 split data, the models' accuracy was as follows: The results for Decision Tree, XGBoost, Random Forest, Random Forest, and Perceptron Multilayer are 86.37% (with cross-validation) and 86.53% (without cross-validation), 86.28% (with cross-validation) and 86.94% (without cross-validation), and 86.77% (with cross-validation) and 87.02% (without cross-validation). The recommended models' AUC (area under the curve) values are as follows: Decision tree: 0.94; multilayer perceptron: 0.95; XGBoost: 0.95; random forest: 0.95. They concluded that the multilayer perceptron with cross-validation outperformed all other methods in terms of accuracy. Its accuracy was highest at 87.28% [21].

## 3. Methodology

### 3.1 Dataset Description

This study utilized the Cleveland Heart Disease dataset, a frequently used dataset for predicting heart disease [22]. The Kaggle Machine Learning repository is where you can find the Cleveland dataset. In 1988, the Cleveland Clinic Foundation conducted a health research study using the

Cleveland dataset. 1,026 individuals had 76 distinct traits listed in the dataset's first edition. Only 14 of these features—including the target category feature are known to be used by the majority of researchers. These characteristics include blood pressure, cholesterol, blood sugar, age, gender, and numerous other health indicators. Since there are 1025 samples in the data set, 499 of them come from people without heart disease and 526 from those who do, meaning that the data for the target variable we wish to predict is balanced, as Figure 1 illustrates.

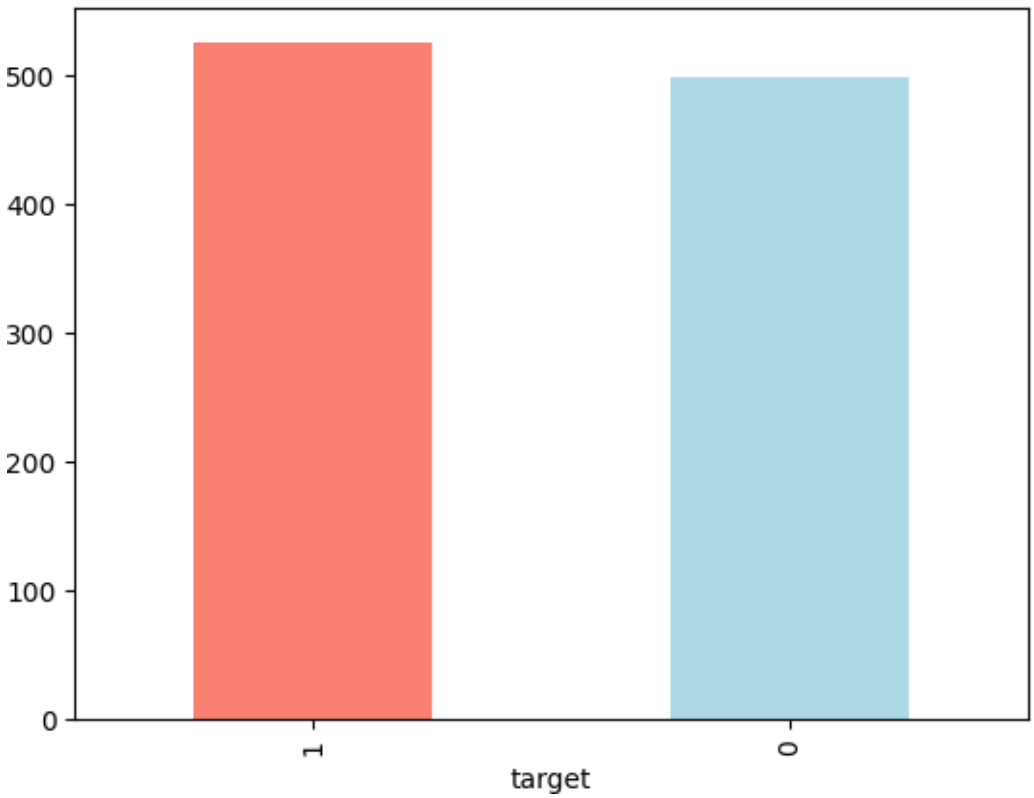



Figure 1: The distribution of classes in dataset

The closer the correlation value is to zero, the weaker the correlation between the feature and the target variable. The closer the correlation value is to -1 or 1, the stronger the correlation. As shown in Figure 2.

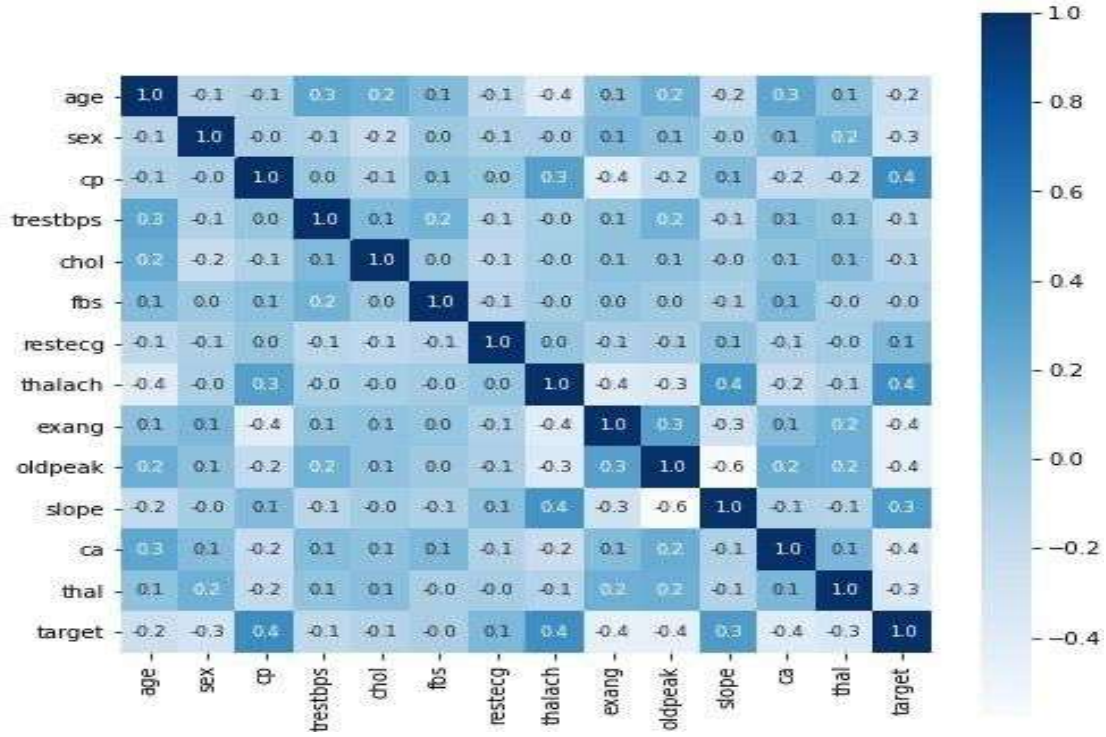



**Figure 2: The amount of correlation between data characteristics**

Fbs is the feature least closely related to the target variable. All other features correlate with the target variable to varying degrees. As shown in Figure 3.

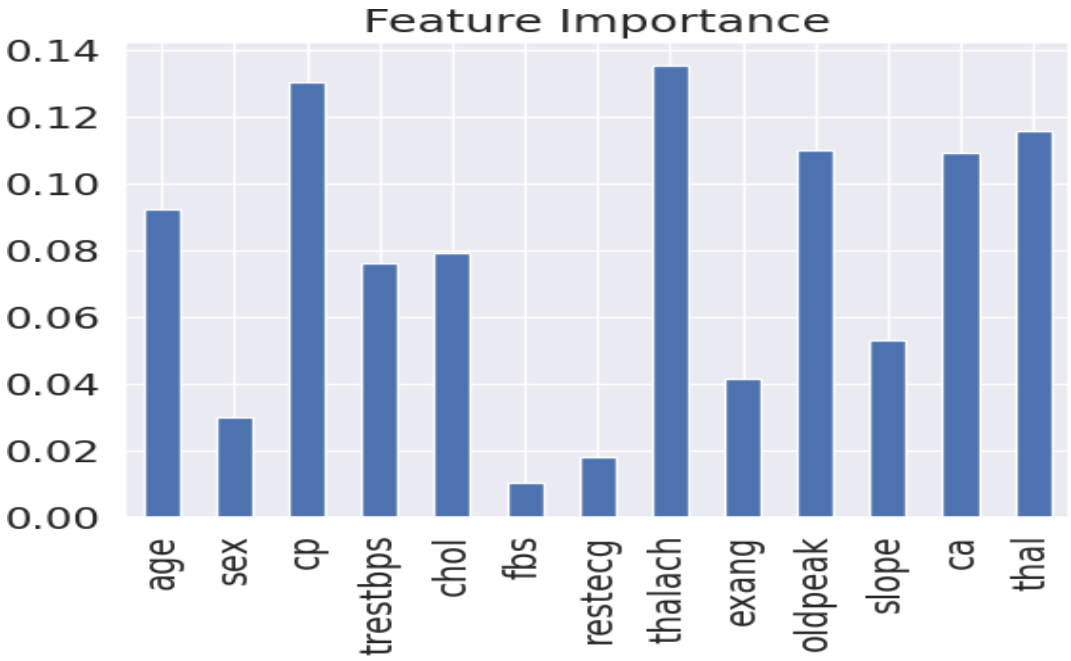



**Figure 3: The effect of data characteristics on heart disease**

## 3.2 Machine Learning Models for Classification

The application of computer algorithms that can be trained to perform a specific task using sample data is known as machine learning. Supervised learning techniques are frequently used to construct machine learning models for use in applications like illness prediction. Accurate labelling of training samples is necessary for supervised learning. The output, in its most basic form, is a binary variable that has a value of 0 for healthy individuals and 1 for patients. It is advised to employ balanced training data from both healthy and ill people in order to produce reliable machine learning models [23].

### 3.2.1 K–Nearest Neighbor (KNN)

It is one of the supervised machine learning algorithms used in the field of classification and regression. The KNN algorithm is considered one of the most important and simplest supervised machine learning algorithms. The KNN algorithm is also one of the algorithms used in classification and prediction tasks. It can also handle anomalous data or outliers with high efficiency. The basic idea behind the KNN algorithm is to determine the class or value of a new sample based on its location from the closest samples of the training dataset.

This algorithm's basic premise is to compute the Euclidean distance between two locations; the algorithm's name comes from the notion that the less the

distance between two points, the more likely it is that they are related. The number of samples represented by the letter K that will be used to classify a location according to how far away it is from its neighbours. K is their number. This indicates that the method calculates the separation between the target point and the three points that are closest to it, assuming that K = 3. In order to classify new data on a given scale based on the output corresponding to the nearest neighbor points, a target point will be classified as belonging to group A if the two nearest points are in group A and the third point alone is in group B. It is necessary to store all currently available data with the corresponding output. This way, K, which expresses the number of points, can be found experimentally by comparing the training error with the test error and determining the point at which the two lines become close, or in a state of balance. At that point, we have reached the ideal value of K. One of the most crucial algorithms for aiding data scientists in finding anomalous points is this one.

The advantages of the KNN algorithm include its lack of a data training process; instead, the data itself serves as a model that will be a reference for future predictions. Consequently, it is highly efficient in terms of time and in improving random models using the available data. The distance between various locations based on various feature data is all that has to be determined for the closest neighbour method to be implemented, and this distance may be calculated using the Manhattan or Euclidean distance functions. Since

there is no data training process, new data can be added at any time because it will not affect the performance of the model. The most popular applications of the closest neighbours technique are for regression and classification. It is a type of supervised learning algorithm. This approach is quite flexible and may also be used to resample datasets and impute missing magnitudes. As the name (Nearest Neighbors) suggests, it considers the K nearest neighbors (data points) to predict the class or continuous value of the new data point [24].

### 3.2.2 Chi-squared Automatic Interaction Detector (CHAID)

One of the most popular statistical techniques for supervised learning in decision tree building, since statistician Cass published it in the late 1970s, is the CHAID algorithm. CHAID refers to the repetitive, automated process of creating trees using Pearson's Chi-square statistic and the corresponding p-value. The CHAID method, a multivariate dependency approach, is primarily used to identify correlations between numerous independent factors and the dependent variable specifically, the chi-square analysis data point. As a result, the technique maximises the dependent variable's variance across groups while minimising its differences among groups [25].

At each stage of the algorithm's execution, the CHAID technique essentially tests hypotheses on the independence of two variables. The way the test is conducted and the findings are formulated follows the same logic as

traditional statistical hypothesis testing, with the exception that the software algorithm is supported. The decision tree algorithm CHAID is the earliest in known history. In 1980, Gordon V. Cass raised it. Then, in 1984, CART was discovered, in 1986, ID3 was suggested, and in 1993, C4.5 was declared. Chi-square Automatic Reaction Detection is what the acronym stands for. Chi-square is a metric used to determine a feature's significance. The statistical significance increases with increasing value. CHAID creates decision trees for categorization issues, just as other tools do.

### 3.2.3 C&R TREE

An example of a decision support tool is a decision tree, which is a diagram similar to a tree that shows decisions and their anticipated outcomes, including the likelihood of reaching specific outcomes, and is a one-way diagram to display the algorithm. Decision trees are generally used in operations research, especially in decision analysis, to help determine the strategy that will lead to achieving a goal. Similar in Form to a flowchart, a decision tree has inner nodes that represent "tests" for various qualities (e.g., whether a coin comes up heads or tails), branches that reflect test results, and final nodes that indicate the choice taken after taking into consideration all the attributes. The categorization rules are represented by the route from the root to the last node [26].

Although decision trees are a supervised learning method that may be used to solve regression and classification problems equally well, they are frequently chosen for usage in the former. This classifier is tree-structured, with core nodes standing in for data set attributes, branches for decision rules, and leaf nodes for each outcome.

## 4. **Result**

All experiments presented in this article were performed in MATLAB using the Deep Learning Toolkit. To identify ECG data, researchers usually utilize TensorFlow, an extensive open-source machine learning framework, together with Python and the Keras Deep Learning module. For software development and implementation, Python is the ideal tool; nevertheless, MATLAB has several capabilities and functionalities that make it a more advantageous choice during the research stage. Matlab offers a flexible two-way interface with Python and several other programming languages. This enables multiple teams to collaborate by utilizing MATLAB methods in information systems and programs. In addition, MATLAB makes it simple to create and test algorithms, write computational programs, debug swiftly, access a vast library of pre-built algorithms, create apps with a graphical user interface, and much more.

The important algorithms were tested to obtain the most accurate and best algorithm for detecting and predicting heart disease. When compared to the

other machine learning algorithms that were evaluated, the CHAID algorithm performed better in predicting heart disease, with an accuracy of 96.74%. The C&R TREE method came in second with an accuracy of 96.09%, while K-Nearest Neighbour came in third with a performance of 75.66%, as seen in Figure 4.

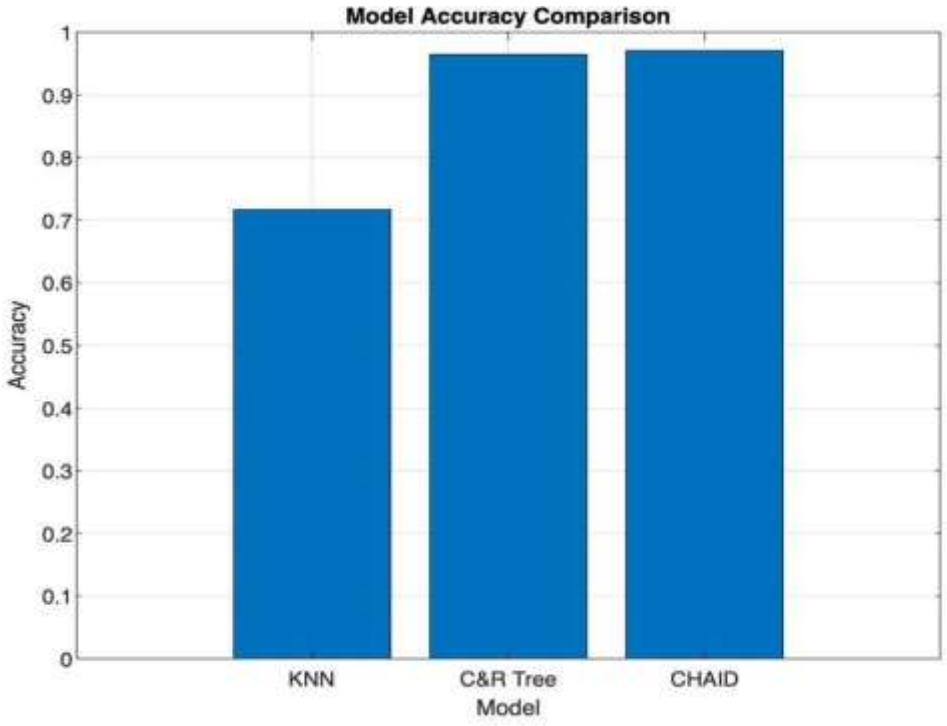



**Figure 4**: **The accuracy of the algorithms for predicting heart disease**.

## 5. Conclusion and Future Work

This study aims to develop a machine-learning model. Selecting an exact algorithm with the capacity to forecast cardiac disease properly was the main objective of this work. The algorithm can be a valuable tool in creating focused diagnostic and treatment plans for the condition.

To address these challenges and identify the most effective algorithms for heart disease prediction, we utilized the Cleveland Heart Disease dataset. We employed three machine learning models KNN, CHAID, and C&R tree to compare their performance. Notably, the CHAID model achieved the highest performance accuracy of 96.74%. The C&R TREE method ranked second with an accuracy of 96.09%, followed by K-Nearest Neighbour with a performance of 75.66%. This indicates that CHAID is the optimum method for classifying heart disease.

In future work, suggest that future researchers use another machine learning model on this dataset and expand the work by using another dataset with CHAID to ensure Form performance.

**References**

[1] WHO. (2017). Cardiovascular Diseases. World Health Organization. Retrieved from http://www.who.int/mediacentre/factsheets/fs317/en

[2] Liu, R., Ren, C., Fu, M., Chu, Z., & Guo, J. (2022). Platelet detection based on improved YOLO_v3. Cyborg Bionic Systems, 2022, 9780569.

[3] L. S. Ashour et al., "Non-overlapping Patch-Based Pre-trained CNN for Breast Cancer Classification," Iraqi Journal for Computer Science and Mathematics, vol. 6, no. 2, Jun. 2025, doi: 10.52866/2788-7421.1271.

[4] A. Abed Mohammed, P. Sumari, and kassem Attabi, "Hybrid K-means and Principal Component Analysis (PCA) for Diabetes Prediction,"

International Journal of Computing and Digital Systems, vol. 15, no. 1, pp. 1719–1728, Jun. 2024, doi: 10.12785/ijcds/1501121.

[5] Mohamed, A.A.A., Hançerliog˘ ullari, A., Rahebi, J., Ray, M.K., & Roy, S. (2023). Colon disease diagnosis with convolutional neural network and grasshopper optimization algorithm. Diagnostics, 13(1728).

[6] Palaniappan, S., & Awang, R. (2008). Intelligent heart disease prediction system using data mining techniques. In IEEE/ACS International Conference on Computer Systems and Applications (pp. 108–115). Doha, Qatar.

[7] Ghumbre, S., Patil, C., & Ghatol, A. (2011). Heart disease diagnosis using support vector machine. In International Conference on Computer Science and Information Technology (pp. 84–88). Pattaya, Thailand: Planetary Scientific Research Centre.

[8] Amma, N. (2012). Cardiovascular disease prediction system using genetic algorithm and neural network. In International Conference on Computing, Communication and Applications (pp. 1–5). Dindigul, Tamilnadu, India: IEEE.

[9] Amin, S., Agarwal, K., & Beg, R. (2013). Genetic neural network-based data mining in prediction of heart disease using risk factors. In IEEE Conference on Information Communication Technologies (pp. 1227–1231). Thuckalay, Tamil Nadu, India.

[10] Waghulde, N., & Patil, N. (2014). Genetic neural approach for heart disease prediction. International Journal of Advanced Computer Research, 4(3), 778.

[11] Masethe, H., & Masethe, M. (2014). Prediction of heart disease using classification algorithms. In Proceedings of the World Congress on Engineering and Computer Science (pp. 22–24). San Francisco, USA: International Association of Engineers (IAENG).

[12] Khateeb, N., & Usman, M. (2017). Efficient heart disease prediction system using k-nearest neighbor classification technique. In Proceedings of the International Conference on Big Data and Internet of Things (BDIOT) (pp. 21–26). New York, NY, USA: ACM.

[13] Pouriyeh, S., et al. (2017). A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease. In Proceedings of IEEE Symposium on Computers and Communications (ISCC) (pp. 204–207). Heraklion, Greece: IEEE.

[14] Zriqat, I., Altamimi, A., & Azzeh, M. (2017). A comparative study for predicting heart diseases using data mining classification methods. International Journal of Computer Science and Information Security (IJCSIS), 14(12), 868–879.

[15] Liu, X., et al. (2017). A hybrid classification system for heart disease diagnosis. Computational and Mathematical Methods in Medicine, 2017, 1–11.

[16] Garg, A., Sharma, B., & Khan, R. (2020). Heart disease prediction using machine learning techniques.

[17] Sharma, V., Yadav, S., & Gupta, M. (2020). Heart disease prediction using machine learning techniques.

[18] Williams, R., Shongwe, T., Hasan, A. N., & Rameshar, V. (2021). Heart disease prediction using machine learning techniques.

[19] Kavitha, M., Gnaneswar, G., Dinesh, R., Sai, Y. R., & Suraj, R. S. (2021). Heart disease prediction using hybrid machine learning model.

[20] Thirumalai, C., Srivastava, G., & Mohan, S. (2022). Effective heart disease prediction using hybrid machine learning techniques.

[21] Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective heart disease prediction using machine learning techniques.

[22] UCI Machine Learning Repository. (n.d.). Heart Disease Data Set. Retrieved from http://archive.ics.uci.edu/ml/datasets/Heart+Disease

[23] Hazra, A., Mandal, S., Gupta, A., & Mukherjee, A. (2017). Heart disease diagnosis and prediction using machine learning and data mining techniques: A review. Advances in Computational Sciences and Technology, 10, 2137–2159.

[24] Imandoust, S., Bafandeh, S., & Bolandraftar, M. (2013). Application of K-Nearest Neighbor (KNN) approach for predicting economic events: Theoretical background. Journal of Engineering Research and Applications, 3(5), 605–610.

[25] Rokach, L., & Maimon, O. (2008). Data mining with decision trees: Theory and applications. New Jersey, USA: World Scientific.

[26] Patel, J., Upadhyay, P., & Patel, D. (2016). Heart disease prediction using machine learning and data mining techniques. Journal of Computer Science & Electronics, 7, 129–137.

# Sentinel–Wear: A Provenance–First, Tiered Edge–Federated Framework for Trustworthy AI Security in Consumer Wearables and Remote Patient Monitoring

**Muthana Naser Hussein**[1]

[1]Dhi Qar Education Directorate ,Ministry of Education, Correspondence, muthananaser6@gmail.com

**مثنى ناصر حسين حسن السايول – المديرية العامة لتربية ذي قار، قسم تربية قلعة سكر**

**Abstract**. Artificial intelligence is increasingly proposed as the answer to Internet of Things (IoT) security, and healthcare is routinely cited as one of its highest-stakes proving grounds. Yet most AI-for-IoT-security work treats healthcare wearables and remote patient monitoring (RPM) as just another resource-constrained IoT segment, optimizing classifier accuracy against generic attack datasets. This paper argues that this framing misses the failure modes that actually matter in healthcare wearables: adversarial manipulation at the physiological-signal level, alarm fatigue caused by security and clinical alerts that lack provenance, and a regulatory and liability structure — FDA premarket cybersecurity rules, HIPAA, and clinician trust — that most technical proposals treat as an afterthought rather than a design constraint. We propose Sentinel-Wear, a three-tier architecture (on-device

signal-integrity attestation, edge/gateway federated learning with robust aggregation and differential privacy, and cloud-side explainable adjudication) organized around a single governing principle: every security-relevant decision must carry an auditable evidence trail from raw sensor sample to clinician-facing alert. We detail the design of each tier, walk through a concrete spoofing scenario, and lay out an evaluation methodology grounded in figures reported in the empirical literature on adversarial attacks against physiological classifiers, federated learning with differential privacy, TinyML on wearable hardware, and clinical alarm fatigue. We close by discussing the framework's limitations and the open engineering and policy questions it does not resolve.

**Keywords**: healthcare IoT, wearable security, remote patient monitoring, adversarial machine learning, federated learning, differential privacy, explainable AI, alarm fatigue, medical device cybersecurity

## 1. **Introduction**

Consumer and clinical-grade wearables — smartwatches, patches, continuous glucose monitors, wearable ECG and pulse-oximetry devices — have become a primary channel through which physiological data leaves the body and enters a decision pipeline that may include a phone app, a cloud model, and eventually a clinician. Recent surveys of this space describe rapid growth in AI-enabled diagnostic and monitoring capability, from

cardiovascular risk prediction to fall and disability detection to multi-modal monitoring in oncology care [2, 3, 4, 5, 6, 7, 8]. The same surveys, and the broader IoT security literature, treat "security" for these devices largely as a special case of general IoT security: detect network intrusions, harden protocols, apply access control.

That framing is incomplete for two reasons specific to this domain. First, the attack surface that matters most for a wearable is not only the network stack but the physiological signal itself. A growing body of adversarial-machine-learning research shows that small, often imperceptible perturbations to ECG, EEG, PPG, and radar-derived vital-sign data can flip a classifier's output — for example, adversarial perturbations to single-lead ECG have been shown to fool arrhythmia classifiers with a 74% success rate while remaining invisible to trained cardiac specialists [9], and both contact-based (PPG, EEG) and contactless (Doppler radar, Wi-Fi) vital-sign sensing pipelines have been shown to be spoofable or adversarially evadable [9, 10, 11, 12]. A defense that only watches network traffic never sees this class of attack. Second, the consequence of a false or missed alert in this domain is not a dropped packet but a patient-safety event or a trust failure: hospital and remote-monitoring alarm systems are documented to be 80–99% non-actionable in several studies [23, 24, 25, 26, 27], and this alarm burden is independently linked to slower clinician response and missed true events [23, 25]. An AI security layer that adds its own undifferentiated stream of alerts

on top of an already saturated clinical channel is not obviously a net safety improvement, even if its detection accuracy looks good in isolation.

A third gap is regulatory. Since 2023, the U.S. Food and Drug Administration has required a Software Bill of Materials (SBOM) and a documented cybersecurity risk management process for "cyber devices" — network-connected medical devices — as a condition of premarket approval [29, 30]. Very little of the AI-for-IoT-security literature, including recent surveys that explicitly call for cross-domain and policy-aware frameworks [1], operationalizes what that means for the design of the AI components themselves: which model updates are traceable to which data source, which detection decisions are auditable, and which privacy budget was spent on which cohort.

This paper makes three contributions:

1. We articulate a threat model for AI-enabled healthcare wearables and RPM that separates physiological-signal-level attacks, federated-learning-pipeline attacks, and alert-channel attacks — a decomposition that existing general-purpose IoT threat models do not make explicit (Section 3).

2. We propose **Sentinel-Wear**, a three-tier framework whose organizing principle is *provenance-first design*: every alert or model update carries a structured, auditable record of which tier produced it, what

evidence supported it, and what privacy/robustness guarantees applied (Section 4).

3. We lay out a concrete, falsifiable evaluation plan for the framework, grounded in effect sizes reported in the existing empirical literature on adversarial robustness, federated differential privacy, TinyML energy budgets, and alarm-fatigue reduction, and we are explicit about what remains unproven (Section 5).

We position this work as a design and argument paper, not an implementation report: no component of Sentinel-Wear has been built or benchmarked by the authors. Section 6 discusses this limitation directly, along with the engineering and ethical questions the framework leaves open.

## 2. Background and Related Work

### 2.1 AI in wearable health monitoring and RPM

AI-enabled wearables now span disease-specific monitoring (cardiovascular risk, cancer-therapy side effects, disability and fall detection) and general-purpose vital-sign tracking, typically combining on-body sensors with cloud- or edge-based classifiers for anomaly detection and early warning [2, 3, 4, 5, 6, 7, 8]. A representative example is a cancer-care remote-monitoring platform that aggregated over two million data points across roughly six thousand patient-days from fifty patients, using thirty extracted features across multiple modalities to predict adverse events during systemic

therapy [8]. Systematic reviews of AI in wearable health monitoring emphasize that closed-loop systems — continuous sensing, analysis, and alerting — are becoming the default architecture, and that deep learning models such as CNNs and LSTMs are the dominant modeling choice for physiological time series [6]. None of this literature centers security; when it is discussed, it is typically framed as data confidentiality in transit and at rest, rather than as robustness of the inference itself.

2.2 Physiological-signal-level adversarial threats

A distinct and directly relevant literature examines adversarial machine learning against physiological classifiers. A systematic review of adversarial attacks and defenses in physiological computing catalogs both evasion and poisoning attacks across ECG, EEG, PPG, blood pressure, $SpO_2$, and other modalities, and documents that white-box evasion attacks against an ECG-based arrhythmia classifier achieved a 74% success rate at flipping the diagnosis of correctly classified recordings, while producing waveforms rated as unaltered by board-certified specialists [9]. The same review reports that universal adversarial perturbations can be computed to fool an EEG-based seizure detector into missing most seizure events, and surveys a broader set of studies covering ten distinct vital signs under both targeted/non-targeted and white-box/black-box threat models [9, 11]. Related work specifically targeting open-source healthcare models finds that transparency and accessibility — properties that are otherwise desirable — increase

susceptibility to adversarial and model-inversion attacks, and identifies adversarial training, differential privacy, and model sanitization as partial (not complete) mitigations [11].

This threat extends beyond contact sensors. A 2025 survey of radio-frequency sensing security documents spoofing attacks on Doppler-radar vital-sign monitors using commodity RF hardware (phase shifters, signal generators) that successfully forge heartbeat and respiration waveforms, and separately documents that standard white-box adversarial methods (FGSM, PGD, MIM) significantly degrade the accuracy of Wi-Fi-based sleep apnea detectors [12]. These findings matter for the framework proposed here because they establish that spoofing is not a hypothetical: it has been demonstrated against the exact signal types (ECG, EEG, radar/Wi-Fi vital signs) that consumer and clinical wearables rely on, using threat models ranging from full white-box access to fully black-box, sensor-level interference.

### 2.3 Data-quality attacks framed as security events

A related but conceptually distinct threat is ordinary sensor unreliability and denial-of-service framed as a security event: a review of machine-learning-based anomaly detection for IoT healthcare describes denial-of-service attacks that flood a wearable with spurious traffic to block legitimate transmission, and spoofing attacks in which an attacker's device impersonates

a legitimate one to gain access to health data [10]. This is a lower-sophistication but higher-frequency threat class than the adversarial-perturbation attacks above, and any practical framework has to handle both without conflating them.

### 2.4 Privacy-preserving distributed learning for health IoT

Federated learning (FL) is widely proposed as the way to train models across wearables or hospitals without centralizing raw physiological data, and a large recent literature specifically targets healthcare and IoMT applications [13, 14, 15, 16, 17, 18, 19, 20, 21]. A 2024–2025 review focused specifically on FL in smart healthcare highlights that its main open security risks are adversarial attacks, data poisoning, and model inversion, and that differential privacy (DP) and secure multiparty computation are the leading mitigations, with a known trade-off against model accuracy [13]. Several proposed systems combine FL with DP explicitly for wearable or medical-IoT data — for instance, an adaptive differential-privacy federated architecture for medical IoT is designed around a hierarchical structure in which edge servers mediate between devices and a central aggregator specifically to support HIPAA-relevant compliance goals [21], and a differential-privacy federated learning model for wearable biomedical monitoring reports that adding DP and secure-computation layers protects data but introduces an accuracy cost that must be tuned per deployment [16, 19]. This confirms that the privacy/accuracy trade-off is a first-class design variable, not a detail to be

resolved later — which is one reason we treat the privacy budget as an object with its own audit trail in Section 4.

### 2.5 Alarm fatigue and explainable false-alarm suppression

A substantial patient-safety literature, independent of the security literature, documents that clinical monitoring systems generate overwhelmingly non-actionable alarms — estimates across multiple systematic reviews and algorithm-validation studies converge on roughly 72–99% of alarms being technically false or clinically insignificant [23, 24, 25, 26, 27]. This is directly relevant to AI security design because a naive anomaly- or intrusion-detector bolted onto a wearable adds to this same alert stream, and the clinical-informatics literature treats alarm volume itself as a patient-safety hazard, associated with desensitization, delayed response, and missed true events [24, 25]. Recent work has begun applying explainable AI specifically to alarm triage: a two-stage early-warning framework uses a first-stage predictive model followed by a second stage that applies SHAP-based feature attribution, filtered through domain knowledge, to suppress low-confidence alerts, reporting an 11.5-percentage-point increase in positive predictive value on external validation [28]. Closer to the concerns of this paper, a provenance-guided, multi-agent false-positive suppression system for RPM has recently been proposed that explicitly reasons about the *source* of an anomalous reading — not just its statistical signature — before triaging it to a clinician [22]. Sentinel-Wear shares this provenance-first instinct but

extends it upstream: rather than suppressing false positives after a signal has already been classified as anomalous, our device tier attempts to establish signal integrity *before* any anomaly classifier runs, so that a spoofed or corrupted signal is flagged as such rather than being fed into the same detection pipeline as genuine physiological data.

### 2.6 Regulatory and supply-chain context

Since March 2023, FDA guidance under Section 524B of the Food, Drug, and Cosmetic Act has required manufacturers of network-connected "cyber devices" to submit a Software Bill of Materials — a machine-readable inventory of proprietary, open-source, and third-party software components — along with a documented secure product development process and vulnerability-management plan, as a condition of premarket submission [29, 30]. This applies to many consumer-adjacent wearables that make clinical claims, and it is a structural constraint that a security *architecture*, not just a security *policy document*, has to be able to satisfy: components need to be individually identifiable, versioned, and attributable to a vendor, and the system needs a way to prove which model version and which data sources produced a given alert.

### 2.7 Gap and positioning

Table 1 summarizes how the strands above relate to one another and to this paper. No single line of existing work combines (a) explicit handling of

physiological-signal-level adversarial threats, (b) privacy-preserving distributed learning with an accountable privacy budget, (c) alarm-fatigue-aware, provenance-based alerting, and (d) a structural answer to SBOM/audit obligations, inside one architecture that also respects the compute and energy budget of a wearable. General IoT security surveys, including the one whose title most closely resembles our topic [1], address several of these individually — adversarial robustness, federated learning, XAI, and regulatory alignment are all discussed — but as a catalog of techniques rather than as an integrated architecture with a stated design principle, and without treating the physiological signal itself, as opposed to network traffic, as the primary threatened asset. That is the gap Sentinel-Wear is designed to fill.

**Table 1. Positioning of Sentinel-Wear relative to related work**

| Concern | Representative prior work | Sentinel-Wear's treatment |
|---|---|---|
| Physiological-signal adversarial evasion/spoofing | [9, 11, 12] (attack demonstrations, largely without a deployed countermeasure) | Dedicated device-tier signal-integrity attestation *before* classification (Sec. 4.3.1) |
| Federated learning privacy/robustness | [13, 14, 15, 16, 17, 18, 19, 20, 21] (FL + DP designs, generally not adversary-aware at the aggregation step) | Robust aggregation + DP with an explicit, auditable privacy ledger (Sec. 4.3.2) |

| | | |
|---|---|---|
| Alarm fatigue / false-alarm suppression | [22, 23, 24, 25, 26, 27, 28] (post-hoc triage of already-generated alerts) | Provenance carried from signal-integrity check through classification into triage, not reconstructed after the fact (Sec. 4.3.3) |
| Regulatory/supply-chain accountability | [29, 30] (compliance requirements, not an architecture) | SBOM-linked component registry treated as a first-class system component (Sec. 4.3.4) |
| Resource constraints of wearable hardware | [31, 32, 33, 34] (TinyML feasibility studies, not security-specific) | Used as the feasibility basis for the device tier's power/memory budget (Sec. 4.2, Sec. 5) |

## 3. **Threat Model**

We define the asset under protection as the *integrity and provenance of the physiological data pipeline*, from sensor to clinician-facing decision — not merely the confidentiality of data in transit, which is the more conventional IoT security framing. We assume an attacker may have one or more of the following capabilities, drawn from the demonstrated attacks surveyed in Section 2: (i) physical or RF proximity to the wearable or its wireless link, sufficient to inject spoofed sensor readings [10, 12]; (ii) white-box or black-box query access to a deployed classifier, sufficient to craft adversarial

perturbations [9, 11]; (iii) control of one or more participants in a federated learning cohort, sufficient to poison model updates [11, 13]; and (iv) no special access at all, but the ability to exploit an unmanaged alert channel to degrade clinician trust over time (an availability/trust attack rather than a confidentiality attack).

We organize the resulting threats into four classes:

**T1 — Physiological signal spoofing and adversarial evasion**. An attacker forges or perturbs a physiological signal so that a downstream classifier produces an incorrect diagnosis or misses a true event. This includes both sensor-level spoofing (RF/radar signal injection, replayed PPG waveforms) and model-level adversarial perturbation (imperceptible input modification against a known or inferred classifier) [9, 10, 11, 12].

**T2 — Federated pipeline poisoning**. A compromised or malicious device in a federated cohort submits corrupted gradients or model weights to bias the shared model, degrade its accuracy for other users, or implant a backdoor [11, 13].

**T3 — Alert-channel abuse and alarm-fatigue induction**. Even without directly corrupting any signal or model, an attacker (or, more commonly, a poorly tuned defensive system) generates enough spurious alerts to desensitize clinicians or caregivers, creating a window in which a genuine event is missed. We treat this as a security-relevant threat, not only

a usability issue, because the empirical alarm-fatigue literature ties elevated alarm volume directly to missed true-positive events [24, 25].

**T4 — Supply-chain and firmware compromise**. A vulnerability in a third-party library, an unpatched firmware component, or an unauthorized model update compromises the device or the pipeline. This is the threat class the FDA's SBOM requirement is designed to make auditable after the fact [29, 30]; we argue it also needs to be *addressed structurally* at design time.

Table 2 summarizes these classes together with the tier of Sentinel-Wear primarily responsible for mitigating each.

**Table 2**. **Threat classes and primary mitigating tier**

| Threat | Primary mitigating tier | Mechanism |
|---|---|---|
| T1 — Signal spoofing / adversarial evasion | Device tier | Cross-modal plausibility and signal-integrity attestation prior to classification |
| T2 — Federated pipeline poisoning | Edge/gateway tier | Robust aggregation (outlier-resistant averaging) + differential privacy |
| T3 — Alert-channel abuse / alarm fatigue | Cloud/clinical tier | Provenance-tagged, evidence-bundled triage rather than raw alert forwarding |
| T4 — Supply-chain / firmware compromise | Cross-cutting (all tiers) | SBOM-linked component registry and update-attestation ledger |

## 4. **Proposed Framework**: **Sentinel-Wear**

## 4.1 Design principles

Sentinel-Wear is built around one governing principle — **provenance-first design**: no alert, model update, or automated action should reach a clinician or an aggregation step without a structured record of where it came from, what checks it passed, and what confidence and privacy guarantees apply to it. This principle is operationalized through four subordinate design commitments:

1. **Resource-tiered placement**. Checks are placed at the tier that can perform them cheaply and early, rather than centralizing all inference in the cloud. Signal-integrity checks belong on-device because they must run before an untrustworthy signal is allowed to influence anything downstream.

2. **Privacy-by-architecture**, **not privacy-by-policy**. Differential privacy and local data retention are structural properties of the edge tier, not configuration flags that can be silently disabled.

3. **Human-in-the-loop by default**, **full autonomy by exception**. The framework is designed to reduce clinician alert burden, not to replace clinical judgment; full closed-loop automated response (e.g., automatic care-team escalation) is reserved for a narrow set of high-confidence, high-severity cases.

4. **Auditability as a first-class output**. Every component's outputs are designed to be replayable against an SBOM-linked component registry, so that a regulator or hospital security team can reconstruct why a given decision was made.

4.2 Architecture overview

Sentinel-Wear has three tiers, summarized in Figure 1.

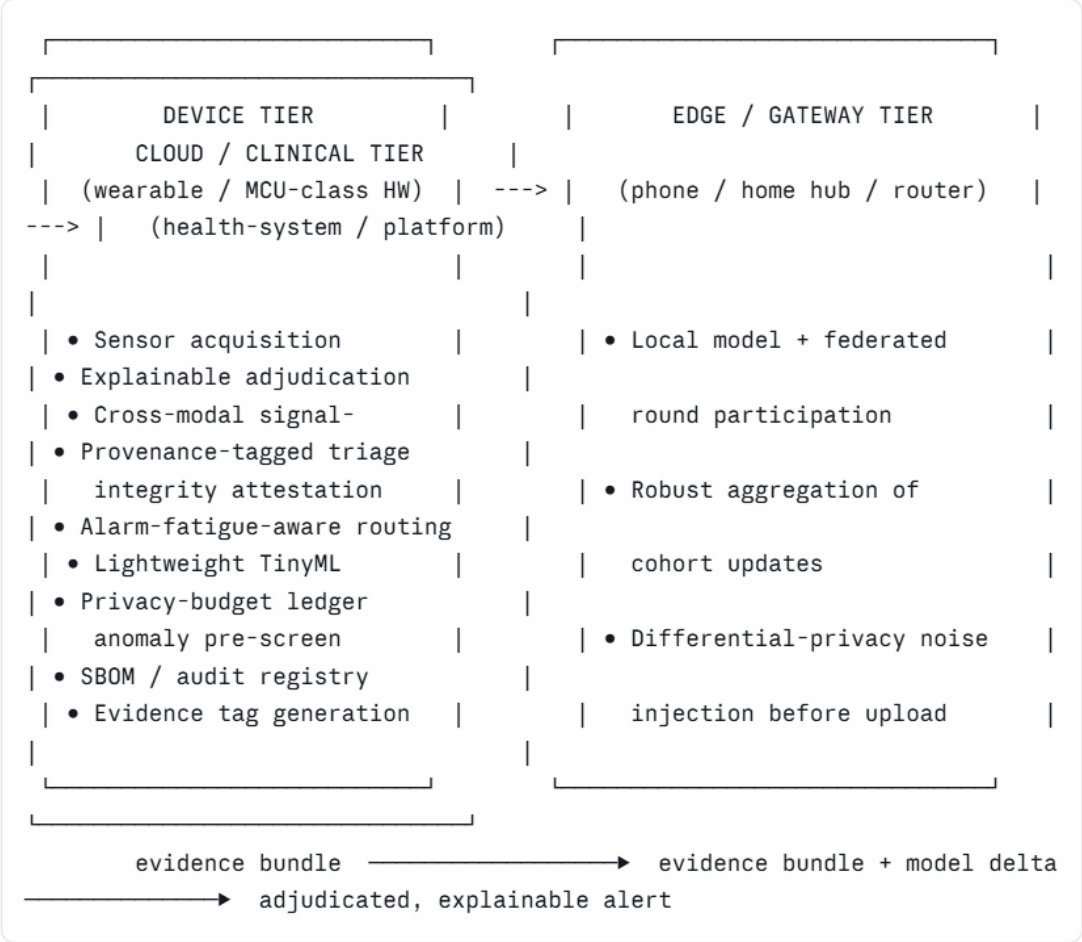



**Figure 1**. **Sentinel-Wear three-tier architecture**. Each arrow carries not just data or a model delta, but an accompanying evidence bundle: a

structured record of which checks were run, what they found, and what privacy operations were applied.

4.3 Component details

4.3.1 Device tier: signal-integrity attestation before classification

The device tier's job is to answer one question before any anomaly or diagnostic classifier runs: *is this plausibly a genuine physiological signal from this wearer, right now?* This is distinct from — and prior to — asking whether the signal looks clinically abnormal. We propose two lightweight, complementary checks, both designed to fit within the memory and power budgets demonstrated for TinyML deployments on wearable-class microcontrollers, where optimized on-device ECG anomaly models have been shown to run at roughly 0.024 mW with 92% accuracy [31], and autoencoder-based on-device arrhythmia detectors have been fit into roughly 180 KB with single-digit-millisecond inference latency [33]:

- **Cross-modal plausibility**. Independent sensor modalities on the same device (e.g., PPG-derived heart rate vs. accelerometer-derived motion, or ECG vs. impedance-based respiration) are checked for physiologically consistent covariation. A heart-rate spike with no corresponding motion or respiration change, or a waveform morphology inconsistent with any plausible cardiac cycle, is flagged as *integrity-suspect* rather than being passed to the diagnostic classifier as an ordinary anomaly.

- **Lightweight replay/injection screening**. Lightweight statistical checks (e.g., waveform periodicity and noise-floor consistency, in the spirit of the fully on-device Z-score anomaly detectors shown feasible on low-cost microcontrollers with kilobyte-scale memory footprints [32]) screen for the signatures of replayed or synthetically injected signals, without requiring cloud connectivity or heavy compute.

Both checks are intentionally cheap and imperfect — they are not expected to catch a sophisticated white-box adversarial perturbation crafted against the specific downstream classifier. Their purpose is to raise the cost of the *easy* attacks (sensor replay, RF spoofing, naive injection) essentially for free, and, critically, to attach an **integrity tag** to every reading that propagates downstream regardless of outcome. A reading that fails integrity screening is not discarded silently; it is forwarded with an explicit "integrity-suspect" tag, because a spoofing attempt is itself a security-relevant event that the clinical tier should be able to see.

4.3.2 Edge/gateway tier: robust, differentially private federated learning

The edge tier — realistically a smartphone or home hub with more compute and connectivity than the wearable itself — hosts the local model and participates in periodic federated learning rounds, following the general pattern established in the healthcare FL literature [13, 14, 15, 16, 17]. Two properties are added specifically to address T2 (federated pipeline poisoning):

- **Robust aggregation**. Rather than a plain federated average, the aggregator applies an outlier-resistant combination rule (e.g., coordinate-wise trimmed mean or norm-clipping before averaging) so that a small number of compromised or malfunctioning participants cannot disproportionately bias the global model — a mitigation consistent with what recent FL-security reviews identify as a necessary complement to privacy mechanisms, since privacy and robustness are shown to be separable concerns that are often conflated [13].

- **Auditable differential privacy**. Calibrated noise is added to model updates before they leave the edge tier, following the general DP-federated pattern used in prior wearable and medical-IoT proposals [16, 19, 21]. What we add is a **privacy ledger**: each participant's cumulative privacy budget expenditure is logged locally and summarized (not raw data, only budget metadata) to the cloud tier, so that an auditor can verify that a stated privacy guarantee was actually maintained over time rather than taking a one-time design claim on faith. This directly targets the regulatory-traceability gap identified in Section 2.6 and connects the accuracy/privacy trade-off documented in the literature [13, 16, 19] to a concrete, inspectable artifact rather than leaving it as an internal hyperparameter.

4.3.3 Cloud/clinical tier: provenance-tagged explainable adjudication

The cloud tier receives evidence bundles — not raw alerts — from the edge tier. Each bundle contains: the classifier's output and confidence, the device-tier integrity tag, a feature-attribution summary (in the spirit of the SHAP-based attribution used in recent tiered early-warning work [28]), and a reference to the exact model and data-processing version that produced it (tying back to the SBOM registry, Section 4.3.4). A triage layer — analogous in spirit to the provenance-guided suppression recently proposed for RPM false positives [22], but operating on a richer evidence bundle that already includes device-tier integrity information — uses this bundle to route the event into one of three categories:

1. **Suppressed with logged rationale** — low-confidence, integrity-clean readings consistent with known benign patterns are not surfaced to a clinician at all, but are retained in the audit log.

2. **Routed for clinician review** — the default path for genuine clinical anomalies, presented together with the evidence bundle so the clinician sees *why* the system flagged it, addressing the "black-box" trust problem documented in interpretability research on IoT and medical AI.

3. **Flagged as a security event** — readings tagged integrity-suspect at the device tier, or associated with anomalous federated-update patterns at the edge tier, are routed to a security/IT channel distinct from the clinical

alert channel, so that a spoofing attempt does not simply masquerade as one more clinical false alarm to be tuned away.

This three-way split is the mechanism by which Sentinel-Wear addresses T3: it does not merely try to reduce the *number* of alerts (as most alarm-fatigue interventions do), it changes *what kind of information* accompanies each alert that does get through, and it keeps security-relevant anomalies from being silently absorbed into the same statistical noise the clinical triage layer is trying to suppress.

4.3.4 Cross-cutting: SBOM-linked component registry

Every software and model component across all three tiers — firmware modules, ML model versions, third-party libraries — is registered with a version identifier, aligned with the FDA's SBOM expectations for cyber devices [29, 30]. This registry is what the evidence bundles in Section 4.3.3 reference, and what the privacy ledger in Section 4.3.2 reports against. Practically, this means an audit query such as "show every alert produced by model version X while it was exposed to cohort Y" or "confirm the differential-privacy budget spent on participant Z in the last quarter" is answerable by construction, rather than requiring forensic reconstruction after an incident.

4.4 Worked example: a spoofed heart-rate scenario

To make the architecture concrete, consider an attacker attempting to trigger a false "cardiac event" alert on a target's wearable using a portable RF signal generator, in the style of the radar-spoofing attacks documented in the RF sensing security literature [12]. Under Sentinel-Wear: (1) the device tier's cross-modal check notices that the injected heart-rate spike has no corresponding accelerometer or respiration correlate and tags the reading integrity-suspect; (2) the reading, together with its tag, is forwarded (not discarded) to the edge tier, which includes the tag unmodified in the evidence bundle it sends upstream; (3) the cloud tier's triage layer routes the event to the *security* channel rather than the clinical channel, because the integrity tag — not the clinical severity — determines the routing; (4) a clinician never sees a spurious cardiac alert, but a security operator does see a spoofing attempt, with a full evidence trail tying it to a specific device and time window. Contrast this with a conventional pipeline that only classifies clinical severity: the spoofed spike would either trigger an ordinary (and, over repeated attempts, fatigue-inducing) clinical alarm, or be missed entirely if the classifier's threshold happens not to be crossed.

## 5. **Evaluation Plan**

We have not implemented Sentinel-Wear; this section specifies how it should be evaluated, and what results from the component literature suggest about feasibility, so that the proposal is falsifiable rather than purely descriptive.

**Datasets**. Device-tier and classification benchmarking should draw on established physiological datasets already used in the cited component literature — e.g., MIT-BIH-style ECG arrhythmia corpora for the signal-integrity and classification components [31, 33], and multi-modal activity/wearable datasets such as those used in disability and motion-pattern detection work [2] for the cross-modal plausibility check. Federated and privacy components should be evaluated on a partitioned, non-IID split of such datasets across simulated participants, consistent with the cohort sizes used in comparable healthcare FL studies (e.g., federated cohorts on the order of tens of participants have been used to evaluate accuracy under DP in prior wearable-FL work [16, 19]).

**Metrics**.

- *Device tier*: true/false positive rate of the signal-integrity check against a labeled set of genuine vs. spoofed/replayed signals; added inference latency and energy draw, benchmarked against the sub-milliwatt, kilobyte-scale footprints reported for comparable on-device models [31, 32, 33] to confirm the integrity check does not itself blow the device's power budget.

- *Edge tier*: model accuracy under an increasing fraction of poisoned participants, with and without robust aggregation; accuracy degradation as a function of DP noise (privacy budget $\varepsilon$), benchmarked against the accuracy/privacy trade-off curves reported in prior DP-FL healthcare work

[13, 16, 19] as a sanity check that our robust-aggregation addition does not itself erase the privacy accounting.

- *Clinical tier*: reduction in clinician-facing alert volume and change in positive predictive value relative to a non-provenance-aware baseline, using the same class of metric used in the tiered early-warning literature, which reported an 11.5-point PPV improvement from attribution-based filtering alone [28]; a target worth testing is whether adding device-tier integrity tags produces an additional, separable improvement on top of that baseline.

- *Cross-cutting*: end-to-end time from a simulated spoofing event to correct security-channel routing, and completeness of the audit trail (percentage of alerts for which the full evidence bundle can be reconstructed from the SBOM-linked registry).

**Expected trade-offs**. Based on the component literature, we expect three tensions to dominate: (i) tighter differential-privacy budgets will reduce classification accuracy, consistent with prior findings in this space [13, 16, 19], and the privacy ledger is intended to make this trade-off visible and auditable rather than eliminate it; (ii) more conservative signal-integrity thresholds at the device tier will reduce false triggering of T1-style spoofing detection but risk suppressing genuine, unusual-but-real physiological events, so this threshold should be tuned against clinical, not just security, outcomes; (iii) robust aggregation methods that improve poisoning

resistance can themselves slightly reduce accuracy on legitimately heterogeneous (non-IID) participant data, a known general property of outlier-resistant aggregation that any implementation will need to characterize empirically rather than assume away.

## 6. **Discussion**

**What this paper does not show**. No component of Sentinel-Wear has been implemented or benchmarked by the authors; the evaluation plan in Section 5 is a specification, not a result. The framework's central claims — that provenance-first design meaningfully reduces alarm fatigue beyond attribution-based filtering alone, and that lightweight on-device integrity checks catch a useful fraction of real-world spoofing attempts without excessive engineering cost — are testable hypotheses, not established findings.

**Engineering complexity**. Provenance-first design adds real cost: every tier must generate, transmit, and store structured evidence, which is more engineering than a bare classifier. For a resource-constrained wearable, even the lightweight checks proposed in Section 4.3.1 compete for the same power and memory budget as the primary diagnostic function, and the added telemetry between tiers increases both bandwidth use and the attack surface of the evidence-bundle format itself, which would need its own integrity protection.

**Clinical workflow integration**. Routing security events to a distinct channel from clinical alerts (Section 4.3.3) presumes a receiving security/IT function exists and is staffed appropriately in the deploying health system or consumer platform — not a safe assumption for smaller clinics or purely consumer-facing wearable products, where "security team" may not exist as a distinct role.

**Equity and access**. Continuous on-device inference, federated participation, and cloud-side triage all assume a baseline of connectivity, device compute, and platform support that is not evenly distributed; a security architecture that only ships on premium hardware risks concentrating its safety benefits among users who already have the most access to care.

**Open regulatory questions**. SBOM and audit-trail support (Section 4.3.4) helps satisfy documentation requirements, but does not resolve harder questions this framework surfaces: who is liable when the triage layer suppresses an alert that should have reached a clinician, and how "explainable" an attribution needs to be to satisfy both a clinician's practical judgment and a regulator's audit standard are open questions this paper does not attempt to settle.

## 7. **Conclusion and Future Work**

We have argued that AI security for healthcare wearables and remote patient monitoring needs a framing distinct from general-purpose IoT security: the physiological signal itself, not only the network, is the primary attack surface; alarm fatigue is a security-relevant failure mode, not merely a usability concern; and regulatory accountability needs to be a structural property of the architecture. Sentinel-Wear operationalizes this through a three-tier, provenance-first design in which every alert and model update carries an auditable evidence trail from sensor to clinician. The framework's individual components — TinyML signal screening, robust federated aggregation with differentially private, auditable updates, and attribution-based clinical triage — each have support in the existing empirical literature, but their integration, and the specific claim that provenance-first design measurably reduces both missed spoofing attempts and clinician alert burden simultaneously, remains to be built and tested. The most immediate next step is a reference implementation of the device tier alone, benchmarked for power and latency on real wearable-class hardware against a labeled spoofing dataset, since that tier's feasibility is the load-bearing assumption for everything built on top of it.

## References

[1] Chen, Q., Tan, L., Tang, J., & Qu, X. (2025). AI-Enabled IoT Security: A Survey on Advances, Challenges, and Cross-Domain Collaborative Frameworks. In *Proceedings of the 2025 8th International Conference on*

*Computer Information Science and Artificial Intelligence (CISAI 2025)*. ACM. https://doi.org/10.1145/3773365.3773619

[2] Real-time e-health framework for efficient AI-driven disability monitoring using secured Internet of Medical Things. *Computing*, 107 (2025). https://link.springer.com/article/10.1007/s00607-025-01519-7

[3] IoT and AI for Remote Patient Monitoring and Diagnostics. ResearchGate (2024). https://www.researchgate.net/publication/385284519_IoT_and_AI_For_Remote_Patient_Monitoring_and_Diagnostics

[4] Enhancing remote patient monitoring with AI-driven IoMT and cloud computing technologies. *Scientific Reports* (2025). https://www.nature.com/articles/s41598-025-09727-z

[5] Advancing Remote and Continuous Cardiovascular Patient Monitoring through a Novel and Resource-Efficient IoT-Driven Framework. arXiv:2505.03409. https://arxiv.org/pdf/2505.03409

[6] Role of artificial intelligence in health monitoring using IoT based wearable sensors: A survey. *ScienceDirect* (2025). https://www.sciencedirect.com/science/article/abs/pii/S2542660525002744

[7] Smart Healthcare at Home: A Review of AI-Enabled Wearables and Diagnostics Through the Lens of the Pi-CON Methodology. *Sensors*, 25(19), 6067 (2025). https://www.mdpi.com/1424-8220/25/19/6067

[8] Multi-Modal AI for Remote Patient Monitoring in Cancer Care. arXiv:2512.00949. https://arxiv.org/pdf/2512.00949

[9] Adversarial Attacks and Defenses in Physiological Computing: A Systematic Review. arXiv:2102.02729; also published as *National Science Open*. https://arxiv.org/pdf/2102.02729

[10] Anomaly detection in IoT-based healthcare: machine learning for enhanced security. *PMC*. https://pmc.ncbi.nlm.nih.gov/articles/PMC10928137/

[11] Adversarial Attacks to Machine Learning-Based Smart Healthcare Systems. ResearchGate. https://www.researchgate.net/publication/348784294_Adversarial_Attacks_to_Machine_Learning-Based_Smart_Healthcare_Systems

[12] RF Sensing Security and Malicious Exploitation: A Comprehensive Survey. arXiv:2504.10969. https://arxiv.org/pdf/2504.10969

[13] Federated Learning in Smart Healthcare: A Comprehensive Review on Privacy, Security, and Predictive Analytics with IoT Integration. *Healthcare (MDPI)*, 12(24), 2587 (2024). https://www.mdpi.com/2227-9032/12/24/2587

[14] Federated Learning for Healthcare Data Privacy: A Case Study in Multi-Hospital Collaboration. Preprints.org (2025). https://www.preprints.org/manuscript/202509.0984

[15] Federated Learning for Privacy-Preserving Wearable Health Data Analytics. ResearchGate (2025). https://www.researchgate.net/publication/399898864_Federated_Learning_for_Privacy-Preserving_Wearable_Health_Data_Analytics

[16] Privacy-Preserving Federated Deep Learning for Wearable IoT-Based Biomedical Monitoring (DPFL-HIoT). ResearchGate. https://www.researchgate.net/publication/348250162_Privacy-Preserving_Federated_Deep_Learning_for_Wearable_IoT-Based_Biomedical_Monitoring

[17] Federated Deep Learning for Internet of Medical Things: A Comprehensive Survey of Architectures, Healthcare Applications, Privacy Preservation, Security, and Future Research Directions. *IJCT*. https://ijctjournal.org/federated-deep-learning-internet-medical-things/

[18] Linkage on Security, Privacy and Fairness in Federated Learning: New Balances and New Perspectives. arXiv:2406.10884. https://arxiv.org/pdf/2406.10884

[19] Inclusive, Differentially Private Federated Learning for Clinical Data. arXiv:2505.22108. https://arxiv.org/pdf/2505.22108

[20] A Privacy-Preserving Health Monitoring Framework Using Federated Learning on Wearable Sensor Data. *MDPI* (2025). https://www.mdpi.com/2673-4591/118/1/73

[21] Adaptive Differential Privacy in Federated Edge AI for Medical IoT: Energy-Efficient, HIPAA-Compliant Frameworks for Distributed Real-Time Diagnostics. *Journal of Information Systems Engineering and Management*, 10(34s) (2025). https://jisem-journal.com/index.php/journal/article/view/5773

[22] Veritas-RPM: Provenance-Guided Multi-Agent False Positive Suppression for Remote Patient Monitoring. arXiv:2604.16081. https://arxiv.org/pdf/2604.16081

[23] Reducing Alarm Fatigue Among Critical Care Practitioners Using AI Solutions. ResearchGate (2024). https://www.researchgate.net/publication/387398912_Reducing_Alarm_Fatigue_Among_Critical_Care_Practitioners_Using_AI_Solutions

[24] Artificial Intelligence for Reducing Alarm Fatigue in Hospitals: A Systematic Review. In *NCDHWS 2026*, Springer. https://link.springer.com/chapter/10.1007/978-3-032-28819-6_37

[25] Computational approaches to alleviate alarm fatigue in intensive care medicine: A systematic literature review. *PMC*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9424650/

[26] Artificial Intelligence Technologies for Coping with Alarm Fatigue in Hospital Environments Because of Sensory Overload: Algorithm

Development and Validation. *Journal of Medical Internet Research*, 21(11), e15406 (2019). https://www.jmir.org/2019/11/e15406/

[27] Detecting False Alarms by Analyzing Alarm-Context Information: Algorithm Development and Validation. *PMC*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7270842/

[28] Artificial Intelligence-Powered Tiered Early Warning Framework Addressing High False Alarm Rates for In-Hospital Mortality Prediction (AI-TEW). Scholarly Commons. https://scholarlycommons.libraryinfo.bhs.org/all_works/11616/

[29] FDA Finalizes Premarket Cybersecurity Guidance for Medical Devices. King & Spalding (2023). https://www.kslaw.com/news-and-insights/fda-finalizes-premarket-cybersecurity-guidance-for-medical-devices

[30] US FDA Cybersecurity Premarket Guidance (2023): Definition & Risk Management in Medical Devices Explained. Kusari. https://www.kusari.dev/learning-center/us-fda-cybersecurity-premarket-guidance-2023

[31] Reliable ECG Anomaly Detection on Edge Devices for Internet of Medical Things Applications. *PMC*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12031393/

[32] Fully Autonomous Z-Score-Based TinyML Anomaly Detection on Resource-Constrained MCUs Using Power Side-Channel Data. arXiv:2604.08581. https://arxiv.org/abs/2604.08581

[33] ArrythML: An Autoencoder-Based TinyML Approach for On-Device Arrhythmia Detection on Resource-Constrained Embedded Systems. arXiv:2606.02256. https://arxiv.org/pdf/2606.02256

[34] Tiny Machine Learning and On-Device Inference: A Survey of Applications, Challenges, and Future Directions. *PMC*. https://pmc.ncbi.nlm.nih.gov/articles/PMC12115890/

# برعاية كريمة واشراف السادة الافاضل

الأستاذ الدكتور أسامة إسماعيل المشهداني رئيس جامعة نينوى.
الأستاذ الدكتور كوسار محمد علي رئيس جامعة السليمانية.
الأستاذ الدكتور وسام صلاح عبد الحسين رئيس جامعة قرطبة.
الأستاذ الدكتور حاتم هذال عبدالحميد مدير عام دائرة البحوث والدراسات.
الأستاذ رباح إبراهيم آل جعفر مدير عام دائرة التعليم الديني والدراسات الإسلامية.
الأستاذ الدكتور صالح عبد المهدي الخالدي عميد كلية ابن خلدون الجامعة.
الأستاذة الدكتورة سوزان صابر حيدر عميد كلية الإدارة والاقتصاد - جامعة السليمانية.
الدكتور علي مهدي كاظم عميد كلية العلوم - جامعة قرطبة.
الدكتور علاء ثابت جاسم رئيس مؤسسة أجيال الرافدين لتطوير التعليم.
الأستاذ المساعد الدكتور علي عبودي نعمة المدير التنفيذي لمؤسسة أجيال الرافدين لتطوير التعليم.
الدكتور أحمد جميل مجيد المدرس مدير مركز وعي للاستشارات وبناء القدرات.


الرقيم السومري
للنشر والتوزيع
Scriptum Synthesis Publisher (SSSP)

+964 0777 52 62 892
IRAQ - DAIYALA - BAGHUBA
ISBN: 978-9922-7961-61
email : manager@sss-publisher.com
https://doi.org/10.62909/978.9922.7961.61
Sumerian scriptum synthesis publisher (SSSP)