IJIRIS: International Journal of Innovative Research in Information Security
Volume 10, Issue 04, May 2024
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives

Securing IoT Healthcare Data Using Machine Learning Techniques

Heena Kousar*
Department of Computer Science & Engineering
East Point College of Engineering & Technology
Bangalore, INDIA
heena.k@eastpoint.ac.in
Shammi L
Department of Computer Science & Engineering
East Point College of Engineering & Technology
Bangalore, INDIA
shammil.cse@eastpoint.ac.in
Madhushree R
Department of Computer Science & Engineering
East Point College of Engineering & Technology
Bangalore, INDIA
madhushree.r@eastpoint.ac.in

Publication History
Manuscript Reference No: IJIRIS/RS/Vol.10/Issue04/MYIS10115
Research Article | Open Access | Double-Blind Peer-Reviewed | Article ID: IJIRIS/RS/Vol.10/Issue04/MYIS10115
Received: 07, April 2024 | Revised:16, April 2024 | Accepted: 27, April 2024 Published Online: 07, May 2024 Volume2024
Article ID MYIS10115 http://www.ijiris.com/volumes/Vol10/iss-04/36.MYIS10115.pdf
Article Citation: Heena,Shammi,Madhushree(2024). Securing IoT Healthcare Data Using Machine Learning Techniques. International Journal of Innovative Research in Information Security, Volume 10, Issue 04, Pages 303-307
doi:> https://doi.org/10.26562/ijiris.2024.v1003.36
BibTex key: Heena@2024Securing

Abstract: The internet of Things has become a crucial technology, enabling the interconnection of everyday devices like home appliances, medical equipment and vehicles. However, the sensitive data handled by IoT systems, particularly in the healthcare sector, requires robust security measures to prevent tampering or theft. Detecting security threats in IoT environments requires sophisticated techniques. Different machine learning algorithms are used to identify the attacks in IoT network traffic and predict snooping behavior based on unidentified patterns. This paper proposed a method to identify network traffic attacks by employing different strategies utilizing the NT-ToN-IoT dataset. The classifier utilized Naïve Bayes(NB), Random Forest(RF), Support Vector Machine, Artificial Neural Network. These algorithms favoured over centralized methods to construct streamlined security system for IoT devices. Prior to analysis the dataset underwent preprocessing to eliminate irrelevant or missing data. Following this, a feature engineering approach is applied to extract crucial features. The results from applying each classifier demonstrated a significant maximum classification accuracy of 98% achieved by the RF model.

Keywords: IoT (Internet of Things), ML (Machine Learning), RF(Random Forest), ANN (Artificial Neural Network)

I. INTRODUCTION

Integration of the Internet of Things (IoT) in healthcare significantly improved monitoring of patient, cost reduction and innovation in the patient care i,e it has provide various numerous opportunities. However the security and privacy of healthcare data remain critical concerns due to the complexity and heterogeneity of IoT devices, networks and systems. Some of the key challenges are

1. Data Privacy and Confidentiality: Though there is a need to protect massive health data generated by IoT devices from unauthorized access, misuse or breach, it's difficult to ensure compliance with data privacy regulations such as GDPR and HIPAA while simultaneously allowing for data sharing and analysis in healthcare applications.
2. Distributed and Heterogeneous Nature of IoT: Healthcare systems' entire ecosystem has many IoT devices that are decentralized and heterogeneous in nature hence security policies that are constant across all these different IoT devices poses a challenge.
3. Vulnerability to Cyber Attacks: The presence of limited computational resources in IoT devices makes them susceptible to a variety of cyber threats including malware, DDoS attacks and man-in-the-middle attacks. However, ensuring the safety of these devices along with their communication channels can be tough.
4. Data Integrity and Authenticity: Health data gathered from IoT gadgets must be both authentic and have integrity to facilitate reliable decision-making processes as well as patient care. Nevertheless, the integrity of the information may be compromised through tampering with data, spoofing or injection attacks.
5. Scalability and Interoperability: In addition to scalable solutions for security that can handle increased number of Internet of Things (IoT) devices as well as growing volumes of information being generated.

6. Machine Learning Vulnerabilities: ML algorithms used for security and data analysis in healthcare can be susceptible to adversarial attacks, data poisoning, and model extraction attacks, which can compromise their effectiveness and lead to incorrect decisions or data breaches.

In [1], the author presents a model for the Secured Big Data analytics using Edge–Cloud architecture (SBD-EC), which aims to provide distributed and timely computation of a decision-oriented medical system. In order to improve trust levels for integrated health sensors, a model of the IoMT with Big Data Analytics using Edge Cloud architecture has been developed. The computed cost function using greedy heuristics decreases the data retrieval rate based on mobile edges and improves the management of big data analytics. In addition, it is using cost effective security calculations to reduce overheads at the edge of the Internet of Things network for sinks and mobile phones. However, it lacks with the scalability as the number of mobile users increases and connects to multiple clouds.
In order to ensure the security of medical data, [2] the authors propose a decentralized cloud environment using blockchain technology. It provides electronic signature systems for the processing of data. Moreover, the technology of Blockchain allows for data to be divided in blocks and timestamps are created with observation data that has been kept or updated. For the handling of data, it offers digital signature systems. Compared to existing schemes, the proposed work shall improve performance in terms of cost and time.When crafting and deploying IoT solutions, it's essential to prioritize the management of data and information to mitigate potential security risks. Safeguarding data is a primary focus for network devices. Security takes precedence in the IoT domain because unauthorized access or tampering with IoT devices, especially in critical applications, poses significant risks to human life [4].

The IoT environment deals with many private and sensitive health data that must be kept safe from tampering or theft. If safety precautions are not implemented, these dangers and assaults against IoT devices in the health sector might completely destroy this industry. Detecting security threats to an IoT environment requires sophisticated technology; these attacks can be identified using machine learning (ML) techniques, which can also predict snooping behavior based on unidentified patterns. The contributions of our technology are outlined below, as previously elaborated:

The paper proposes a machine learning based approach to detect attacks on IoT network traffic. The researchers utilized the NF-ToN-IoT dataset which contains 1,379,274 network flow samples with 80.4% being attack samples and 19.6% benign. Proposing an optimal Machine Learning (ML) model for intrusion detection in cyber-attacks entails utilizing the NF-ToN-IoT dataset and applying classification techniques such as Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), and Naive Bayes (NB) algorithms.

The subsequent sections of this research are structured as follows: Literature Review: Section II summarizes the existing literature on intrusion detection utilizing machine learning and deep learning techniques with diverse datasets. Model Architecture: Section III offers a detailed explanation of the proposed methodology and architecture. Implementation and Result: Section IV presents the implementation and Results of the machine learning algorithms utilized in this study, along with the results obtained from each. Results and Discussion: Finally, Sections V present the conclusion drawn from this research and outline potential directions for future work.

II. LITERATURE REVIEW

The study [5] primarily investigated the impact of machine learning on flow-based anomaly detection in Software-Defined Networking (SDN). The authors introduced two distinct approaches to intrusion detection systems, leveraging deep neural networks (DNN) and machine learning. In the first approach, the NSL-KDD dataset was utilized, and feature selection using the Random Forest (RF) classifier yielded an accuracy rate of 82%. Conversely, in the second approach, when combined with DNN-based IDS, the accuracy increased to 88%.

In [6] the authors introduce a novel deep learning-based intrusion detection system tailored for IoT networks and devices. The [7] authors proposed a framework for Structural Health Monitoring (SHM) utilizing IoT technologies with intelligent and consistent observation. Additionally, they outlined a strategy for data routing in collaboration with big data analytics. This framework demonstrated improved scalability and low latency performance. In the realm of smart medical systems, the Internet of Medical Things (IoMT), a subset of IoT, serves various functions including patient remote monitoring, medication management, and treatment assessment. However, ensuring the confidentiality of patient records within medical systems is critical within an IoMT environment. Thus, safeguarding confidential health information (PHI) stands as a vital aspect of healthcare data protection. The authors of [8] presented LPME, a lightweight privacy-preserving medical diagnosis mechanism designed for edge computing. LPME remains the Extreme Gradient Boosting (XGBoost) model within an edge-cloud framework and leverages encrypted model parameters rather than local data to reduce computational burdens. Additionally, this approach ensures secure diagnosis at the edge, safeguarding sensitive data while enabling prompt detection.

The authors of [9] proposed an algorithm designed to forecast patients' present health condition while maintaining continuous professional monitoring. They explored additional parameters and methodologies, such as K-nearest neighbor, logistic regression, support vector machines, random forests, and Adaboost classifiers, employing the UCI heart disease dataset. Their endeavor led to the creation of a tool beneficial for patients, healthcare professionals, and the healthcare system overall. This decision-making method achieved an impressive accuracy level of 93%.

In [10], IoMT systems are characterized by their management of substantial data volumes crucial for illness diagnosis, prediction, and monitoring. Due to limited storage and processing capabilities of certain IoMT devices, medical data regarding patients often requires transmission to cloud storage and external computer devices. However, such an

approach introduces security and privacy risks. To mitigate these challenges, a solution is proposed employing a swarm neural network-based approach for intruder identification within IoMT. This model aims to effectively evaluate healthcare data and detect intruders during data transmission. Performance evaluation conducted using the NF-ToN-IoT dataset revealed an accuracy rate of 89.0% for the proposed model.

III. MODEL ARCHITECTURE

In this phase, the models were trained using chosen features that depicted attack behaviors. Afterwards, these models were tested by comparing test data with the training data to gauge their accuracy. A model was considered ready only if the accuracy test produced satisfactory results; otherwise, it underwent retraining until an acceptable level of accuracy was attained. To compare performance, algorithms such as Artificial Neural Network (ANN), Random Forest (RF), Support Vector Machine (SVM), and Naive Bayes (NB) classification techniques were employed.

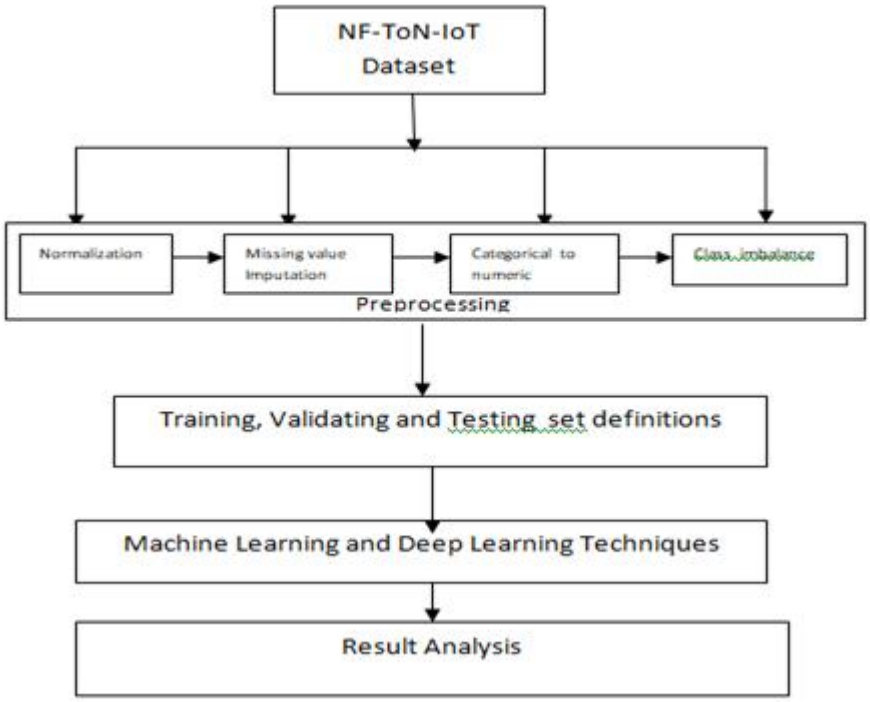

Fig 1. Architecture of the proposed framework

As illustrated in Fig.1, our framework initiates with the acquisition of the NF-ToN-IoT dataset. Subsequently, a crucial step involves data pre-processing, which encompasses normalization, imputation of missing values, conversion of categorical attributes to numerical ones, and addressing class imbalance. Following pre-processing, the dataset is divided into three distinct sets: a training set (70%), a validation set (15%), and a testing set (15%). The training set is utilized for training the ML models (namely, NB, RF, ANN, and SVM). Post-training, the models carry out predictions on the testing set to ascertain their accuracy in generating results. Various evaluation parameters such as accuracy, precision, recall, and F1 score are employed for assessment. Data Pre-processing is a process of converting unprocessed data into a format that can be read, accessed and analyzed. Prior to employing ML and DL algorithms, pre-processing plays a crucial role in ensuring or enhancing the overall performance or accuracy of any system. During the pre-processing phase, data is cleaned to serve a variety of purposes. Some ML algorithms require data to be in a specific format to ensure compatibility with various algorithms. To tackle these issues, the following pre-processing steps were undertaken:

i. The NF-ToN-IoT dataset utilized in this research posed several challenges, including missing values, categorical attributes, and class imbalance.
ii. Upon examination, it was determined that the NF-ToN-IoT dataset did not contain any missing values. Consequently, the dataset was affirmed to be of high quality and value.
iii. The NF-ToN-IoT dataset includes various categorical features, necessitating the conversion of these categorical attributes into numerical values. This conversion process was executed using Label Encoder
iv. Class imbalance occurs when certain classes are significantly more prevalent than others. In such cases, standard classifiers often neglect the minority classes as they are overshadowed by the abundance of the majority classes.
To tackle this issue, the SMOTENN (Synthetic Minority Over-sampling Technique and Edited Nearest Neighbors) method was implemented to balance the classes within the dataset. The selected features play a crucial role in enhancing the performance of the intrusion detection task. Therefore, the seven features outlined in Table IV were carefully chosen as they were deemed most influential in facilitating the classification process. Increasing the number of features beyond these would not significantly impact the classification process. By removing non-essential features, the risk of over fitting is reduced, computational speed is improved, and model accuracy is enhanced.

IV. IMPLEMENTATION AND RESULT

In this paper, the models utilized were trained on an NF-TON-IoT dataset. Subsequently, a set of data extracted from these models was employed to evaluate the accuracy of the trained models by accurately categorizing all data into its

respective labels. Accuracy, Precision, F1-Score, and Recall are the metrics employed to assess the effectiveness of the algorithms. Each metric can be defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1 Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

Table 1: Comparison of the performance of the 5 different classifier on the NF-ToN Dataset

Classifier	Accuracy	Precision	Recall	F1 Score
SVM Classifier	0.8234	0.8943	0.8234	0.8027
NB Classfier	0.7825	0.8027	0.7925	0.7751
RF classifier	0.9839	0.9838	0.9839	0.9838
ANN classifier	0.9299	0.9302	0.9299	0.9298

Table 1 illustrates that the Random Forest (RF) classifier achieved the highest accuracy (0.98), followed by the Artificial Neural Network (ANN) model (0.93). Conversely, the Support Vector Machine (SVM) model exhibited a lower accuracy of 0.82, while the Naïve Bayes (NB) classifier achieved the lowest accuracy at 0.78. In terms of other metrics such as precision and recall, they are summarized by the F1 score. The Random Forest (RF) Model obtained the highest F1 score (0.98), followed by the Artificial Neural Network (ANN) 0.93, respectively. The Support Vector Machine (SVM) model yielded a lower F1 score (0.80), whereas the Naïve Bayes (NB) classifier achieved the lowest F1 score at 0.77.Fig. 2 depicts a visual representation of the performance of the classifiers in terms of accuracy, precision, recall, and F1 score.

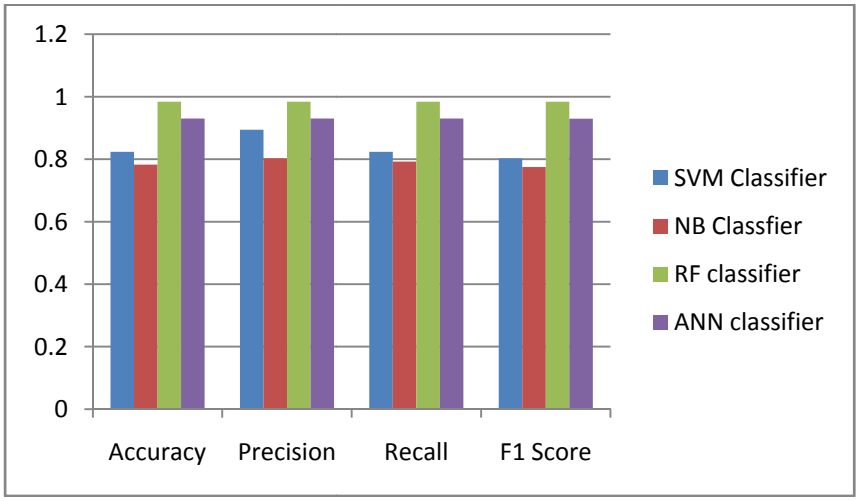

Fig 2: Comparison of performance of the different classifiers on the NFToN-IoT dataset.

V. CONCLUSION

The Internet of Things (IoT) environment encompasses a vast amount of private and sensitive health data that must be safeguarded against tampering or theft. Without adequate safety measures, these threats and attacks against IoT devices within the healthcare sector could potentially devastate the industry. Such attacks are often motivated by financial gain, whether through the sale of stolen data or by holding victim data hostage for ransom. Embracing technology, particularly IoT, in the healthcare sector represents a significant step forward in providing enhanced services for patients and facilitating communication and the sharing of crucial files or tasks. Hence, the implementation of IoT in healthcare is imperative. Indeed, the utilization of IoT presents various privacy and security concerns. Without proper safety measures, threats and attacks against IoT devices within the healthcare sector pose a significant risk to the industry's integrity. These attacks often aim for financial gain, either through the sale of stolen data or by extortion. Therefore, implementing a robust security system alongside IoT deployment becomes paramount. In this study, five different classifiers were employed to forecast potential attacks within IoT services. The NF-ToN-IoT dataset was chosen for this purpose, and underwent pre-processing before being partitioned into training and testing data sets. Upon evaluation, it was observed that the Random Forest (RF) model outperformed others, achieving the highest accuracy (0.98) and F1 score (0.98). Following with the Artificial Neural Network (ANN) model ranking second in performance. Future endeavors will focus on experimenting with models employing diverse algorithms and datasets. Additionally, there are plans to explore combinations of multiple deep and machine learning algorithms to develop models yielding the highest accuracy rates and minimizing loss rates. This pursuit aims to achieve optimal results in classifying attacks on IoT devices within the electronic health sector. Furthermore, there will be a comparison between different datasets and algorithms to assess their effectiveness.

REFERENCES

[1]. Amjad Rehman 1 , Khalid Haseeb 2 , Tanzila Saba 1 , Jaime Lloret 3,4,* and Usman Tariq ,"Secured Big Data Analytics for Decision-Oriented Medical System Using Internet of Things", Electronics 2021, 10, 1273
[2]. Nagasubramanian, G.; Sakthivel, R.K.; Patan, R.; Gandomi, A.H.; Sankayya, M.; Balusamy, B. "Securing e-health records using keyless signature infrastructure blockchain technology in the cloud. Neural Computing. Appl." 2020, 32, 639–647
[3]. Haifa Khaled Alanazi1 , A. A. Abd El-Aziz2 , Hedi Hamdi "Securing IoT Devices in e-Health using Machine Learning Techniques", International Journal of Advanced Computer Science and Applications,September 2023.
[4]. Alenoghena, Caroline Omoanatse, et al. "eHealth: A survey of architectures, developments in mHealth, security concerns and solutions." International Journal of Environmental Research and PublicHealth 19.20 (2022): 13071.
[5]. Dey, Samrat Kumar, and Md Mahbubur Rahman. "Effects of machine learning approach in flow-based anomaly detection on software-defined networking." Symmetry 12.1 (2019):.
[6]. Awajan, Albara. "A novel deep learning-based intrusion detection system for IOT networks." Computers 12.2 (2023): 34.
[7]. Tokognon, C.A.; Gao, B.; Tian, G.Y.; Yan, Y. Structural health monitoring framework based on Internet of Things: A survey. IEEE Internet Things J. 2017, 4, 619–635
[8]. Ma,Z.;Ma,J.; Miao, Y.; Liu, X.; Choo, K.-K.R.; Yang, R.; Wang, X. Lightweight privacy-preserving medical diagnosis in edge computing. IEEE Trans. Serv. Comput. 2020.
[9]. Chola, Channabasava, et al. "IoT based intelligent computer-aided diagnosis and decision making system for health care." 2021 International Conference on Information Technology (ICIT). IEEE, 2021.
[10]. Awotunde, Joseph Bamidele, et al. "A deep learning-based intrusion detection technique for a secured IoMT system." International Conference on Informatics and Intelligent Applications. Cham: Springer International Publishing, 2021.

