Journal of Administrative and Economic Sciences • Vol 18, Issue 1 (2025) • p 607-618

Secure E-healthcare System Based on Blockchain Technology and Ensemble Models

Adel ALTI

Department of Management Information Systems, College of Business and Economics,
Qassim University, Buraidah • Saudi Arabia

a.alti@qu.edu.sa

Abstract:

Ensuring the security and privacy of e-healthcare systems is now crucial due to the increasing occurrence of various cyberattacks in vulnerable environments, including Distributed Denial of Service (DDoS). Blockchain technology and ensemble models offer one of the most promising solutions to ensure data privacy and integrity within secure e-healthcare system. In this paper, blockchain and ensemble models' capabilities are employed to develop a secure e-healthcare system. The DDoS attack detection mechanism utilizes ensemble model training and smart contracts to significantly enhance attacks detection efficiency and accuracy. We assess the approach's performance by evaluating two ensemble models, Random Forest (RF) and Extreme Gradient Boosting (XGBoost) focusing on DDoS attacks detection using the recent dataset CICIoMT2024. The combination of blockchain and ensemble models result in accurate detection of DDoS attacks exceeding 90%. This improvement signifies an unprecedented innovation in ensemble models in blockchain-enabled e-healthcare systems. Moreover, in terms of prediction time, the DT model excels with the fastest prediction time.

Keywords: Blockchain, Internet of Medical Things, DDoS, Ensemble Model, Accuracy, e-healthcare System.

1. Introduction

The Internet of Medical Things (IoMT) is a powerful tool for e-health applications. It is a cutting-edge technology that includes various interconnected medical sensors and devices to monitor personal health data [**1**]. Continuous, real-time ambulatory health data monitoring is crucial for assessing patients' health status, improving life quality and ensuring timely medical interventions. Wireless data transmission is also invaluable for medical diagnosis. However, transmission of health data faces several threats and attacks. Distributed Denial of Service (DDoS) attacks are a major threat to health data transmission [**2**]. These attacks may lead to serious problems such as identity theft, alteration of medical data, and loss of confidential information. Significant efforts are still required to develop a new robust e-healthcare system that aligns with ongoing progress and provides secured services by integrating artificial intelligence with blockchain technology.

Blockchain technology has become one of the most important strategies for modern healthcare, with the goal of providing secure and trustworthy services. In addition, it has gained great popularity, emerging as indispensable tool in healthcare sector. Developing new decentralized and transactive technology that leverage digital technologies for data processing improves the security of continuous data transactions. Blockchain can be public, private and hybrid. Public blockchains operate with unlimited user access, while private blockchains allow greater user control. The potential for e-healthcare applications has been significantly expanded by recent research studies [**6 - 15**]. Studies have demonstrated that blockchain technology excels in protecting health data. However, the blockchain network faces significant challenges, such as the identification of security vulnerabilities, DDoS attacks, and anomalies [**3**]. The blockchain suffered from a DDoS attack generating 4000.00 transactions per second, resulting in several hours of downtime and a network rollback supported by 80% of validators [**4**]. In addition, the blockchain was subject to DDoS attack that overwhelmed its sequencer, causing a 45-minute outage without risking user funds [**5**]. These incidents highlight the vulnerabilities in blockchain networks when faced with high transaction volumes over a short period. Therefore, implementing secure and efficient IoMT systems in a blockchain environment is crucial to efficiently detect DDoS attacks.

This paper presents a secure e-healthcare system built on ensemble models and blockchain technology. The system is designed to detect different types of attacks. The effectiveness of ensemble models is demonstrated in their DDoS detection capabilities. This ability not only improves the security of e-healthcare applications but also aids in predicting different types of attacks. By integrating blockchain technology and ensemble models, the system enhances security in the e-healthcare sector. The system highlights the significance of employing ensemble model through two models: Extreme Gradient Boosting (XGBoost) and Random Forest (RF) to evaluate the prediction time and accuracy. It considers CICIoMT2024 dataset [**6**] for detecting DDoS attack. These innovations are crucial to enhancing the overall performance and robustness of healthcare applications and represent a significant advance in the field of e-healthcare applications.

The paper is structured as follows: Section 2 presents related work and its limitations. Section 3 presents the design and implementation of secure healthcare applications based on blockchain and ensemble models. Section 4 is dedicated to the discussion of our results. Finally, Section 5 concludes our study.

2. Related Work

Challenges in e-healthcare systems are continually increasing, highlighting the need for improved security measures. This demand has driven researchers to develop techniques that enhance both the effectiveness and transparency of healthcare data and transaction safety. In this section, we review some recent research aimed at enhancing the security of e-healthcare system through the use of blockchain technology and ensemble models.

2.1 Blockchain to secure e-healthcare Systems

Reffad *et al.* [**7**] proposed a secure distributed mobile-fog-cloud service approach based on Diffie Hellman-RSA and blockchain to maintain the privacy of documents and health data stored in the Cloud. A smart contract is used to reinforce the security of healthcare applications. Wu *et al.* [**8**] proposed SmartCheck, a smart contract solution for healthcare applications. By leveraging blockchain technology, SmartCheck aims to identify potential threats in smart contracts by analyzing and flagging vulnerabilities. In addition, smart contracts are used to fully automate SCM operations, validating transactions through the Proof of Work (PoW) consensus mechanism. Key features of the solution include the implementation of a ContractFuzzer that generates test inputs based on the ABI specification to identify security vulnerabilities during the contract's runtime. Kumar *et al.* [**9**] presented a combined blockchain and IoT-based fog solution to effectively monitor DDoS attacks across mobile and wearable devices. It extracts relevant features from the collected data, such as traffic volume, transaction frequency, and behavioral patterns of IoT devices and blockchain transactions. The proposed solution was evaluated using BoT-IoT dataset, demonstrating its suitability for detecting and mitigating DDoS attacks. The findings showed that XGBoost outperforms in binary attack detection, while RF excels in multi-attack detection. Moreover, RF requires less training and testing time compared to XGBoost on distributed fog nodes. This work considers a limited number of attacks and disregards the security aspects of blockchain itself within the transaction's legitimacy guarantee mechanisms. Saad *et al.* [**10**] presented a blockchain-based approach to ensure the safety of consumer devices in the context of e-healthcare systems. The proposed approach addresses these issues by measuring temperature, latency, bandwidth, Block Transaction Rate (BTR), and block loss percentages to identify DDoS attacks.

2.2 Ensemble Models Learning for Attack Detection in e-healthcare Systems

The process of ensemble model learning involves the integration of multiple individual models to generate predictions that are more accurate compared to those of a single model. In other world, combining two or more machine learning models can lead to improved predictions by leveraging the strengths of each model. In the domain of attack detection, Latif *et al.* [**11**] attempted to detect DoS, R2L, U2R, and probe attacks in the e-healthcare system using decision tree technique. The Wisconsin dataset was used, which consists of nine attributes. Using Decision Trees, the system successfully detected each type of attack at 99.49% precision.

Nsaif *et al.* [**12**] attempted to identify DDoS attacks through sniffing techniques using both machine learning-based classifiers and K-means clustering. Although not as quick as the other two, the study managed to detect DDoS attack at a precision rate of 86.29%. Recently, an approach combining supervised dimensionality reduction and classification model was proposed to detect spoofing attacks in a cloud-centric healthcare solution [**13**]. Feature representations were appended to give an overall feature representation. This study estimated that feature representation and the learned decision boundaries performed better than PCA model.

Moulahi *et al.* [16] proposed a system that integrates blockchain technology with Federated Learning (FL) to strengthen security and safeguard data privacy. By incorporating FL into the blockchain framework, this solution ensures the confidentiality of patient information.

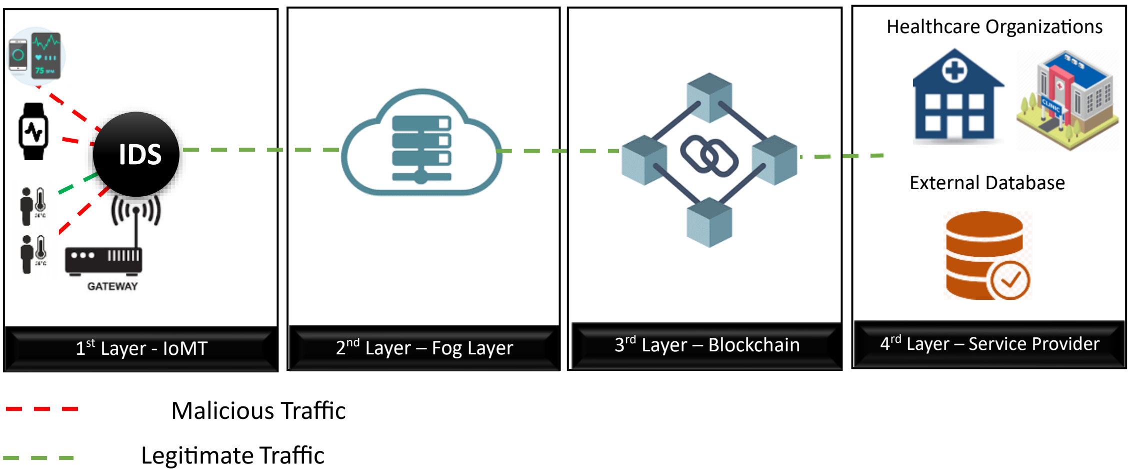
In another study, Alabdulatif *et al.* [17] combined machine learning models with blockchain technology by storing the gradients of trained models on-chain. They also reconstructed the decision functions of Support Vector Machines (SVM) and Multi-Layer Perceptron (MLP) models directly on the blockchain.

Alsayegh *et al.* [18] introduced an asymmetric searchable encryption and proxy re-encryption technique leveraging the patient’s public key to protect patient confidentiality during third-party access. Their results revealed that both the record length and key size significantly influence the computational and communication costs of the encryption algorithm.

3. Methodology

Fig.1 illustrates the pivotal elements of our approach to secure e-health information system, we opted for blockchain and ensemble model: data acquisition, pre-processing, ensemble model classification, and secure blockchain storage. Data acquisition is facilitated by medical sensors positioned near the Fog layer. These sensors capture health data, which is then collected by a microcontroller and wirelessly transmitted via a getaway to a fog server. The getaway consists of an IDS that filters the data generated by sensors and medical devices to determine the nature of traffic (legitimate, malicious). This involves filtering to eliminate noisy data, checking of missing data, and removal of duplicate data. The fog layer runs the blockchain which is a decentralized platform, manages the healthcare data, and is connected to healthcare providers or an external storage source.

Figure 1: The proposed approach to secure e-healthcare system.



3.1 Operating principle of the proposed approach

As shown in Fig. 2, the e-health information system consists of several components, wherein each component collaborates autonomously to ensure security and data privacy. The health data (i.e. blood pressure, temperature, patient location, ECG, etc.) and transactions are generated by the sensors and medical devices, analyzed in real-time, and filtered by ensemble models using the IDS component, and the validated transactions are stored into blockchain within the health data. This integration of blockchain technology and ensemble models is one way to effectively ensure security and data privacy.

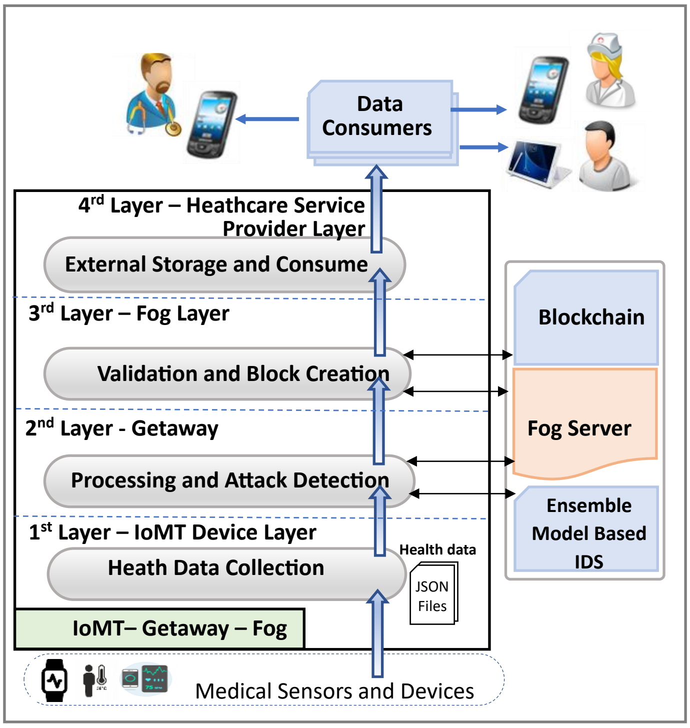

Data Consumers
4rd Layer – Heathcare Service Provider Layer
External Storage and Consume
3rd Layer – Fog Layer
Validation and Block Creation
2nd Layer - Getaway
Processing and Attack Detection
1st Layer – IoMT Device Layer
Heath Data Collection
Health data
JSON Files
IoMT– Getaway – Fog
Medical Sensors and Devices
Blockchain
Fog Server
Ensemble Model Based IDS

Figure 2: The operating principle of the proposed approach.

3.2 Dataset

In this paper, ensemble model is pre-trained on CICIoMT2024 dataset from the Institute of Canadian, Cybersecurity Laboratory by Dadkhah *et al.* [6]. This dataset includes malicious and benign traffic with a mix of forty real and simulated medical devices. It contains a large amount of traffic records, totaling 8775013 samples, with 45 numerical features divided into five classes: Spoofing, Recon, MQTT, DoS, DDoS, and benign traffic. The work was used 4971919 records and distributed across different sizes.

3.3 Preprocessing

Medical data closely linked to sensors, are susceptible to noise, making data preprocessing essential for subsequent noisy elimination processes. This improves data quality and reliability. This section describes the preprocessing steps used in this study: missing data verification, numerical conversion, normalization, data duplication removal, and data split. First, we checked the missing data to prevent the model from handling noisy data and thus speeding up the training phase. Then, we ensured uniform data types were significant for a consistent analysis. Therefore, we converted all values in the records to numeric types. This step safeguards against potential issues arising from inconsistent data types. Next, we normalized the data using a standard scaler to eliminate outlier values. We also identified and removed 5.119 duplicate rows to prevent overfitting problems. Finally, the dataset is separated into two subsets, one for training and one for testing. The training phase comprises 80% of records and the testing phase 20% of records from a total of 4.971.919 records.

3.4 Attack detection with ensemble Model

The proposed secure e-health information system has an IDS component as shown in Fig. 1, which gives it control over the incoming data and transactions. The IDS recognizes that the transactions contain legitimate traffic and notifies the system administrator. The system administrator determines whether the transaction is a real threat or a false alert and takes appropriate action. This study aimed to apply various ensemble models including, Random Forest (RF) and Extreme Gradient Boosting (XGBoost) for detecting DDoS attacks within e-healthcare system.

- RF used a concept called collective intelligence. Its builds a bunch of decision trees independently, which are simple predictors and aggregates their results into a single result using a majority vote.
- XGBoost is a collection of decision trees. The trees are built additively to improve deficiencies of the previous trees. XGBoost uses a method boosting which combines weak learners sequentially and minimizes the gradient of loss function to achieve accurate results.

4. Results and Discussions

In the proposed model, we used Google Colab and Python platform. Google Colab is cloud-based environment used for our analysis. It used Python programming language and leveraged essential libraries such as TensorFlow, Keras, and Kaggle. The implementation consists of collecting health data and transactions during device sensing, the training phase using DDoS and benign traffic samples from CICIoMT2024 dataset [**6**] to evaluate two models: XGBoost and RF with existing approaches. After training, the models were tested against different evaluation metrics to assess their prediction capabilities. In the rest of this section, we will detail the evaluation results of each model across attack classification, using various metrics such as accuracy, precision, recall, F1-score, and prediction time.

4.1 Evaluation Metrics

The metrics used are accuracy, precision, F1 score, recall, confusion matrix, and prediction time. These metrics are often used in the prediction of attacks and the classification of traffic data.

- **Accuracy:** is a widely used metric for predictions. It determines the percentage of correct predictions out of the total number of predictions. One way to calculate the accuracy of ensemble model is defined by Equation 1:

$$Accuracy = \frac{\#Correct\ Predictions}{\#Total\ Predictions} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

  where True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). The higher the accuracy values, the better the model performance.

- **Precision:** It is defined as the percentage of positive predictions compared to all predicted positive cases.

$$Accuracy = \frac{TP}{(TP + TN)} \tag{2}$$

- **Recall:** It is defined as the percentage of positive cases that are correctly identified.

$$Recall = \frac{TP}{(TP + FN)} \tag{3}$$

- **F1-Score:** is an important metric to evaluate the balance between precision and recall.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

- **Prediction time** is another metric that is often used to evaluate ensemble model. It is the time required to predict the attacks. In the context of performance, prediction time helps determine the effort needed for attack training and testing. The lower the time values, the better the model performance.

4.2 Comparison of Attack Prediction-Based Ensemble Learning and Machine Learning

The proposed approach underwent comprehensive testing with five types of attacks all of them were evaluated with two ensemble models, XGBoost, RF in comparison with other relevant studies [**11**][**12**][**16**][**18**]. In [**11**], the approach utilized decision tree as a straightforward predictor for attack detection, while [**12**] utilized the K-means clustering method to identify attacks. In [**16**], the approach utilized federated learning and blockchain technologies while [**18**] utilized blockchain with decision functions. Fig. 3 displays the accuracy rate of five types of attacks using three different ensemble and machine learning models. A single iteration corresponds to a 5-minute session in which the training algorithm was applied, and biofeedback was given approximately 2000 times.

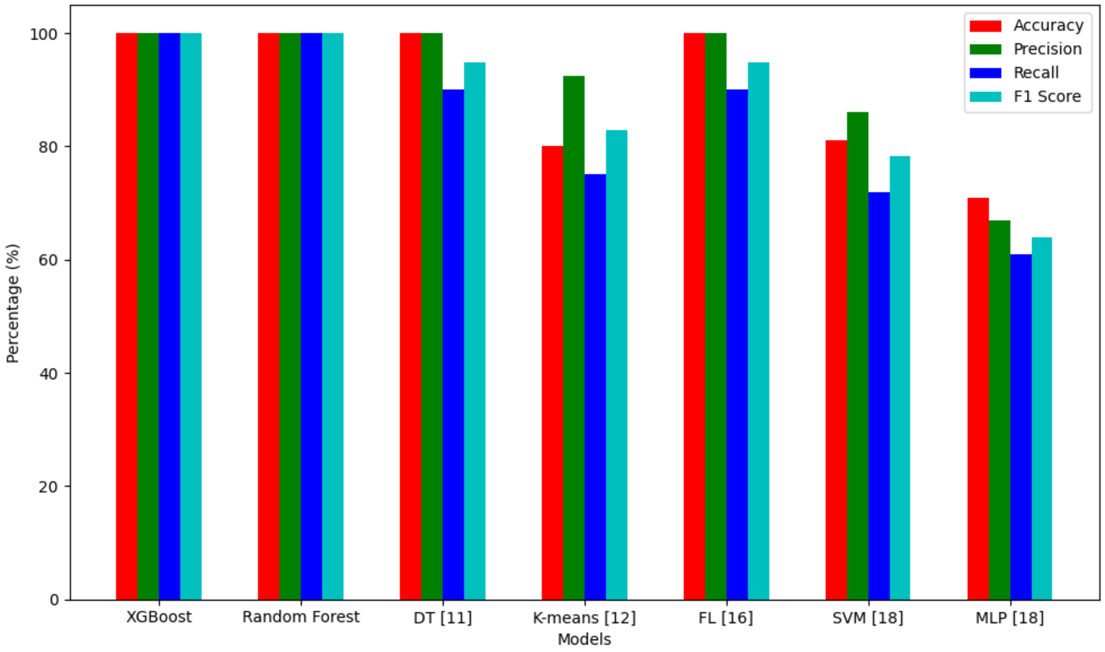


Figure 3: Comparison of accuracy, precision, recall, and F1 Score for various models.

From Fig.3, it is evident that XGBoost and RF achieves the highest accuracy, precision, recall, and F1 score. These results are achieved due to the contribution of the ensemble learning models, which is evident from the comparison with other related work in [**11**][**12**][**16**][**18**]. It appears clear that the DT based attack detection system performs worse than RF and XGBoost in almost all tests, which indicates its low performance in these cases. For SVM [18], accuracy (81%) and precision (86%) are good, with recall at 72% and an F1 score of 78.34%. While the SVM performs reasonably well, it doesn't achieve the same level of performance as XGBoost or Random Forest, likely due to less sensitivity to true positives compared to these models, leading to lower recall and F1 score. For Federated Learning (FL) [18], accuracy (100%), with recall at 90% and an F1 score of 94.74%, is exceptional and outperforms SVM, MLP, K-means and DT.

From this experiment, it is evident that ensemble learning models' selection improves the performance more than DT-based attack detection system.

4.3 Evaluation of Prediction Time-based Ensemble Learning and Machine Learning

The prediction time comparison of our proposed approach with other works [**11**][**12**][**16**][**18**] in Table 1 shows prediction times difference, indicating that DT based model [**11**], along with the relevant extracted features, outperforms the XGBoost and RF models in predicting attacks and assessing risks.

Table 1. Prediction time of binary and multi-class classification.

Model	Binary Classification (ms)	Multi- Classification (ms)
XGBoost	1	3
RF	1.2	1.4
DT [**11**]	0.043	0.064
K-Means [**12**]	0.810	0.920
FL [**16**]	-	2.026
SVM DF [**18**]	-	1.6808

Overall, these results are justified by DT, which efficiently builds complete conditions for different types of attacks. In addition, RF and XGBoost, Federated Learning (FL), SVM Decision Function (DF) aggregate variety of attacks predictions which are also slower to build.

4.4 Evaluation of Storage Requirement and Gas Consumption for Transactions

We implemented our blockchain using the Docker and Ganache frameworks, configuring the blockchain parameters with a maximum gas limit of 6,721,975 per block. To minimize gas fees, we maintained an average of 4.31 and 4.32 events, storing data directly on the blockchain network instead of using smart contracts.

The gas consumption for transactions and storage costs of our approach were compared with other blockchain approaches in [16] and [18], as presented in Table 2. The findings show that Smart Contract Federated Learning (FL) [16] requires the least storage, making it highly suitable for environments with limited storage capacity, such as edge devices or decentralized networks in e-healthcare. In contrast, smart contract and Decision Function (DF) [18] requires a moderate storage infrastructure. The proposed Blockchain Ensemble Learning (EL) strikes a balance, offering both storage efficiency and the ability to handle CICIoMT2024 datasets, making it suitable for moderately storage-constrained scenarios. However, although the proposed approach is scalable and performs well in critical attacks, it could face challenges with large-scale data storage demands.

Table 2. Comparison of average gas consumption and storage cost (n number of transactions).

Model	Storage Requirement (Kb)	Average Gas Consumption (eth)	Total Execution Cost (s)
Blockchain FL [**16**]	[n ×20, n ×30]	4 150 540	2.2700
Smart Contract DF [**18**]	[n ×24, n × 32]	3 320 656	1.6808
Proposed Blockchain EL	[n ×64, n ×128]	5 519 242	2.7100

In conclusion, for e-healthcare applications:

- Blockchain FL [**16**] is ideal for fast, low-latency use cases, such as IoT-enabled healthcare monitoring.
- Smart Contract DF [**18**] is best suited for distributed systems with minimal infrastructure and low-cost requirements.
- Proposed Blockchain EL is optimal for security, feature-rich systems where execution costs are a secondary concern.

4.5 Limitations

The model was mainly trained and validated on CICIoMT2024 dataset, which mainly contains data from critical patients. To improve the generalizability of the model across different user groups, it is important to include additional data covering a spectrum of age, gender, skin color, and health status. This will ensure the adaptability of the model to a wider population. It is anticipated that future advances will enable real-time data capture, data pre-processing and attack prediction using deep learning, overcoming the limitations of ensemble models and offline learning tasks.

5. Conclusion

The integration of blockchain technology with ensemble models provided e-healthcare system with enhanced security and prediction of attacks. In this paper, a hybrid approach was proposed to detect DDoS attacks using ensemble models based on generated traffic data, efficiently meets all quality standards of the original data. Moreover, standard metrics were used to validate the ensemble models performance in terms of accuracy, precision, recall and F1-score. Additionally, data generated is trained using standard ensemble models through binary classification, which classify sensed data as malicious or legitimate. This approach aimed to evaluate ensemble models to achieve accurate predictions, reduce attacks and store validated transitions in the blockchain. The CICIoMT2024 dataset is employed the training and validation based on three different ensemble models: XGBoost and RF models also were compared. The ensemble models showed remarkable effectiveness in predicting attacks. Future work will also include applying blockchain and deep neural networks like RNNs and transformers to enhance prediction accuracy and ensure effective analysis.

References

1. S. Gupta, B. Yadav, B. Gupta, "Security of IoT-based e-healthcare applications using blockchain," in Technology for Cyber Physical Systems, Springer, 2022.
2. C. Butpheng, K. H. Yeh, and H. Xiong, "Security and privacy in IoT-cloud-based e-health systems—A comprehensive review," Symmetry, 2020.
3. P. Thanalakshmi, A. Rishikhesh, J. Marion Marceline, G. P. Joshi, and W. Cho, "A quantum-resistant blockchain system: A comparative analysis," Mathematics, vol. 11, no. 18, 2023.
4. S. Team, "9-14 network outage: Initial overview," https://solana.com/fr/ news/9-14-network-outage-initial-overview, July 2024, accessed: 2024- 08-30.
5. Halborn, "How blockchain DDoS attacks work," https://www.halborn.com/blog/how-blockchain-ddos-attacks-work, October 2021, accessed: 2024- 08-30.
6. S. Dadkhah, E. Carlos Pinto Neto, R. Ferreira, R. Chukwuka Molokwu, S. Sadeghi, and A. Ghorbani, "CICIoMT2024: Attack vectors in healthcare devices-a multi-protocol dataset for assessing IoMT device security," Internet of Things, vol. 28, 2024, 101351, https://doi.org/10.1016/j.iot.2024.101351
7. H. Reffad, A. Djenaoui, A. Alti, "Distributed Secure Services Based on IoT and Blockchain for e-Health Remote Care," *Proc. Int. Conf. Computer Science's Complex Systems and Their Applications, Oum El Bouaghi (Algeria)*. CEUR-WS.org, 2021.
8. G. Wu, H. Wang, X. Lai, M. Wang, D. He, and S. Chan, "A comprehensive survey of smart contract security: State of the art and research directions," Journal of Network and Computer Applications, vol. 226, p. 103882, 2024.
9. R. Kumar, P. Kumar, R. Tripathi, G. P. Gupta, S. Garg, and M. M. Hassan, "A distributed intrusion detection system to detect DDoS attacks in blockchain-enabled IoT network," Journal of Parallel and Distributed Computing, vol. 164, pp. 55–68, 2022.
10. M. Saad, L. Njilla, C. Kamhoua, J. Kim, D. Nyang, and A. Mohaisen, "Mempool optimization for defending against ddos attacks in pow- based blockchain systems," in 2019 IEEE international conference on blockchain and cryptocurrency (ICBC). IEEE, 2019, pp. 285–292.
11. S. Lakshminarasimman, S. Ruswin, and K. Sundarakantham, "Detecting DDoS attacks using decision tree algorithm," *IEEE fourth international conference on signal processing, communication and networking (ICSCN)*, pp. 1-6, 2017.

12. J.M. Nsaif, M.T. Gaata, "K-Means clustering-based semi-supervised for DDoS attacks classification." Bulletin of Electrical Engineering and Informatics 11.6 (2022): 3570-3576.

13. C. Kim, S.Y. Chang, D. Lee, J. Kim, K. Park, J. Kim, "Reliable detection of location spoofing and variation attacks," *IEEE Access* 11 (2023): 10813-10825.

14. K. G. Arachchige, P. Branch, and J. But, "An analysis of blockchain-based iot sensor network distributed denial of service attacks," Sensors, vol. 24, no. 10, 2024.

15. K. Dwivedi, A. Agrawal, A. Bhatia, and K. Tiwari, "A novel classification of attacks on blockchain layers: Vulnerabilities, attacks, mitigations, and research directions," arXiv preprint arXiv:2404.18090, 2024.

16. W. Moulahi, I. Jdey, T. Moulahi, M. Alawida, A. Alabdulatif. "A blockchain-based federated learning mechanism for privacy preservation of healthcare IoT data." *Computers in Biology and Medicine*, vol. 167, 2023,107630.

17. M. Alsayegh, T. Moulahi, A. Alabdulatif, P. Lorenz. "Towards secure searchable electronic health records using consortium blockchain." *Network*, vol. 2, no. 2, pp., 239-256, 2022.

18. A. Alabdulatif, M. Al Asqah, T. Moulahi, S. Zidi. "Leveraging artificial intelligence in blockchain-based e-health for safer decision-making framework." *Applied Sciences*, vol. 13, no. 2, 2023, 1035.