JCBI
Journal of Computing & Biomedical Informatics Volume 04 Issue 01
ISSN: 2710 - 1606 2022

Research Article
https://doi.org/10.56979/401/2022/80

Cyberattacks Detection in IoMT using Machine Learning Techniques

Haseeb Tauqeer[1], Muhammad Munwar Iqbal[1]* Aatka Ali[2], Shakir Zaman[1] and Muhammad Umar Chaudhry[3]

[1]Department of Computer Science, University of Engineering and Technology Taxila, Pakistan.
[2] Department of Computer Science, Air University Islamabad, Multan Campus, Multan.
[3] Department of Computer Science, MNS-University of Agriculture, Multan.
*Corresponding Author: Muhammad Munwar Iqbal. Email: munwariq@gmail.com.

Received: September 11, 2022 **Accepted:** November 11, 2022 **Published:** December 29, 2022.

Abstract: Information and Communication Technology (ICT) has changed the computing paradigm. Various new channels for communication are created through these developments, and the Internet of Things (IoT) is one of them. Internet of Medical Things (IoMT) is a part of IoT in which medical devices are connected through a network. IoMT has resolved many traditional health-related problems and has some security concerns. This article uses three Machine Learning algorithms, Random Forest, Gradient Boosting, and Support Vector Machine (SVM), to detect cyberattacks. Machine Learning models are best for performing cyberattack detection. Proposed Machine Learning models are evaluated on the WUSTL EHMS 2020 dataset, which consists of main in-the-middle, data injection, and spoofing attacks. The evaluation of the result analysis shows that the proposed Machine Learning models outperformed existing techniques.

Keywords: IoMT, Security, Medical Devices, ML, Random Forest, Gradient Boosting, SVM.

1. Introduction

The Internet of Things (IoT) is a framework comprising physical devices operated by various sensors or software. These devices are operated to make sure they connect with other devices and share essential information between them. IoT is one of the most evolving technology, and almost every industry is taking benefits from IoT. Medical is one of the most common sectors that use IoT to make the working of a medical center or hospital as smart as IoT works. IoT in healthcare is known as the Internet of Medical Things (IoMT) [1].

IoMT, known as Medical IoT or Healthcare IoT, is a group of medical applications and devices connected with the healthcare system through online servers. IoMT applications are very beneficial for hospitals, where doctors, nurses, or medical staff constantly use these devices to check out their patients all day and night [2]. With the help of IoMT, there is no need for patients to make a physical appearance in the hospitals because the doctors can quickly examine them by medical applications and continue their treatment online. IoMT is a game changer for the medical industry because it drops medical costs and improves the quality of patient care.

IoMT devices provide many benefits, but also it has security issues. Many scenarios are discussed in which hackers or cybercriminals can access data or hijack the network. Medical devices are connected to various applications, allowing hackers to access valuable data. Most medical devices depend on open-source wireless communication, by which network attacks can easily damage the network [3]. In a survey in 2020, it was proposed that 34% of medical institutes become prey to cybercriminals or hackers. In this single year, almost 18 million patients' record was affected by attackers. These days, prominent medical

institutes are attacked by attackers, and many small medical clinics are at risk too. So it is imperative to come up with some solutions that are used to protect the medical institutes and their valuable data stored on different clouds.

Many approaches are used to counter various attacks in the IoMT environment. In this research article Machine learning-based approach is used. Machine Learning is a sub-branch of Artificial Intelligence (AI) that focuses on data and algorithms to provide humans with an environment for learning. Machine Learning is pivotal in cybersecurity because it can quickly analyze millions of records. While providing security, the machine learning algorithms continuously monitor the network or learn from the records. Then ML algorithms find the patterns from these learnings, which are used to detect cyber-attacks and predict where malware nodes are online to provide safe browsing for users [4].

Machine Learning (ML) has several algorithms used for classification and regression. These algorithms are used to protect the network from different attacks. This article uses three ML-based algorithms: Random Forest (RF), Gradient Boosting (GB), and SVM. Random Forest solves each problem by creating a Decision Tree. Random Forest uses the Ensemble learning approach. It is an approach that is used to combine many classifiers to solve complex problems [5]. Gradient Boosting relied on the possibility of reducing prediction error by combining the previous model with the new model [6]. The main goal of the SVM algorithm is to divide the classes of datasets in such a way as to find the best Maximum Marginal Hyperplane (MMH). MMH is the optimal line between two classes used as a separator to increase the distance between those classes [7].

WUSTL EHMS 2020 dataset is used in this research which contains medical records. This dataset has network and biometric features and a class label to distinguish between normal and attack nodes. Three types of attacks are used in this dataset: data injection, main in the middle, and spoofing attacks in the IoMT environment [8]. We applied some preprocessing steps to the dataset in the proposed research, like cleansing and feature selection. And then, Random Forest, Gradient Boosting, and SVM are applied to preprocessed datasets one by one. Then analyze the results of both algorithms and compare them with existing models.

2. Literature Review

The authors [9] have said that IoT plays a significant role. Moreover, the IoMT is a particular type of IoT in which medical sensors, devices, and applications are connected. IoMT helps the healthcare department and also provides better care for its patients. With many benefits, the IoMT has some significant security and privacy concerns from attackers who get access to stored data. In this article, they proposed Deep Neural Network (DNN) based Intrusion Detection System (IDS) to avoid Sybil attacks. The authors said the proposed model gains 15% higher accuracy than existing models.

IoMT addresses a considerable number of limitations in the traditional healthcare system. Quality of treatment is enhanced by IoMT and improves patients' health. But critical fault lines of the IoMT network were exposed when a cyber-attack on an Indiana hospital in 2018 happened. So to come up from those attacks, the authors [10] proposed an ensemble learning and fog computing-based architecture for the security of the IoMT network. The authors have said that most existing approaches are evaluated on NSL-KDD and KDD CUP99 datasets. They used a realistic dataset named ToN-IoT because those datasets have a shortage of recent IoMT attacks. In the results, they indicated that their model gained 96% of accuracy.

The advancement in ICT has changed the complete structure of the computing environment. Through these advancements, many new channels are created. IoT is one of them. IoT is emerging to provide an intelligent architecture to almost every field. IoMT is the part of IoT where medical applications can communicate to benefit patients and medical staff. Security is becoming a hurdle between the IoMT environment and its users. So, the authors utilized the Deep Recurrent Neural Network (DRNN) and Supervised Machine Learning (SML) models to create an effective IDS system for the IoMT environment. In the analysis, the authors show that the proposed model outperformed existing techniques [11].

In recent years Supervisory Control and Data Acquisition (SCDA) have been used in power grids and remote monitoring systems. But, these SCDA systems always prey to various cyberattacks, and many solutions are also not applicable to these systems because of their distinct behavior. So, in this paper, the authors [12] used Deep Learning based approach to identify cyberattacks that are used to target SCDA systems. Feature importance of cyberattack detection is also performed in this study, along with malicious

attack detection. According to the authors, the proposed model outperformed form traditional Deep Learning algorithms, including Random Forest, SVM, OneR, Naïve Bayes, and Adaboost.

The authors [13] have proposed that healthcare applications are achieving popularity in Smart Cities with the help of IoMT. With remote data access, these applications provide benefits to both patients and medical staff. The authors have said that many research solutions are becoming dangerous for information leakage and damaging the network. So to counter those faulted solutions, they proposed an SDN-based model to improve the data delivery rate. Furthermore, they proposed centralized-based Software Defined Networking (SDN) architecture to overcome the threats between sensors. They also used the Unsupervised Machine Learning approach to reduce network communication overhead. Results showed that this model increases the system's performance.

Nowadays, Machine Learning based IoMT environments are overgrowing for healthcare environments. Machine Learning is essential for the IoMT environment to detect malicious nodes. However, some traditional learning architectures are becoming critical for IoMT. So in this article, the authors [14] proposed the Federated Learning-based Blockchain Enabled Task Scheduling (FL-BETS) framework. The main purpose of the FL-BETS is to ensure the security protection and analysis of fraud data at different levels. Some mathematical models are introduced in this study. In the end, they anticipated that the proposed model outperformed all traditional ML and Blockchain models.

The introduction of IoT systems to medical applications has allowed the medical staff to handle patients remotely. Nevertheless, security is a big challenge because the information and confidentiality of patients are at high risk. But when it comes to intrusion detection, Machine Learning approaches to provide practical solutions. Most IoMT datasets have biometric or network features, but the dataset used in this article combines both. The dataset is created with Enhanced Healthcare Monitoring Systems (EHMS) testbed, which monitors the network and biometric features. And then, the authors used different ML models to test and train the dataset. In experimental analysis, the authors show that the proposed model improved the performance from 7 % to 15 % [8].

IoT is developing increasingly at present but also faces some serious threats for Distributed Denial of Service (DDoS) attacks. And also, the existing security solutions are very expensive. So to counter this, in this article, the authors proposed an Adaptive Machine Learning SDN-enabled DDoS attacks detection and mitigation (AMLSDM) framework. The proposed model uses different Machine Learning approaches with a multilayer feed-forwarding approach. In the first layer, SVM, Random Forest, Naïve Bayes, KNN, and Logistic Regression are used to detect DDoS attacks. The Ensemble Voting algorithm accumulates the first layer's performance, which takes input from the first layer. The third layer performs the measurement of live network traffic for the detection of DDoS attacks. In experimental analysis, the authors showed that the proposed model achieved higher accuracy than other state-of-the-art approaches[15].

The authors [16] have said that cyberattacks are more dangerous due to the excessive use of IoT networks. The use of fog computing in IoT networks provides security from all those malicious attacks. Fog computing aims to detect attacks on IoT networks in a short period. However, the problem with fog computing is that it does not have enough resources to detect the attacks on time. So in this paper, the authors proposed an Ensemble Machine Learning model to detect attacks on fog devices. The NSL-KDD is used to evaluate this model. In the result analysis, the authors used different evaluation metrics to show the effectiveness of their model.

The rapid growth in the Internet of Medical Things (IoMT) provides multiple healthcare benefits. From all those benefits, the IoMT network still has significant security and privacy concerns because patients' health data is at risk. Evaluation of malware attacks is becoming a big issue that needs to be resolved. So in this paper, the authors [17] proposed Intrusion Detection and Prevention System (IDPS). The proposed model relies on Hypertext Transfer Protocol (HTTP) and Transmission Control Protocol (TCP) protocols. HTTP is adopted by ICT healthcare applications, while IoMT adopts the TCP protocol. The experimental results show the effectiveness against HTTP and TCP attacks.

The authors [18] deliberated that IoMT is a biological network for healthcare applications. It is moderately improving the healthcare industry by enabling the flawless communication of medical devices. IoMT aims to deliver the finest quality medical services at a minimum price. With the advancement in IoMT, it also contains privacy and security concerns. Because of these privacy and security concerns, the patients may phase data damage. To ensure the security of IoMT, the authors proposed Bayesian

Optimization and Extreme Learning Machine (ELM). In the end, the authors showed that this model achieved higher accuracy than old approaches.

The architecture of the IoMT is the interconnection of medical applications. It has effectively solved several traditional medical-related glitches. It also has some downsides, like the patient's critical information being in danger due to a lack of security. So there is a need for efficient solutions for IoMT devices to provide security to the network. In the past, Machine Learning algorithms played a vital role in detecting attacks. But now a day, the investigation through Machine Learning algorithms is reduced. So to fill that gap, the authors [19] projected a Machine Learning model for the IoMT network to detect intrusions. Various ML models are used in the proposed model, like K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Artificial Neural Network (ANN). These ML models were evaluated on the Bot-IoT dataset. The best approach is to build efficient IDS that can detect attacks from the IoMT network by comparing the proposed ML approaches' results.

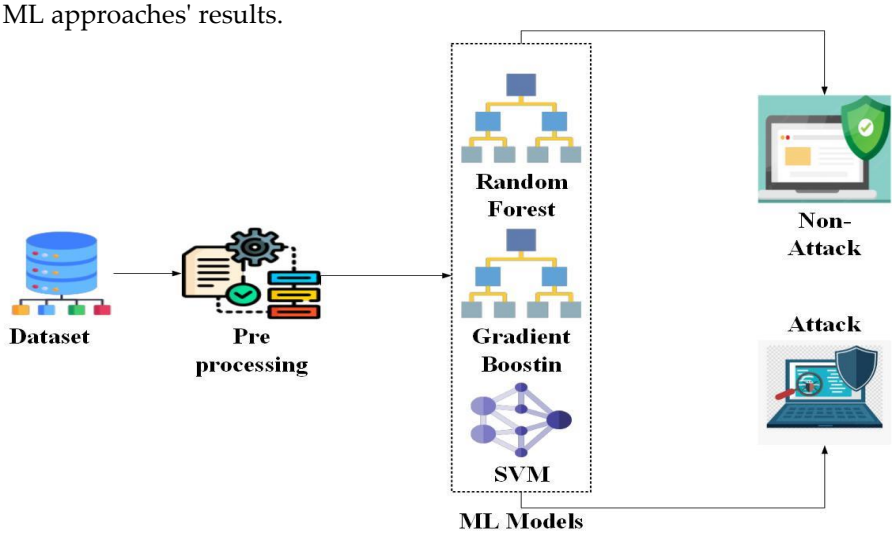

Figure 1. Proposed a diagram to detect cyberattacks

3. Proposed Model

This section is a brief introduction to our proposed model. First, an IoMT dataset with normal and attack records is collected. Then some preprocessing steps are followed on this dataset in order to get the best possible results. Then three Machine Learning models, RF, GB, and SVM are applied to this dataset.

Figure 1 shows the architecture of proposed ML models to detect cyberattacks. The proposed model's first step is to get the WUSTL EHMS 2020 dataset and apply preprocessing steps. Then preprocessed dataset is given to proposed Machine Learning algorithms, and then evaluation of these algorithms is performed by Accuracy, Precision, Recall, F1 Score, and Loss.

3.1. Dataset

Researchers of Washington University in St. Louis (WUSTL) created a dataset using Enhanced Healthcare Monitoring System (EHMS) testbeds named WUSTL EHMS 2020 is used in this research. The dataset has 16,318 records, of which 14,272 records are normal and 2046 are attacks. This dataset has both network and biometrics features and one class label. Attacks used in this dataset are spoofing, data injection, and main in the middle. Preprocessing steps, cleansing, and feature selection are performed on this dataset.

3.2. Preprocessing of Dataset

Preprocessing of the dataset is essential because the data it contains is processed effectively. Two preprocessing steps are followed in this proposed research, cleansing and feature selection. Simple python code is used for cleansing. First, the dataset is loaded using the panda library and performed with different functions. Then Dir and Flags features of this dataset are deleted because these features have no related records. The Feature Selection step is performed using the Mutual Information function. By using this function, only those features that have non-negative values are selected for processing. After selecting those features, the preprocessed dataset is provided to ML algorithms for training and testing.

3.3. Machine Learning models to detect Cyberattacks in IoMT

This article proposed three Machine Learning based models to detect various cyberattacks (main in the middle, spoofing, and data injection) in the IoMT environment. These models are given below:

3.3.1. Random Forest

Random Forest is also known as an ensemble learning model. Decision trees are built on numerous samples to hold the majority vote for classification and the average for regression. The main benefit of Random Forest is that this algorithm can handle both categorical and continuous variables in the dataset, one at classification and the other at regression. It is also known as the bagging technique. Still, the difference is bagging classifier use all features of the dataset, and the Random Forest uses few features according to the nature of the problem.

3.3.2. Gradient Boosting

Gradient Boosting is usually known for its accuracy and prediction speed with large datasets. It reduces the overall prediction error by combining the next possible model with the previous model. Gradient Boosting is also used as a bagging classifier. The difference between bagging and boosting is that bagging is used for over-fitting, while boosting reduces bias. The contradiction between RF and GB is that the former creates the trees independently and later creates one tree at a time.

3.3.3. SVM

SVM is commonly used for classification problems, and the Kernel Function performs the transformation of data in SVM. This function takes raw data as input and transforms it into processing data. The hyperplane is used by SVM to divide datasets into two classes. These classes are divided by an optimal line known as the Maximum Marginal Hyperplane (MMH).

3.4. Evaluation Metrics

Accuracy, Precision, Recall, F1 Score, and Loss are the evaluation metrics used by Machine Learning models to evaluate performance. Accuracy demonstrates that the Machine Learning model predicts the correct value. At the same time, precision is how well the proposed model predicts the correct value. The recall is how the ML model correctly classifies positive values. F1 score measures the accuracy of the model on a dataset. Loss is used to showing the ominous predictions of the model. Loss is also known as Mean Square Error (MSE). The equations for all the above metrics are given below.

$$Accuracy=\frac{TP+TN}{Total\ Prediction} \tag{1}$$

$$Precision=\frac{TP}{TP+FP} \tag{2}$$

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

$$F1\ Score=2*\frac{Precision*Recall}{Precision+Recall} \tag{4}$$

$$MSE=\frac{1}{n}\sum_{i=1}^{n}(Yi-\bar{Yi})^2 \tag{5}$$

4. Results

This section discussed the results of the proposed Machine Learning models. Random Forest, Gradient Boosting, and SVM are the proposed models to detect cyberattacks in the IoMT environment. The performance of these models is evaluated using Accuracy, Precision, Recall, F1 Score, and Loss.

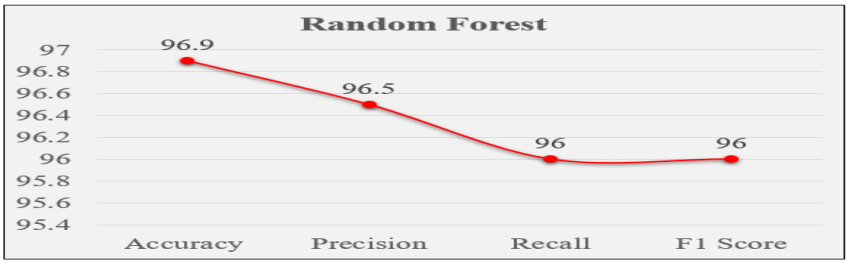

Figure 2. Performance metrics for Random Forest

Figure 2 shows the performance metrics of the proposed Random Forest model. In this model, we used 30 estimators and completed them with the best possible results. In comparison, showed that the proposed model got higher results. The proposed Random Forest model achieved 96.9% of accuracy, while other metrics are 96.5%, 96%, and 96%, successively.

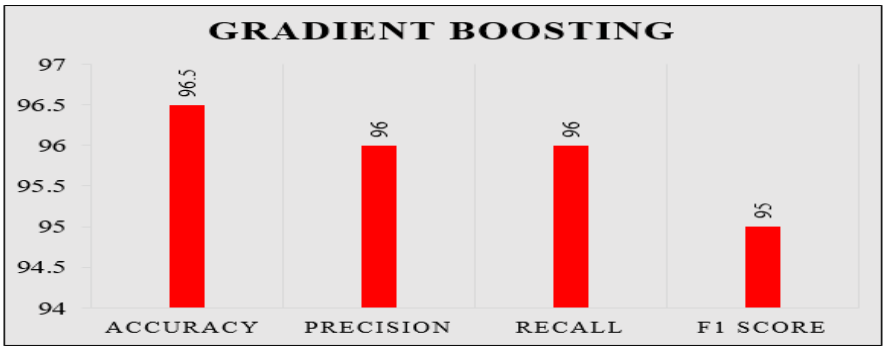

Figure 3. Performance Metrics for Gradient Boosting

Figure 3 demonstrates the performance measurement metrics for Gradient Boosting. It is used 37 estimators and a 0.05 learning rate to get desired results. Proposed Gradient Boosting achieved 96.5% of accuracy. Furthermore, the precision is 96%, recall is 96%, and the F1 score is 95%.

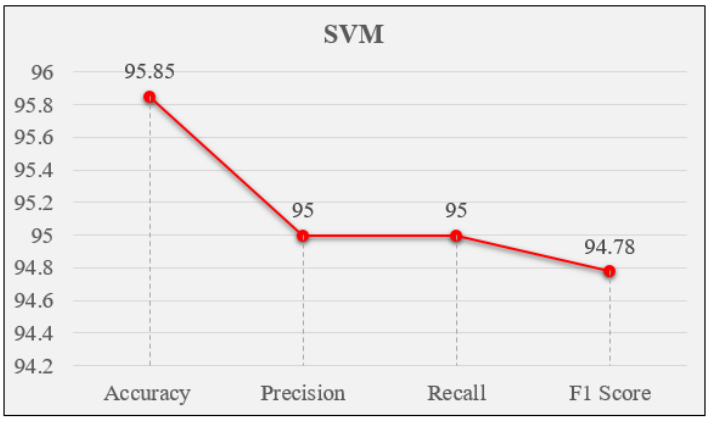

Figure 4. Performance metrics for SVM

Figure 4 shows the performance metrics for the proposed SVM model. RBF kernel is used in the proposed model, also known as the linear kernel. The proposed SVM model achieved 95.85% accuracy and precision, and recall and F1 scores were 95%, 95%, and 94.78%, respectively.

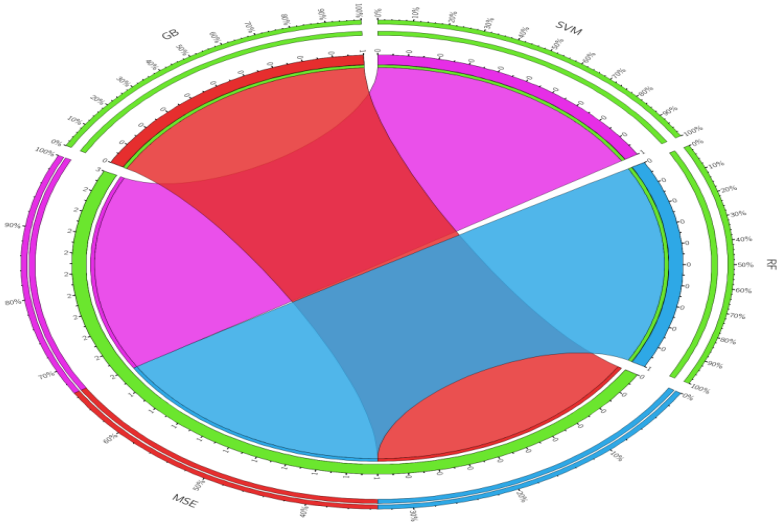

Figure 5. Mean Square Error comparison for ML models

Figure 5 shows the Mean Square Error (MSE) for proposed Machine Learning models. MSE is also known as Loss rate. MSE is used to show the bad predictions of the model. The model is considered as good if its value is near zero. The Loss rate for the proposed Random Forest, Gradient Boosting, and SVM are 0.0295, 0.0368, and 0.0561, respectively.

4.1. Comparison of Proposed Models with Existing Techniques

Figure 6 shows the performance comparison between proposed Machine Learning models with existing techniques. It clearly shows that the proposed RF and GB outperformed the traditional model with an accuracy rate of 96.9% and 96.5%. The proposed SVM model also provides better accuracy only behind Kumar et al. [10], which is 95.85. Grammatikis et al. [17] used different ML models from all the Random Forests and achieved a higher accuracy rate of 94.45%. Hady et al. [8] used WUSTL EHMS 2020 and evaluated it on different Machine Learning algorithms from which the Random Forest model gained the best accuracy of 92.27%. We also used that dataset, but our results were higher than Hady et al. [8] because of preprocessing steps followed on the dataset. Kumar et al. ToN-IoT dataset and performed various Machine Learning models. On the first level, the authors used Decision Tree, Naïve Base and Random Forest; on the other level, they used XGBoost and achieved an accuracy of 96.35%.

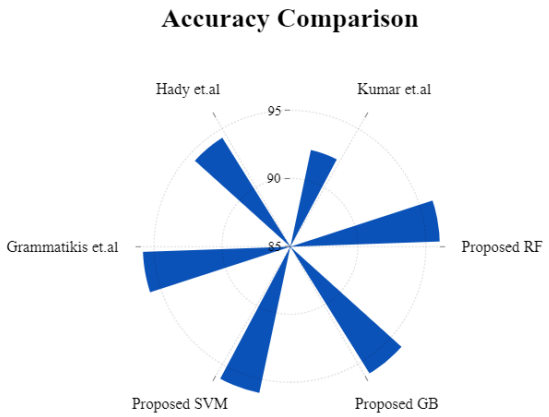

Figure 6. Accuracy comparison between the proposed model and existing technique

5. Discussion

We use different Machine Learning models in this research work to detect cyberattacks. These Machine Learning models are evaluated on WUSTL EHMS 2020 and achieved the best results compared to old approaches which are shown in Figure 6. Before this research, WUSTL EHMS 2020 was evaluated on different Machine Learning algorithms from which Random Forest got 92.27% of accuracy, higher than all [8]. In comparison, we demonstrate that this accuracy is less than the proposed models used in this research because of preprocessed steps followed on the dataset.

6. Conclusion

In this research article, we used three Machine Learning based algorithms, Random Forest, Gradient Boosting, and SVM. Machine Learning based models are used because they detect cyberattacks efficiently. These Machine Learning models are evaluated on WUSTL EHMS 2020 dataset, which contains both the network and biometric features. This dataset contains data injection, spoofing, and main in-the-middle attacks. The dataset is passed through preprocessing steps, Cleansing, and Feature Selection in order to achieve higher results. Previously, when Machine Learning models were applied to this dataset, it achieved 92.27% accuracy with Random Forest. Following preprocessing steps, the proposed Machine Learning models achieved higher accuracy with 96.9% for Random Forest, 96.5% for Gradient Boosting, and 95.85% for SVM. Some of the future directions that can enhance this research are another authentic dataset that should be used to enhance the performance of the models. Deep Learning algorithms can be applied to this dataset to get more efficient results than the proposed models. The proposed model can also be used for IoT-based approaches like Smart Home and Industry 5.0.

Conflict of Interest: The authors stated that they have no conflicts to report about this study.

References

1. S. Vishnu, S. J. Ramson, and R. Jegan, "Internet of medical things (IoMT)-An overview," in 2020 5th international conference on devices, circuits and systems (ICDCS), 2020, pp. 101-104.
2. A. H. M. Aman, W. H. Hassan, S. Sameen, Z. S. Attarbashi, M. Alizadeh, and L. A. Latiff, "IoMT amid COVID-19 pandemic: Application, architecture, technology, and security," Journal of Network and Computer Applications, vol. 174, p. 102886, 2021.
3. S. C. Sethuraman, V. Vijayakumar, and S. Walczak, "Cyber attacks on healthcare devices using unmanned aerial vehicles," Journal of medical systems, vol. 44, pp. 1-10, 2020.
4. I. H. Sarker, A. Kayes, S. Badsha, H. Alqahtani, P. Watters, and A. Ng, "Cybersecurity data science: an overview from machine learning perspective," Journal of Big data, vol. 7, pp. 1-29, 2020.
5. A. B. Shaik and S. Srinivasan, "A brief survey on random forest ensembles in classification model," in International Conference on Innovative Computing and Communications, 2019, pp. 253-260.
6. A. S. Dyer, D. Zaengle, J. R. Nelson, R. Duran, M. Wenzlick, P. C. Wingo, et al., "Applied machine learning model comparison: Predicting offshore platform integrity with gradient boosting algorithms and neural networks," Marine Structures, vol. 83, p. 103152, 2022.
7. W. Xie, Y. She, and Q. Guo, "Research on multiple classification based on improved SVM algorithm for balanced binary decision tree," Scientific Programming, vol. 2021, 2021.
8. A. A. Hady, A. Ghubaish, T. Salman, D. Unal, and R. Jain, "Intrusion detection system for healthcare systems using medical and network data: A comparison study," IEEE Access, vol. 8, pp. 106576-106584, 2020.
9. S. P. RM, P. K. R. Maddikunta, M. Parimala, S. Koppu, T. R. Gadekallu, C. L. Chowdhary, et al., "An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture," Computer Communications, vol. 160, pp. 139-149, 2020.
10. P. Kumar, G. P. Gupta, and R. Tripathi, "An ensemble learning and fog-cloud architecture-driven cyber-attack detection framework for IoMT networks," Computer Communications, vol. 166, pp. 110-124, 2021.
11. Y. K. Saheed and M. O. Arowolo, "Efficient cyber attack detection on the internet of medical things-smart environment based on deep recurrent neural network and machine learning algorithms," IEEE Access, vol. 9, pp. 161546-161554, 2021.
12. W. Wang, F. Harrou, B. Bouyeddou, S.-M. Senouci, and Y. Sun, "A stacked deep learning approach to cyber-attacks detection in industrial systems: application to power system and gas pipeline systems," Cluster Computing, vol. 25, pp. 561-578, 2022.
13. K. Haseeb, I. Ahmad, I. I. Awan, J. Lloret, and I. Bosch, "A machine learning SDN-enabled big data model for IoMT systems," Electronics, vol. 10, p. 2228, 2021.
14. A. Lakhan, M. A. Mohammed, J. Nedoma, R. Martinek, P. Tiwari, A. Vidyarthi, et al., "Federated-Learning Based Privacy Preservation and Fraud-Enabled Blockchain IoMT System for Healthcare," IEEE Journal of Biomedical and Health Informatics, 2022.
15. M. Aslam, D. Ye, A. Tariq, M. Asad, M. Hanif, D. Ndzi, et al., "Adaptive Machine Learning Based Distributed Denial-of-Services Attacks Detection and Mitigation System for SDN-Enabled IoT," Sensors, vol. 22, p. 2697, 2022.
16. V. Tomer and S. Sharma, "Detecting IoT Attacks Using an Ensemble Machine Learning Model," Future Internet, vol. 14, p. 102, 2022.
17. P. Radoglou-Grammatikis, P. Sarigiannidis, G. Efstathopoulos, T. Lagkas, G. Fragulis, and A. Sarigiannidis, "A self-learning approach for detecting intrusions in healthcare systems," in ICC 2021-IEEE International Conference on Communications, 2021, pp. 1-6.
18. J. Nayak, S. K. Meher, A. Souri, B. Naik, and S. Vimal, "Extreme learning machine and bayesian optimization-driven intelligent framework for IoMT cyber-attack detection," The Journal of Supercomputing, pp. 1-26, 2022.
19. A. Binbusayyis, H. Alaskar, T. Vaiyapuri, and M. Dinesh, "An investigation and comparison of machine learning approaches for intrusion detection in IoMT network," The Journal of Supercomputing, pp. 1-20, 2022.