


Hybrid Machine Learning and Blockchain Technology for Early Detection of Cyberattacks in Healthcare Systems



Faisal Ghazi Abdiwi 

Department of Computer Science & Information Systems, Al-Mansour University College, Baghdad 10067, Iraq

Corresponding Author Email: faisal.ghazi@muc.edu.iq

Copyright: ©2024 The author. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijssse.140622>

ABSTRACT

Received: 11 October 2024

Revised: 15 November 2024

Accepted: 29 November 2024

Available online: 31 December 2024

Keywords:

blockchain, machine learning, health care, balancing techniques

This study aims to improve the safety of healthcare data through applied machine learning and blockchain technology. This research aims to develop an all-in-one platform that uses deep learning to identify and handle cyber-attacks and apply blockchain technologies to protect the information through a Proof of Work (PoW) based mechanism. The in-one procedure creates stands with the appropriate information, reveals the ovule with PoW, and subsequently joins the stands to the chain. With this, it becomes possible to enhance data security, making it less prone to alteration and hacking attempts. The suggested model was applied to the CIC IoMT 2024 dataset that includes 40 actual and simulated IoT medical devices and 18 types of cyberattacks and utilized Wi-Fi, MQTT, and Bluetooth protocols. After normalizing the dataset with RandomOverSampler, the model recorded 100% accuracy. There is a need for a fusion of deep learning and fewer resources requiring light techniques while upholding the planning of distribution of a non-centralized system of data, particularly a distribution management system. In general, through this research, we seek to provide an innovative and advanced contribution to enhancing healthcare data security, taking advantage of modern technologies to combat increasing cyber threats.

1. INTRODUCTION

The healthcare system faces many challenges in managing large volumes of data generated daily by numerous medical tools, electronic health records, telemedicine, and mobile health services that may require efficient, fast, and secure data analysis techniques to deliver high-quality services to patients [1].

Machine learning and blockchain technologies are highly valued across various fields. This study focuses on the classification of advanced machine-learning techniques and effective blockchain methods for processing large data volumes. Given that the healthcare system handles highly sensitive data concerning patients' health and personal information, it necessitates robust security services. Blockchain technology can serve as an effective solution for this purpose [2, 3].

The use of machine learning (ML) algorithms is increasing dramatically due to rapid digitalization. In the current digital age, machine learning algorithms are being applied in fields such as healthcare [4]. However, to anticipate or resolve specific issues, these algorithms must be properly trained. It is quite possible that training data sets can be altered, leading to biased outcomes. Meanwhile, the advent of blockchain technology (BC) has spurred a digital revolution in several industries, including supply chain, healthcare, and finance [5].

Intelligent healthcare systems use it to provide patients with control and transparency over their medical records. Nevertheless, the integration of blockchain technology and healthcare still faces numerous challenges, including concerns over patient data storage, security, and privacy [6]. Regarding security, new attacks aim to compromise various elements of the blockchain network, including nodes, wallets, consensus algorithms, and smart contracts (SC) [7]. Thus, to enhance scalability and learning, blockchain technologies offer many opportunities for the healthcare system to achieve its goals, such as reducing healthcare costs, ensuring effective diagnosis, providing timely treatment, ensuring transparency in regulatory reports, and guaranteeing effective health data management by building a security system capable of pre-detecting cyber threats that could periodically threaten medical systems [8, 9]. Therefore, a solution for AI-based healthcare systems using blockchain technologies has been proposed to enhance privacy and security and address the accompanying challenges [10].

This research paper consists of six sections. Section 2 reviews the most important literary works related to the topic. Section 3 presents the theoretical background and the most prominent techniques used. Section 4 discusses the proposed methodology, while Section 5 presents the results and their discussion. Section 6 concludes the paper and provides future recommendations.

2. LITERATURE REVIEW

Some previous research studies have dealt with the classification and segmentation of brain tumors through the design and implementation of various programs based on Artificial Intelligence.

The study proposed by Ali et al. [11] integrates hybrid deep learning models with a permissions-based blockchain framework to create scalable and secure healthcare solutions. The framework facilitates seamless data sharing and collaboration among healthcare providers while guaranteeing that only authorized entities have access to and the ability to edit sensitive health information. Hybrid deep learning models also make it possible to analyze massive amounts of healthcare data in real time, which speeds up illness prediction, therapy suggestions, and diagnosis. Integrating blockchain with hybrid deep learning has many benefits, including privacy, interoperability, scalability, and improved healthcare decision-making.

Gadekallu et al. [12] proposed a blockchain-based method to protect datasets generated by an IoT device and used in e-health applications. The proposed fix for the above issue entails the usage of a private cloud blockchain. The researchers created a framework that can be used by data set owners to protect their data during inspection.

The "model malware validation" (PoPMV) algorithm was presented by Mohammed et al. [13] and is intended for the ICPS blockchain. To get rewards and feedback in real-time, it makes use of a deep learning model (LSTM) and reinforcement learning techniques. Mitigating vulnerabilities, increasing processing speed, identifying known and unknown attacks, and enhancing ICPS capability are the main objectives. Simulations reveal that the suggested method is superior to current blockchain frameworks, exhibiting dynamic microservice allocation and a 30% overall attack detection security improvement.

Rathee et al. [14] presented a security framework for multimedia data in healthcare using blockchain technology by creating a hash for each data so that any data change, modification, or drug hack is reflected to all users of the blockchain network. The results were analyzed against the traditional approach and validated with improved simulation results achieving an 86% success rate against product drop, counterfeiting, wormhole, and probabilistic authentication scenarios thanks to blockchain technology.

The study proposed by Mohammed et al. [15] suggests a system architecture based on machine learning (ML) for identifying fraudulent transactions and attacks in the BC network. The system entails two steps: (1) using machine learning to verify sensor-derived medical data and prevent erroneous data from entering the blockchain network, and (2) verifying blockchain transactions with the same machine learning algorithm whereby regular transactions are stored, while anomalous transactions are flagged as new attacks in the assaults database. Six machine learning algorithms (KNN, Naive Bayes, SVM, Random Forest, Decision Tree, and Logistic Regression) and two datasets were used in the construction of the system. The outcomes show that by obtaining the best accuracy, execution speed, and scalability, the Random Forest method fared better than the others. As a result, it was regarded as the finest.

Kumar et al. [16] combined deep learning (DL) techniques with blockchain technology and permissionless smart

contracts to create a novel, safe, and effective data-sharing framework called PBDL. To be more precise, PBDL comprises a blockchain system for connected entity registration, verification via zero-knowledge evidence, and verification by a consensus process based on smart contracts. Secondly, a novel deep learning system that combines self-attention-based bidirectional long-term memory (SA-BiLSTM) and sparse variational encoder (SSVAE) is proposed using the validated data. Under this method, SA-BiLSTM detects and enhances the attack detection process while SSVAE encrypts or transforms healthcare data into a new format. The superiority of the PBDL framework over current state-of-the-art methodologies is confirmed by security analyses and experimental findings utilizing IoT-Botnet and ToN-IoT datasets.

Chen et al. [17] described a blockchain-enabled diabetes diagnosis framework that uses several machine-learning classification algorithms to detect the condition early and securely keep track of patients' electronic health records. The Interplanetary File System (IPFS), where patient health data is gathered via wearable sensors, blockchain technology, and symptom-based disease prediction are all combined to create the EHR sharing framework. The EHR Manager receives this data and uses a machine learning model to process it further and to gather the required outputs. With the patient's and their practitioner's permission, the results are then recorded in the blockchain along with the physiological information.

To guarantee the privacy of healthcare data with the advancement of blockchain technology, Ashraf et al. [18] proposed FIDChain IDS, which uses lightweight artificial neural networks (ANN) with a federated learning (FL) method. This distributed ledger allows for the collection of local weights and subsequent broadcasting of the data. After averaging, the global weights are updated, preventing poisoning attacks and giving the distributed system complete transparency and stability with very little overhead. By using a detection model at the edge, which blocks data from its gateway with a shorter detection time and requires less computing and processing capacity as FL handles smaller groups of data, the cloud is protected from attacks.

Kumar et al. [19] developed a blockchain-coordinated deep learning approach, henceforth referred to as "BDSDT", to enable safe data transfer in an Internet of Things-enabled healthcare system. Specifically, by utilizing the Zero Knowledge Proof (ZKP) technique, a novel scalable blockchain architecture was first presented to guarantee data integrity and secure transmission. Next, BDSDT interfaces with the Ethereum smart contract to handle data security concerns and the InterPlanetary File System (IPFS) for off-chain storage to address challenges with data storage costs. In addition, a deep learning architecture for HS network intrusion detection was designed using the verified data. The latter creates an efficient intrusion detection system by combining bidirectional long short-term memory (BiLSTM) with deep sparse autoencoder (DSAE). The proposed BDSDT achieves close to 99% accuracy using both datasets and outperforms state-of-the-art approaches in both non-blockchain and blockchain situations, according to experiments conducted on two public data sources (CICIDS-2017 and ToN-IoT). Table 1 shows the summaries of the relevant literature.

The studies were distinguished by their successful methods of managing health care according to the Aman system. However, there are many loopholes, as shown in Table 2.

Table 1. Summary of literature review

Ref.	Methods	Results	Models
[11]	Integrates hybrid deep learning models with a permissions-based blockchain framework	Secure and scalable healthcare solutions; improve illness prediction, therapy suggestions, and diagnosis	Hybrid deep learning models
[12]	Blockchain-based method to protect datasets generated by IoT devices	Protects data during the inspection using a private cloud blockchain	N/A
[13]	PoPMV algorithm for ICPS blockchain using LSTM and reinforcement learning	Mitigates vulnerabilities increases processing speed and improves attack detection by 30%	Deep learning model (LSTM) and reinforcement learning
[14]	Security framework for multimedia data in healthcare using blockchain technology	Achieves 86% success rate against various attacks	Blockchain technology
[15]	An architecture of the system based on neural networks to detect fraudulent transactions and assaults in the BC network	The Random Forest method demonstrates the best accuracy, execution speed, and scalability	KNN, Naive Bayes, SVM, Random Forest, Decision Tree, Logistic Regression
[16]	PBDL framework combining DL techniques with blockchain and permissionless smart contracts	Superior performance in security and attack detection	SA-BiLSTM and SSSVAE
[17]	Blockchain-enabled diabetes diagnosis framework using ML classification algorithms	Early detection of diabetes and secure EHR management	Various ML classification algorithms
[18]	FID Chain IDS using lightweight ANN with federated learning	Protects cloud from attacks with transparency and stability	Lightweight ANN with FL method
[19]	BSDST approach for safe data transfer in IoT-enabled healthcare system	Achieves close to 99% accuracy; outperforms state-of-the-art approaches	BiLSTM and DSAE

Table 2. Summary of loopholes of literature review

Ref.	Methods
[11]	Limited real-world implementation; potential scaling issues with hybrid deep learning models. They are primarily theoretical or simulation-based, making scalability and applicability in real-world healthcare systems uncertain.
[12]	Focus on private cloud blockchain; lacks consideration of public blockchain benefits.
[13]	Simulation-based results need validation in real-world scenarios.
[14]	Limited to multimedia data; doesn't address non-multimedia healthcare data.
[15]	It focuses on specific ML algorithms and lacks comparison with more recent models.
[16]	Primarily theoretical, it requires more extensive empirical validation.
[17]	Limited to diabetes diagnosis; doesn't cover other chronic diseases.
[18]	Potential overhead with federated learning requires optimization for large-scale deployment.
[19]	Focus on specific datasets; needs evaluation on more diverse and extensive healthcare datasets.

Source: Central Bank and the Ministry of Communications

The literature review highlights significant advancements in leveraging Artificial Intelligence and blockchain technologies for healthcare solutions, yet it underscores unresolved key gaps. These include limited scalability and real-world applicability, as many studies remain theoretical or simulation-based, a narrow focus on specific use cases such as diabetes diagnosis or multimedia security, insufficient exploration of public and hybrid blockchain configurations, and the persistent issue of imbalanced datasets affecting model accuracy. Additionally, a lack of rigorous empirical validation across diverse datasets limits the generalizability of these approaches. This study addresses these gaps by proposing an integrated framework that combines advanced deep learning models, robust blockchain mechanisms such as Proof of Work, and dataset balancing techniques, offering a scalable, secure, and empirically validated solution for cyberattack detection in healthcare systems.

3. THEORETICAL BACKGROUND

This section entails a review of the theoretical aspect used in the design and implementation of the proposed model, with a brief explanation of the most important methods relied upon, which are based on machine learning and deep learning.

3.1 Artificial Intelligence (AI)

Artificial Intelligence covers many areas, such as natural language processing, robotics, expert systems, fuzzy logic, machine learning, and deep learning, which will be explained in the following sub-sections. Figure 1 shows the types of Artificial Intelligence [20].

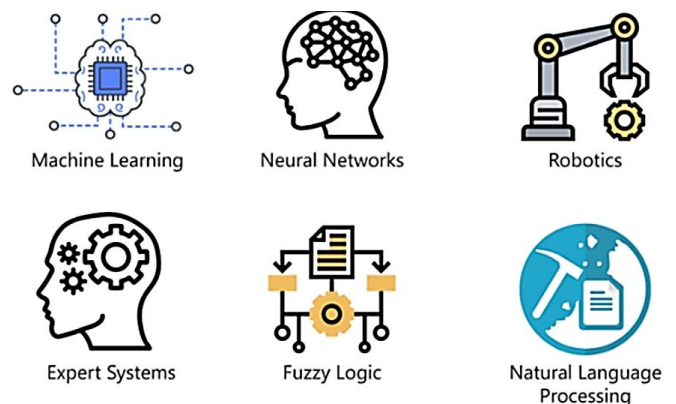


Figure 1. Types of Artificial Intelligence [20]

3.2 Machine learning

Within the arena of Artificial Intelligence, machine learning makes use of statistical models and algorithms to teach computers to learn from data and make choices without the need to explicitly write rules. In short, machine learning algorithms enable computers to recognize patterns, forecast effects, and make choices based on those patterns. Multiple techniques exist for machine learning including supervised learning where the data used for training is monitored, unsupervised learning where the system learns on its own without direct supervision, semi-supervised learning, and reinforcement learning. These techniques can be applied in multiple fields such as image and speech recognition, natural language processing, and analytical predictions [21-24].

Figure 2 shows a general model of supervised learning (classification techniques). However, classification suffers from limitations such as missing data which can cause problems during the training and classification phases. Failure to enter records due to misinterpretation, removal of unwanted data, or hardware malfunction can result in such missing data [25, 26].

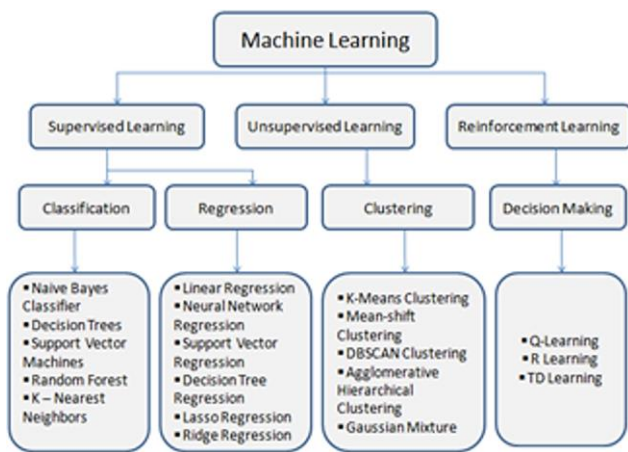


Figure 2. Supervised learning classification techniques [26]

3.3 Machine learning models

ML is an important factor of AI. Algorithms are used to analyze data and make relevant decisions. It consists of several branches. Supervised ML involves the use of training information and feedback from people to analyze the relationship between given inputs to a given output, whilst Unsupervised Learning involves exploring input data without being given an explicit output variable [27].

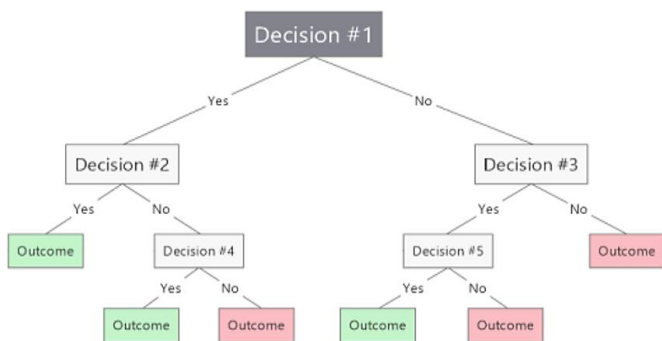


Figure 3. Decision Tree model

3.3.1 Decision Tree (DT)

This is a widely used algorithm in the field of data mining and ML due to its simplicity and effectiveness. The main objective of DT is to produce a model that can calculate the value of a required variable based on multiple input variables [28]. Figure 3 shows a DT model [29].

3.3.2 Support Vector Machines (SVMs)

SVMs are a set of algorithms that belong to the category of supervised ML techniques, which are suitable for classification and regression. To train an SVM, an optimization algorithm is utilized to identify the hyperplane that best separates the classes [30]. Figure 4 illustrates an SVM [30].

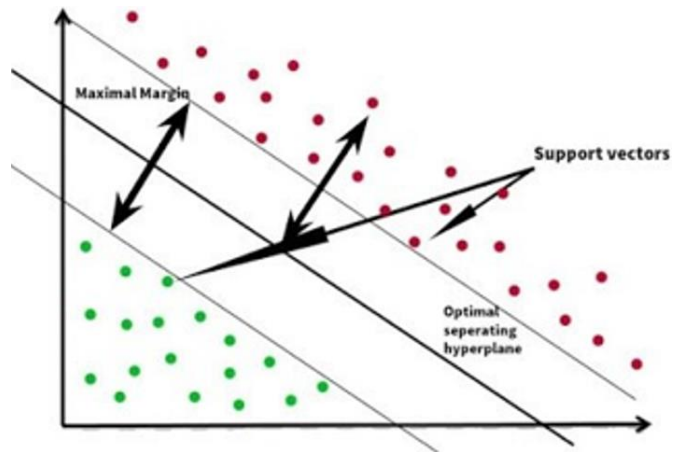


Figure 4. Support vector machine [30]

3.3.3 K-Nearest Neighbor (KNN)

A supervised machine learning algorithm for regression and classification is known as KNN. Predicting a new data point's class label based on the class labels of its closest neighbors is the aim of classification. Predicting the constant target variable's value using the values of its closest neighbors is the aim of regression analysis. The success of K-means algorithm usage is determined by the proper selection of K. If K is too small, then the model may become overly sensitive to noise that is present in the dataset. On the other hand, if K is large, it could end up in an oversimplified decision space that might be able to capture some relationships within the data [31]. Figure 5 illustrates a KNN [31].

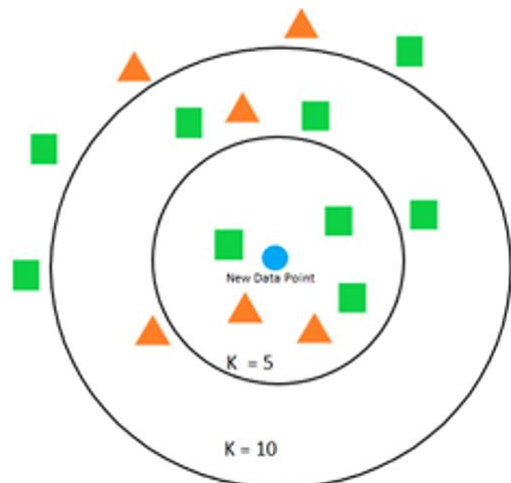


Figure 5. The K-Nearest Neighbor model [31]

3.3.4 Naive Bayes Classifier (NB)

Naive Bayes (NB) is a classification method rooted in probability. It leans on Bayes' theorem, a math formula that helps gauge the likelihood of a theory given certain evidence. The naive part comes from the assumption that the presence or absence of one feature does not hinge on the presence or absence of any other feature [32]. Figure 6 illustrates a NB model [32].

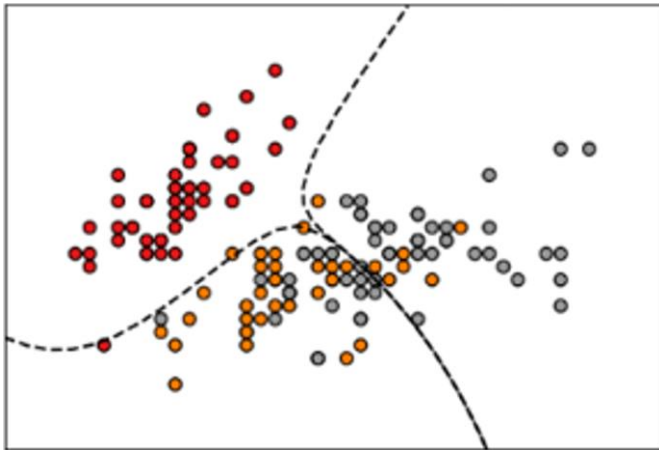


Figure 6. Naïve Bayes classifier [32]

It is short and efficient, making it ideal for tackling massive datasets. However, it might not be carried out properly if the functions are biased [33].

3.4 Ensemble methods in ML

Ensemble methods in machine learning are like putting together a dream team of models to boost accuracy and reliability in predictions. The great thing about ensemble methods is that they gather a bunch of different models and merge their powers. This way, if one model has a weak spot, the others can pick up the slack, making the whole system stronger. Figure 7 shows the ensemble methods [34].

3.5 Blockchain technology

Blockchain technology is usually defined as a decentralized system that records transactions in a time series of connected blocks that form a distributed database. The system is distributed and fault-tolerant, meaning that the data is stored on some independent devices and is publicly accessible.

The blocks in the blockchain as shown in Figure 8 contain

a set of transactions, and each block contains a hash value that points to the previous block, creating a chain that can only be modified by network consensus. This achieves immutability, as records cannot be easily falsified or modified.

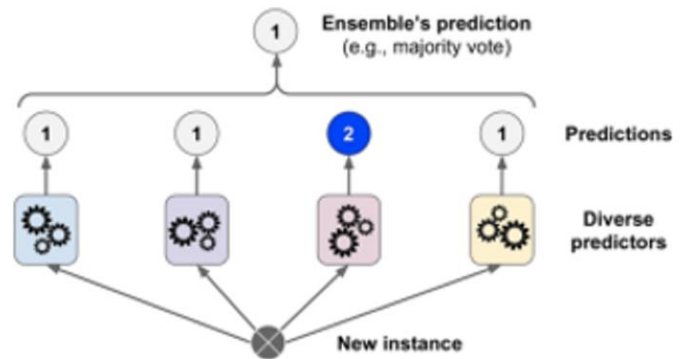


Figure 7. Ensemble methods [34]

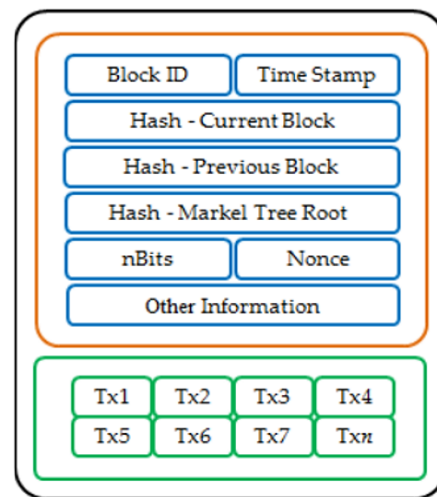


Figure 8. Block structure in blockchain [35]

The digital signature algorithm is important in blockchain, where each user signs their transactions with their private key, allowing others to verify the validity of the transactions using their public key [35].

Proof of Work (PoW), as shown in Figure 9, is one of the consensus methods used in blockchain, where a difficult mathematical problem is solved to prove that the node has duly mined a new block, which ensures the security of the network and facilitates the addition of new blocks.

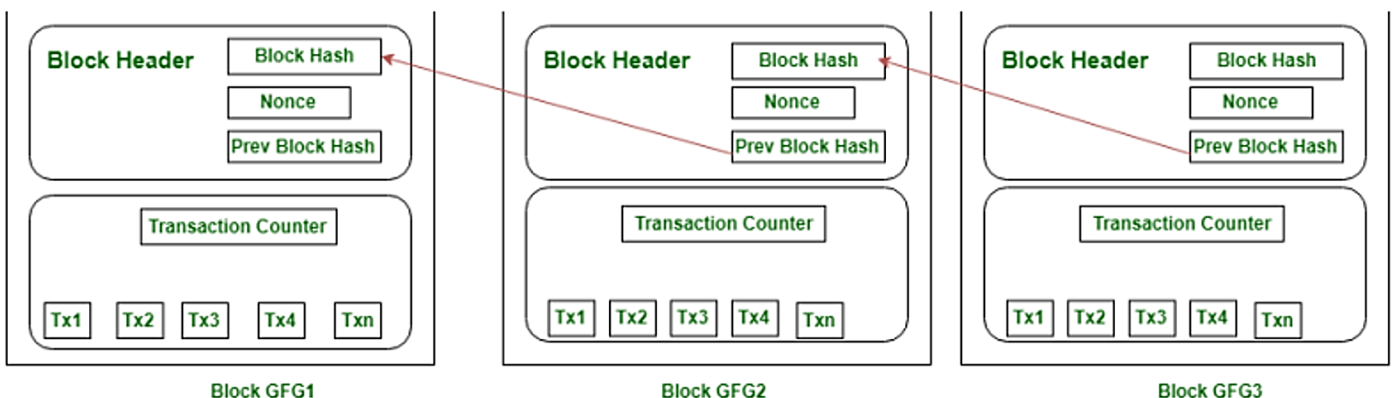


Figure 9. Proof of Work (PoW) [35]

These characteristics of blockchain—immutability, decentralization (no central authority), and robust security—make it a desirable system for many applications requiring security, trust, and transparency, like machine learning-based healthcare system authentication [36].

The most common application of blockchain technology in the healthcare industry is to safeguard patient data, demonstrating the significance of security in this domain. Between July 2021 and June 2022, 692 significant healthcare data breaches were reported. In these breaches, bank account and credit card details, along with health and genetic testing reports, were pilfered. Blockchain technology is perfect for security applications because it offers an unchangeable, transparent, and decentralized record of all medical data. Although it is transparent, it secures identity by protecting patient privacy using sophisticated, safe procedures that maintain the confidentiality of medical information. This technology's decentralized architecture makes it possible for physicians, patients, and other healthcare workers to securely and swiftly exchange information [37].

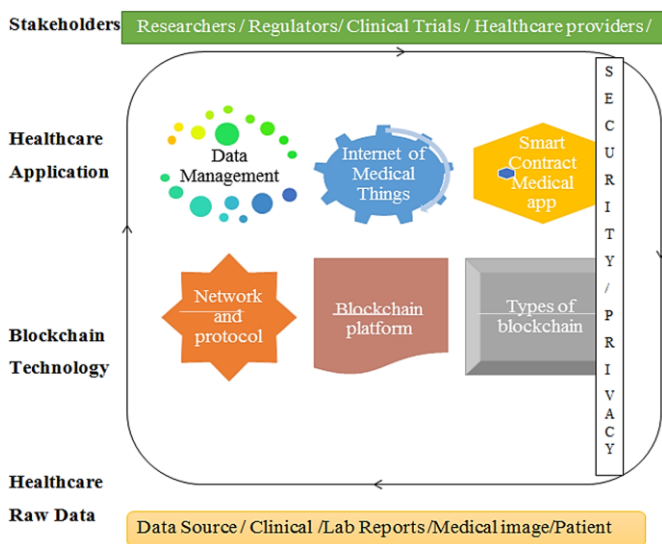


Figure 10. Blockchain-based workflow environment for healthcare application

Blockchain is reinventing traditional identity management in many healthcare applications because of its capacity to securely and openly communicate data, thus modernizing legacy systems, increasing efficiency, and lowering costs in the industry. In theory, blockchain-based healthcare technologies are divided into four layers: primary data sources, blockchain technology, healthcare applications, and healthcare stakeholders. Figure 10 depicts a blockchain-based workflow environment for healthcare applications [38].

Firstly, raw data—obtained from medical devices, labs, and other sources—is combined and subsequently turned into big data. This massive amount of data, which makes up the top layer of the stack, is the basis for blockchain healthcare. Situated at the top of the raw data layer, blockchain technology forms the basis of a secure four-tier healthcare architecture. The elements that make it easy for users to create and manage transactions vary throughout blockchain platforms; these characteristics include consensus methods and protocols [39].

In recent years, numerous blockchain-based platforms have been developed such as Ethereum and Hyperledger. Wallets, digital assets, smart contracts, and signatures are some of the main parts of blockchain technology. Different protocols, such

as distributed, peer-to-peer, and decentralized ones, can be used to communicate amongst various blockchain networks or applications. Stakeholders may use a public, private, or consortium blockchain, depending on their needs [40].

Ensuring that the apps are incorporated into the system as a whole comes after the platform is constructed utilizing blockchain technology. Three areas can be used to group blockchain-based healthcare applications: Internet of Medical Things (IoMT), smart contract medical applications, and data management. Electronic health, data storage, data administration, and worldwide scientific data sharing for research and development (R&D) are all included in data management records. Smart contract medical applications include clinical trials and medications.

Data security, IoT infrastructure for healthcare, and healthcare devices compose the IoMT (Internet of Medical Things). Applications for medicine on the blockchain are shown in Figure 11. The stakeholder layer, which is the highest tier in the hierarchy, consists of people who engage in blockchain-based healthcare applications, including patients, healthcare workers, researchers, and so on. At this level, sharing and managing data while guaranteeing its security and privacy are the key concerns of the users [41].

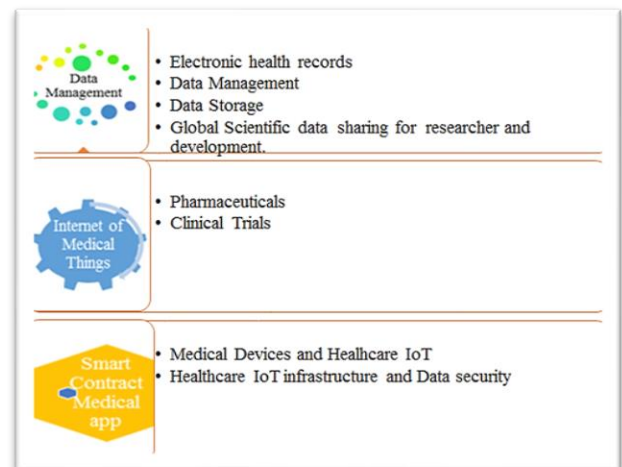


Figure 11. The system's usage in healthcare [41]

3.6 Concept of cyber attacks

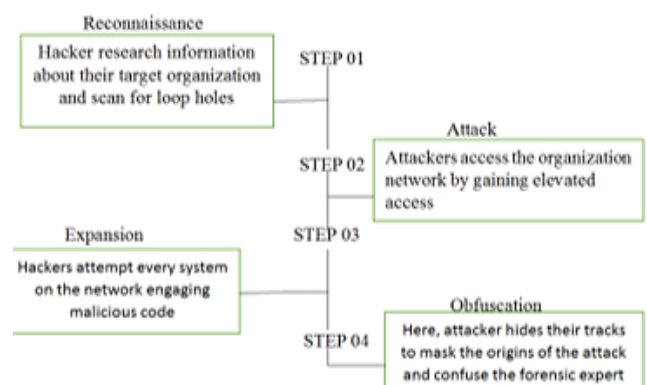


Figure 12. Anatomy of cyber-attacks [44]

The deliberate utilization of computer systems, technology-dependent businesses, and networks is referred to as a cyberattack [42]. Cyberattacks modify computer code, logic,

or data by using malicious code. This causes unfavorable outcomes that may reveal data and encourage cybercrimes such as identity and information theft. Hackers often sneak into business networks covertly to carry out their nefarious deeds, such as stealing important data, using a variety of sophisticated strategies to avoid detection [43]. They often employ complex strategies to encode their attempts to avoid detection by intrusion detection systems. Figure 12 depicts the composition of cyberattacks [44].

Following the attack and subsequent oppression, the attackers might try to download and install malware onto the compromised machine. The four stages of a digital assault are elongation, attack, obfuscation, and reconnaissance [44].

3.7 Dataset

By shedding light on various components involved in designing and implementing security solutions for the IoMT the CIC IoMT 2024 dataset is very useful as a reference. Among the proposed 40 IoT devices posted in this software based on various healthcare protocols like WIFI, MQTT, and Bluetooth, 25 physical and 15 simulated IoT Test (IoMT) attacks have been simulated. These attacks consist of DDoS, DoS, Recon, MQTT, and spoofing; all of which should help researchers develop further ideas on how to improve the healthcare systems beyond mechanisms such as ML. The proposed CIC IoMT 2024 dataset would go a long way in setting the realism bar as far as identifying the security vulnerabilities within the IoMT devices is concerned. This dataset comprises 18 cyberattacks on 40 IoMT devices affecting various health protocols including Wi-Fi, MQTT, and Bluetooth. The research applied an innovative process that generates and captures the IoMT network traffic mimicking a real-hospital network structure. In this case, real and virtual devices as well as modern networking technologies including network taps enable continuous data capture.

3.7.1 Data pre-processing

All the dataset files were successfully merged into one file containing the unified dataset. The final file consists of 7,160,831 rows and 46 features. The new dataset comprises 18 types of attacks, and their details are outlined in Table 3 and Figure 13.

Table 3. Features in the dataset

Feature Name	Samples
TCP_IP-DDoS-UDP	1635956
TCP_IP-DDoS-ICMP	1537476
TCP_IP-DDoS-TCP	804465
TCP_IP-DDoS-SYN	801962
TCP_IP-DoS-UDP	566950
TCP_IP-DoS-SYN	441903
TCP_IP-DoS-ICMP	416292
TCP_IP-DoS-TCP	380384
Benign	192732
MQTT-DDoS-Connect_Flood	173036
Recon-Port_Scan	83981
MQTT-DoS-Publish_Flood	44376
MQTT-DDoS-Publish_Flood	27623
Recon-OS_Scan	16832
ARP_Spoofing	16047
MQTT-DoS-Connect_Flood	12773
MQTT-Malformed_Data	5130
Recon-VulScan	2173
Recon-Ping_Sweep	740

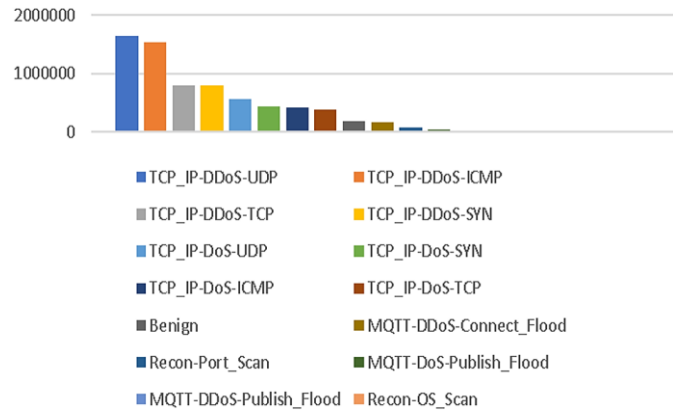


Figure 13. Types of attacks in the CICIOMT2024 dataset

When applying encoding to the dataset and converting it into a binary classification dataset (0,1), benign links amount to 192,732, whereas attack links amount to 6,968,099. This indicates that the categories are imbalanced within the dataset, as depicted in Figure 14.

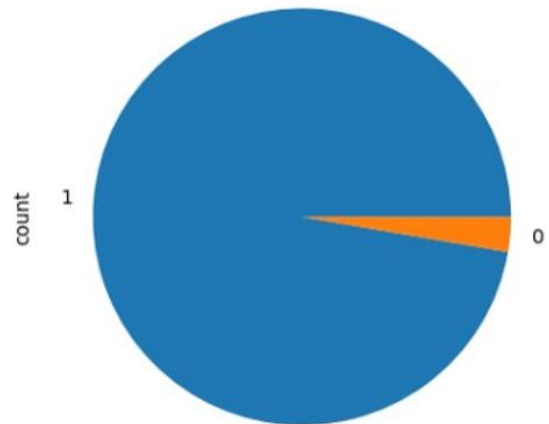


Figure 14. The imbalanced dataset

The CIC IoMT 2024 dataset suffers from class imbalance, making it particularly unsuitable for binary classification. Imbalance causes machine learning models to become more biased towards the most popular category, leading to unreliable predictions in the evaluation phase. Therefore, the approach proposes to balance the dataset (Figure 15) and obtain a balanced set for training the model and evaluating its performance.

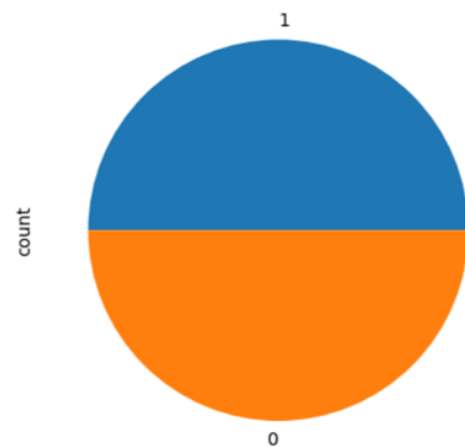


Figure 15. Balanced dataset

This is done using one of the most common techniques for addressing imbalance namely RandomOverSampler. This stage calls for the use of techniques to produce new synthetic samples bringing the number of samples in the minority class to a higher level. All these methods help minimize the variance in the evaluation of the model while making it more effective in handling IoT data that may be as diverse and imbalanced as depicted in the following model.

3.8 Evaluation metrics

Evaluating performance and efficiency is critical. When analyzing machine learning models, various measures are commonly employed to guarantee that the model is properly appraised. For classification tasks, some evaluation measures are available, including Accuracy, Confusion Matrix, Recall, Precision, and F1 Score, which will be described in the following sub-sections [45-48].

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

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

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

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

4. PROPOSED MODEL METHODOLOGY

In this study, we propose a high-fidelity methodology that aims to integrate the concepts of blockchain (Proof of Work), data preprocessing, machine learning model training, and evaluation. It concludes with a system alert mechanism based on model predictions. Each step ensures data integrity (blockchain) and data preparation (normalization and partitioning). The selection of machine learning models and blockchain technologies in this research has been due to the advantage they bring to the provision of a response to particular challenges affecting cybersecurity within healthcare. Specifically, for the high accuracy, scalability, and capacity to deal with complex, multidimensional datasets as witnessed in the CIC IoMT 2024 dataset, preference is given to the Random Forest over other existing machine-based learners that have been proved inferior. Special techniques like RandomOverSampler are also used to handle class imbalance problems. This is meant to ensure that the model is robust and performs well across various cyberattack scenarios.

Robust blockchain technologies, PoW, in particular, are implemented to protect sensitive healthcare information. When it comes to security, PoW makes sure the immutability and decentralization of data, are two factors that are quite important in protecting medical records from being tampered with and having unauthorized access to them. Hence, the choice of blockchain supports the aim of the study to improve transparency and resilience in healthcare systems by deploying a decentralized architecture so as not to have single points of failure. Thus, these technologies form a comprehensive and secure framework for overcoming issues related to scalability, data integrity issues, and reliability related to cyber-threat

detection.

Its performance is evaluated using performance metrics such as confusion matrix, precision, precision, ROC curve, F1 score, area under the curve, and recall. At last, the system generates alerts depending on the model’s forecasts relating to the test data: making alerts concerning possible network assaults in case such a change is suggested or rather claiming that no threat has been recognized.

The extensive approach guarantees proper data handling, correct training of the models, and detection of threats in advance. Figure 16 depicts the essential procedures and processes of the working methodology

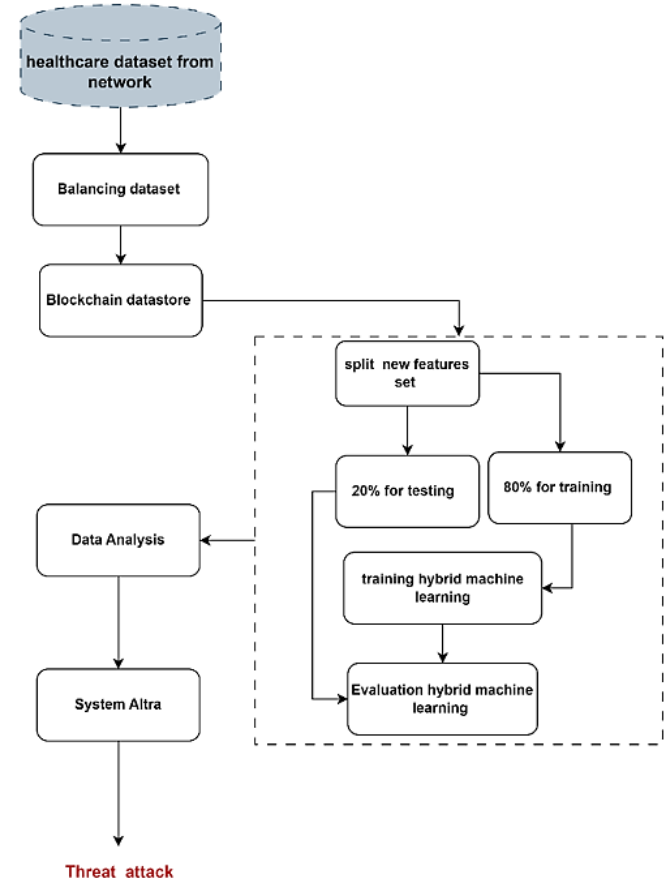


Figure 16. Proposed model architecture

4.1 Blockchain data storage

During the Blockchain Data Storage phase, a Blockchain class is developed together with the help of a Proof of Work (PoW) to ensure data security. This process involves making a new block that contains the information that the system wants to record followed by the computation of hash for this particular block following PoW where a lot of extensive mathematical computation is done to ensure that the information is valid and cannot be altered. Once the requirements of PoW are fulfilled, the block is incorporated into the chain and becomes a permanent part. This makes each block in the chain have a unique identifier making it hard to alter any block without having to affect all the blocks in the chain it affords high security and transparency of the data stored in the blocks. PoW decentralizes control of power amongst the network participants and does not require a single authority. It increases security and minimizes the possibility of an attack that could directly address the healthcare system as it is in this model

4.2 Splitting dataset

First, we randomly shuffled the data. Then, we split the data into training and test sets using "train_test_split", where 20% of the data is used for testing and 80% for training, whilst "random mode=101" ensures that the data is always split in the same way.

5. RESULTS AND ANALYSIS

To test and classify the data in the CIC IoT 2023 dataset, The RandomOverSampler technique was used in the paper to solve the imbalance issue. RandomOverSampler doubles the quantities of data of the sparse category by duplicating the existing samples in every random manner to match the count with that of the coarser category. It is useful for enhancing the performance of the given model because the proposed model is endowed with a small impact on the superiority of the common class while raising the significance of the rare datum.

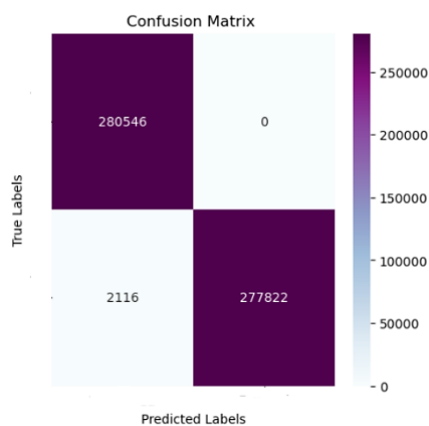


Figure 17. Confusion matrix RandomOverSampler method

	precision	recall	f1-score	support
0	0.99	1.00	1.00	280546
1	1.00	0.99	1.00	279938
accuracy			1.00	560484
macro avg	1.00	1.00	1.00	560484
weighted avg	1.00	1.00	1.00	560484

```
[[280546  0]
 [ 2116 277822]]
```

Figure 18. Report classifier RandomOverSampler method

Table 4 and Figures 17-19 depict the experiment result of the proposed system and of the RandomOverSampler technology in rectifying the imbalanced data set problem.

The model was applied to improve the accuracy of cyberattack classification on the CIC IoT 2023 dataset. Rare classes were randomly oversampled using RandomOverSampler to find a balance between classes, allowing the model to learn more general patterns and deliver high performance across all classes.

The model outperformed remarkably, with an accuracy of 99.7%, a positive precision of 100%, a recall of 99.6%, and an F1 score of 99.6%. This indicates its ability to reduce major errors, especially false positives by not reporting real threats or false negatives by misreporting threats. It supports the fruits of the labor related to early detection of cyberattacks in health IoT systems with reduced risk of charges due to classification errors.

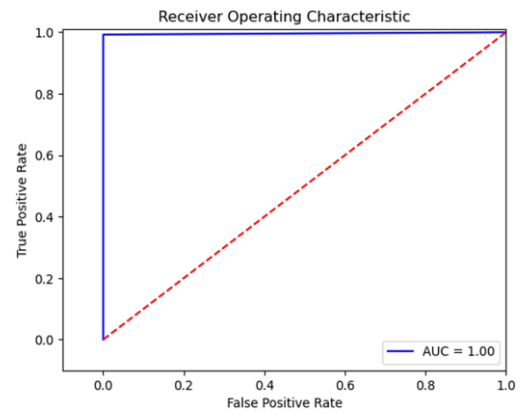


Figure 19. The ROC with RandomOverSampler method

Table 4. Performance of the proposed model with RandomOverSampler method

Performance Metrics	Result (%)
Accuracy	99.7
Precision	100
F1-score	99.6
Recall	99.6

The system has a near-perfect recall of attack cases, ensuring that almost all potential attacks are detected, and excellent accuracy to ensure that benign secure data is not classified as threats; this would increase the number of false positives and cripple the reliability of the system in healthcare environments. While the results are very good, there are some limitations to keep in mind:

1. Overfitting: Since RandomOverSampler generates artificial samples by duplicating some of the original samples, this can lead to overfitting of the model.

2. Real-world cyberattack scenarios: Although the CIC IoT 2023 dataset claims to capture a wide range of the most prevalent network attacks in the IoT world, it is expected that some real-world cyberattack scenarios will be missed to be captured by this dataset.

3. Test environments: The results have not been validated in dynamic real-world test environments. Testing the model in different contexts would be interesting to see if the results hold and can be further generalized.

4. Data accuracy assumption: These results are based on the assumption that the data accurately reflects network traffic and attack patterns; this may not be the case.

This indicates the model's excellent performance under simulation, but further confirmation of its applicability and flexibility in real systems requires further experiments in real and diverse work environments to increase the model's credibility and contribute to improving healthcare cybersecurity.

6. CONCLUSIONS AND FUTURE WORK

The purpose of this research is to increase the protection of healthcare data by applying machine learning as well as blockchain. The CIC IoMT 2024 dataset was checked and assessed, consisting of 18 types of cyber-attacks on 40 medical IoT devices, of which 25 were real, and 15 were simulated devices; the protocols used were Wi-Fi, MQTT, and Bluetooth. The novelty of the research is in proposing and developing an integrated system based on the method of deep

learning to detect and analyze cyber-attacks.

Besides that, efficiency and credibility are met by adopting a Protocol of Work, which is a part of the blockchain setup. This includes developing a new block with the necessary data, calculating the hash by PoW, and joining the block to the chain. That way, safety is enhanced making sure that the data cannot be manipulated and safeguarded from an attack.

We expect significant work in the direction of deep learning transfer techniques like EfficientNetV2 which is smaller and faster than CNNs that have been adapted to new ISIC datasets. Moreover, in analyzing the generative capacity of the generative adversarial networks (hybrid GANs), one finds a direction to enhance and diversify training sets which may also push forward skin cancer diagnosis via deep learning particularly so in the case of pre-trained models.

REFERENCES

- [1] McGuire, S. (2016). World cancer report 2014. Geneva, Switzerland: World Health Organization, international agency for research on cancer, WHO Press, 2015. *Advances in Nutrition*, 7(2): 418-419.
- [2] Chen, J., Ning, C., Zhou, Z., Yu, P., Zhu, Y., Tan, G., Mao, C. (2019). Nanomaterials as photothermal therapeutic agents. *Progress in Materials Science*, 99: 1-26. <https://doi.org/10.1016/j.pmatsci.2018.07.005>
- [3] Ting, D.S., Liu, Y., Burlina, P., Xu, X., Bressler, N.M., Wong, T.Y. (2018). AI for medical imaging goes deep. *Nature Medicine*, 24(5): 539-540. <https://doi.org/10.1038/s41591-018-0029-3>
- [4] Wolf, M., de Boer, A., Sharma, K., Boor, P., et al. (2018). Magnetic resonance imaging T1-and T2-mapping to assess renal structure and function: A systematic review and statement paper. *Nephrology Dialysis Transplantation*, 33(suppl_2): ii41-ii50. <https://doi.org/10.1093/ndt/gfy198>
- [5] Boerma, T. Mathers, C.D. (2015). The World Health Organization and global health estimates: Improving collaboration and capacity. *BMC Medicine*, 13: 1-4.
- [6] Wang, Y., Zu, C., Hu, G., Luo, Y., et al. (2018). Automatic tumor segmentation with deep convolutional neural networks for radiotherapy applications. *Neural Processing Letters*, 48: 1323-1334. <https://doi.org/10.1007/s11063-017-9759-3>
- [7] Jégou, S., Drozdal, M., Vazquez, D., Romero, A., Bengio, Y. (2017). The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation. In 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, pp. 1175-1183. <https://doi.org/10.1109/CVPRW.2017.156>
- [8] Tufail, A.B., Ullah, I., Khan, W.U., Asif, M., et al. (2021). Diagnosis of diabetic retinopathy through retinal fundus images and 3D convolutional neural networks with limited number of samples. *Wireless Communications and Mobile Computing*, 2021(1): 6013448. <https://doi.org/10.1155/2021/6013448>
- [9] Fuzzell, L.N., Perkins, R.B., Christy, S.M., Lake, P.W., Vadaparampil, S.T. (2021). Cervical cancer screening in the United States: Challenges and potential solutions for underscreened groups. *Preventive Medicine*, 144: 106400. <https://doi.org/10.1016/j.ypmed.2020.106400>
- [10] Mehrotra, R., Ansari, M.A., Agrawal, R., Anand, R.S. (2020). A transfer learning approach for AI-based classification of brain tumors. *Machine Learning with Applications*, 2: 100003. <https://doi.org/10.1016/j.mlwa.2020.100003>
- [11] Ali, A., Ali, H., Saeed, A., Ahmed Khan, A., et al. (2023). Blockchain-powered healthcare systems: Enhancing scalability and security with hybrid deep learning. *Sensors*, 23(18): 7740. <https://doi.org/10.3390/s23187740>
- [12] Gadekallu, T.R., Manoj, M.K., Kumar, N., Hakak, S., Bhattacharya, S. (2021). Blockchain-based attack detection on machine learning algorithms for IoT-based e-health applications. *IEEE Internet of Things Magazine*, 4(3): 30-33. <https://doi.org/10.1109/IOTM.1021.2000160>
- [13] Mohammed, M.A., Lakhan, A., Zebari, D.A., Abd Ghani, M.K., et al. (2024). Securing healthcare data in industrial cyber-physical systems using combining deep learning and blockchain technology. *Engineering Applications of Artificial Intelligence*, 129: 107612. <https://doi.org/10.1016/j.engappai.2023.107612>
- [14] Rathee, G., Sharma, A., Saini, H., Kumar, R., Iqbal, R. (2020). A hybrid framework for multimedia data processing in IoT-healthcare using blockchain technology. *Multimedia Tools and Applications*, 79(15): 9711-9733. <https://doi.org/10.1007/s11042-019-07835-3>
- [15] Mohammed, M.A., Boujelben, M., Abid, M. (2023). A novel approach for fraud detection in blockchain-based healthcare networks using machine learning. *Future Internet*, 15(8): 250. <https://doi.org/10.3390/fi15080250>
- [16] Kumar, R., Kumar, P., Tripathi, R., Gupta, G.P., Islam, A.N., Shorfuzzaman, M. (2022). Permissioned blockchain and deep learning for secure and efficient data sharing in industrial healthcare systems. *IEEE Transactions on Industrial Informatics*, 18(11): 8065-8073. <https://doi.org/10.1109/TII.2022.3161631>
- [17] Chen, M., Malook, T., Rehman, A.U., Muhammad, Y., et al. (2021). Blockchain-enabled healthcare system for detection of diabetes. *Journal of Information Security and Applications*, 58: 102771. <https://doi.org/10.1016/j.jisa.2021.102771>
- [18] Ashraf, E., Areed, N.F., Salem, H., Abdelhay, E.H., Farouk, A. (2022). Fidchain: Federated intrusion detection system for blockchain-enabled IoT healthcare applications. *Healthcare*, 10(6): 1110. <https://doi.org/10.3390/healthcare10061110>
- [19] Kumar, P., Kumar, R., Gupta, G.P., Tripathi, R., Jolfaei, A., Islam, A.N. (2023). A blockchain-orchestrated deep learning approach for secure data transmission in IoT-enabled healthcare system. *Journal of Parallel and Distributed Computing*, 172: 69-83. <https://doi.org/10.1016/j.jpdc.2022.10.002>
- [20] Smak Gregoor, A.M., Sangers, T.E., Bakker, L.J., Hollestein, L., Uyl-de Groot, C.A., Nijsten, T., Wakkee, M. (2023). An artificial intelligence based app for skin cancer detection evaluated in a population based setting. *NPJ Digital Medicine*, 6(1): 90. <https://doi.org/10.1038/s41746-023-00831-w>
- [21] Varshney, P., Gupta, C., Girdhar, P., Mohan, A., Agrawal, P., Madaan, V. (2022). Data analysis using machine learning: An experimental study on UFC. In *Machine Learning and Data Science: Fundamentals and Applications*. Scrivener Publishing LLC, pp. 23-46. <https://doi.org/10.1002/9781119776499.ch2>

- [22] Pettit, R.W., Fullem, R., Cheng, C., Amos, C.I. (2021). Artificial intelligence, machine learning, and deep learning for clinical outcome prediction. *Emerging Topics in Life Sciences*, 5(6): 729-745. <https://doi.org/10.1042/ETLS20210246>
- [23] Rajbanshi, S. (2021). Everything you need to know about machine learning. *Analytics Vidhya*. <https://www.analyticsvidhya.com/blog/2021/03/everything-you-need-to-know-about-machine-learning/>.
- [24] An, Q., Rahman, S., Zhou, J., Kang, J.J. (2023). A comprehensive review on machine learning in healthcare industry: Classification, restrictions, opportunities and challenges. *Sensors*, 23(9): 4178. <https://doi.org/10.3390/s23094178>
- [25] Nacem, S., Ali, A., Anam, S., Ahmed, M.M. (2023). An unsupervised machine learning algorithms: Comprehensive review. *International Journal of Computing and Digital Systems*, 13(1): 911-921. <https://doi.org/10.12785/ijcds/130172>
- [26] Khan, A., Qureshi, M., Daniyal, M., Tawiah, K. (2023). A novel study on machine learning algorithm-based cardiovascular disease prediction. *Health & Social Care in the Community*, 2023(1): 1406060. <https://doi.org/10.1155/2023/1406060>
- [27] Wu, Z., Jiang, H., Zhao, K., Li, X. (2020). An adaptive deep transfer learning method for bearing fault diagnosis. *Measurement*, 151: 107227. <https://doi.org/10.1016/j.measurement.2019.107227>
- [28] Yu, C., Han, R., Song, M., Liu, C., Chang, C.I. (2020). A simplified 2D-3D CNN architecture for hyperspectral image classification based on spatial-spectral fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13: 2485-2501. <https://doi.org/10.1109/JSTARS.2020.2983224>
- [29] Hesamian, M.H., Jia, W., He, X., Kennedy, P. (2019). Deep learning techniques for medical image segmentation: Achievements and challenges. *Journal of Digital Imaging*, 32: 582-596. <https://doi.org/10.1007/s10278-019-00227-x>
- [30] Yang, B., Lee, C.G., Lei, Y., Li, N., Lu, N. (2021). Deep partial transfer learning network: A method to selectively transfer diagnostic knowledge across related machines. *Mechanical Systems and Signal Processing*, 156: 107618. <https://doi.org/10.1016/j.ymssp.2021.107618>
- [31] Yang, B., Lei, Y., Jia, F., Xing, S. (2019). An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. *Mechanical Systems and Signal Processing*, 122: 692-706. <https://doi.org/10.1016/j.ymssp.2018.12.051>
- [32] Sarkar, D. (2018). A comprehensive hands-on guide to transfer learning with real-world applications in deep learning. *Towards Data Science*.
- [33] Albelwi, S.A. (2022). Deep architecture based on DenseNet-121 model for weather image recognition. *International Journal of Advanced Computer Science and Applications*, 13(10): 559-565. <https://doi.org/10.14569/IJACSA.2022.0131065>
- [34] ZJE_ANDY. (2020). DenseNet. <https://blog.csdn.net/u014453898/article/details/105670550>.
- [35] Sharma, D.K., Pant, S., Sharma, M., Brahmachari, S. (2020). Cryptocurrency mechanisms for blockchains: Models, characteristics, challenges, and applications. *Handbook of research on blockchain technology*, 323-348. <https://doi.org/10.1016/B978-0-12-819816-2.00013-7>
- [36] Saraf, C., Sabadra, S. (2018). Blockchain platforms: A compendium. In 2018 IEEE International Conference on Innovative Research and Development (ICIRD), Bangkok, Thailand, pp. 1-6. <https://doi.org/10.1109/ICIRD.2018.8376323>
- [37] Alexaki, S., Alexandris, G., Katos, V., Petroulakis, N.E. (2018). Blockchain-based electronic patient records for regulated circular healthcare jurisdictions. In 2018 IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), Barcelona, Spain, pp. 1-6. <https://doi.org/10.1109/CAMAD.2018.8514954>
- [38] Li, H., Zhu, L., Shen, M., Gao, F., Tao, X., Liu, S. (2018). Blockchain-based data preservation system for medical data. *Journal of Medical Systems*, 42: 141. <https://doi.org/10.1007/s10916-018-0997-3>
- [39] Azaria, A., Ekblaw, A., Vieira, T., Lippman, A. (2016). MedRec: Using blockchain for medical data access and permission management. In 2016 2nd International Conference on Open and Big Data (OBD), Vienna, Austria, pp. 25-30. <https://doi.org/10.1109/OBD.2016.11>
- [40] Liang, G., Weller, S.R., Luo, F., Zhao, J., Dong, Z.Y. (2018). Distributed blockchain-based data protection framework for modern power systems against cyber attacks. *IEEE Transactions on Smart Grid*, 10(3): 3162-3173. <https://doi.org/10.1109/TSG.2018.2819663>
- [41] Beal, V. (2024). *Techopedia*. <https://www.techopedia.com/definition/cyberattack>.
- [42] Nwosu, A.U., Goyal, S.B., Bedi, P. (2021). Blockchain transforming cyber-attacks: Healthcare industry. In *Proceedings of the 11th International Conference on Innovations in Bio-Inspired Computing and Applications (IBICA 2020)*, pp. 258-266. https://doi.org/10.1007/978-3-030-73603-3_24
- [43] Powers, D.M. (2020). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*. <https://doi.org/10.48550/arXiv.2010.16061>
- [44] Qi, J., Du, J., Siniscalchi, S.M., Ma, X., Lee, C.H. (2020). On mean absolute error for deep neural network based vector-to-vector regression. *IEEE Signal Processing Letters*, 27: 1485-1489. <https://doi.org/10.1109/LSP.2020.3016837>
- [45] d'Agostino, R., Capezzuto, P., Bruno, G., Cramarossa, F. (1985). Mechanism of etching, polymerization and deposition in RF (radio frequency) discharges. *Pure and Applied Chemistry*, 57(9): 1287-1298. <https://doi.org/10.1351/pac198557091287>
- [46] Kumar, A. (2023). Ensemble methods in machine learning: Examples. <https://vitalflux.com/5-common-ensemble-methods-in-machine-learning/>.
- [47] Wang, W., Lu, Y. (2018). Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model. *IOP Conference Series: Materials Science and Engineering*, 324: 012049. <https://doi.org/10.1088/1757-899X/324/1/012049>
- [48] Dadkhah, S., Neto, E.C.P., Ferreira, R., Molokwu, R.C., Sadeghi, S., Ghorbani, A. (2024). CICIoMT2024: Attack vectors in healthcare devices-a multi-protocol dataset for assessing IoMT device security. *Preprints.org*. <https://doi.org/10.20944/preprints202402.0898.v1>