International Journal of Advances in
Soft Computing and Intelligent Systems (IJASCIS)
2024, Vol 03, Issue 02, 387-402
ISSN: 3048-4987

Original Paper

SECURE INTRUSION DETECTION SYSTEM FOR INTERNET OF MEDICAL THINGS USING RANDOM FOREST CLASSIFIER AND ELLIPTIC CURVE CRYPTOGRAPHY

Arnav Deshmukh[a,*], Satyam Kendre[b]
Dhanashri Wategaonkar[c] , Shlok Sarda[d]
Zaki Shaikh[e]

[a,b,d,e] Student at MIT WPU University, Pune India,

arnav.d.148@gmail.com, satyamkendre1@gmail.com,
sardashlok2004@gmail.com , zakix1502@gmail.com

[c]Assistant Professor at MIT WPU University, Pune, India,
dhanashri.wategaonkar@mitwpu.edu.in

ABSTRACT

Our research aims to enhance IoMT (Internet of Medical Things) security by integrating a Random Forest Classifier-based IDS (Intrusion Detection System) and ECC (Elliptic Curve Cryptography). IDS detects and classifies intrusions in IoMT networks, and with Random Forest Classifier, it efficiently analyzes network traffic patterns with high accuracy. ECC ensures confidentiality and integrity of data transmitted over the network. Experiments in a simulated IoMT environment showed that the Random Forest Classifier achieves a high detection rate for attacks like DoS (Denial of Service) and MITM (Man in the Middle) while maintaining a low false positive rate. ECC provides strong security measures, protecting critical data from unauthorized access. Our research contributes to developing secure IoMT systems, ensuring the integrity of IoMT device communications against evolving cyber threats.

Keywords

Machine Learning, Deep Learning, Encryption, Internet of Things & Healthcare.

1. INTRODUCTION

As the digital transformation of the healthcare industry is growing rapidly, the integration of Machine Learning (ML), Internet of Things (IoT), and cybersecurity in healthcare has become a necessity. Integrating all of them together offers various benefits to the healthcare organization, such as ensuring security and privacy, improving patient care, etc.

For analysis of complex medical data which include electronic health records (EHRs), medical images and genomic data, machine learning has proved to be a powerful tool in healthcare. ML models can help in diagnosing diseases, predicting patient outcomes and giving them personalized treatment plans. Moreover, ML can also be applied in automating administrative tasks, improve resource allocation and optimize hospital operations.

Nowadays, IoT is used massively in healthcare as it helps in connecting medical devices, wearables like smart watches, and have sensors to collect real-time data from patients. The devices which have IoT enabled in them have huge benefits like timely detection of different morbidities, remote patient monitoring, and personalized health related equipment delivery. IoT devices can also improve the overall patient experience, reduce healthcare costs and enhance the efficiency of medical equipment.

Implementing Elliptic Curve Cryptography (ECC) to encrypt sensitive medical data stored in a hospital database involves several steps to ensure robust security and confidentiality. Initially, the hospital would generate ECC key pairs for each authorized user, including healthcare providers and administrators. The public keys would be securely distributed, while the corresponding private keys remain tightly controlled. When storing medical data in the database, ECC encryption would be used to encrypt the critical data with the help of the receiver's public key which makes sure that only the authorized users who have their own private key can decrypt and access the critical information accordingly. Access controls and authentication mechanisms would also be implemented to regulate user permissions and ensure that only authorized personnel can retrieve and decrypt the encrypted data.

Despite all the benefits of ML and IoT in healthcare, their adoption raises concerns regarding two critical aspects i.e. data security and privacy. As IoT is interconnected with healthcare devices, it may increase the risk of cyber-attacks, malware, and unauthorized access to sensitive information. Therefore, due to all these concerns, cybersecurity is an important aspect in healthcare as it is essential to protect the patient data, maintain trust, and comply with all the regulatory requirements.

Considering all these aspects, our research focuses on highlighting the importance of collaborating Machine Learning, IoT, and cybersecurity in healthcare by implementing an intrusion detection system using Random Forest Classifier and integrate it into IoMT devices for malware detection and to protect the data our work aims to integrate ECC (Elliptic Curve Cryptography) into the IoMT devices, which will enhance their operational efficiency and protect the IoMT devices from cyber-attacks.

2. LITERATURE SURVEY

This section briefly reviews the existing Artificial Intelligence and Deep learning methods within the healthcare domain and how the feature selection techniques have been employed.

2.1. Machine Learning and Deep Learning

Strong cybersecurity measures are vitally needed in the healthcare industry, according to Ahmad A. Alzahrani [1], to protect the critical patient data, assuring the integrity of medical systems. The potential of cyber hijinks is real as more medical devices get connected and everything becomes digital. Thus, to give healthcare cybersecurity a major boost, these brainiacs recommend including some Artificial Intelligence (AI), more precisely these hip-sounding Convolutional Neural Networks (CNN). They're all about AI being the cybersecurity professionals' superhero in the face of ever changing cyberthreats. Healthcare organizations may significantly strengthen their defenses and avoid dangers to patient safety and data privacy by hopping on the AI bandwagon. This is the ACO-CNN model, a convolutional neural network based on Ant Colony Optimization. It is expected to be the next big thing, challenging the vulnerabilities in the cybersecurity of healthcare today. It makes the data look good for analysis by cherry-picking the important parts of the dataset using the ACO approach. Min-max normalization is used to the UNBS-NB15 dataset, which simplifies cyber threat detection by scaling the data to a manageable range. By separating regular from shady patterns, standardizing the data makes it easier for the model to play detective and increases its accuracy in identifying attacks. These carefully chosen

characteristics are then combined with the Convolutional Neural Network (CNN), which is an expert at identifying cyberthreats due to its prowess with structured data. Now, the false alarm rate, accuracy and precision of this study helps to size up the ACO-CNN model's performance. Surprise, surprise – the results show it's the MVP, scoring a 99.15% detection accuracy for the UNBS-NB15 dataset, dusting off Integrated-rule based intrusion detection systems (IRIDS) and Artificial Neural Networks (ANN). Plus, it flexes a low 0.55% False Alarm Rate, showing off against the 2.01% and 2.56% from IRIDS and ANN. And it doesn't stop there. For the KDD99 dataset, the ACO-CNN model nails a 99.25% detection accuracy with a cool 1.3% False Alarm Rate, proving it's the real deal at spotting cyber threats. Overall ACO-CNN model presented in the paper suggests a robust solution for enhancing cybersecurity in medical and healthcare, providing an improved detection model and protecting against cyber threats. The paper by Vivek Kumar et.al [2], emphasizes the difficulty and complexity of analyzing healthcare data, especially given its heterogeneous and unstructured character. The vast amount of healthcare data and the necessity of cutting-edge technologies like artificial intelligence to analyze it efficiently are both emphasized in the report. They draw attention to the fact that artificial intelligence technologies are being used more and more in the healthcare industry for tasks including textual clinical report classification, clustering, and recommendation. The study emphasizes how important it is to detect multimorbidity at an early age, as it affects 25% of the global population, to avert serious health problems down the road. The significance of automating the extraction of morbidity indicators from patient clinical records is emphasized in the research, as it will help healthcare workers manage massive amounts of electronic health records. For patients with multimorbidity's who are deemed more vulnerable, they aim to automatically identify these factors indicated in patient clinical records to support healthcare professionals in tasks like personalized monitoring, dynamic forecasting, and individualized treatment recommendations. The research on using deep learning models and sophisticated word embeddings representations for natural language processing (NLP) tasks—which are frequently utilized in solving healthcare issues—is highlighted in this paper. They concentrate on employing bag-of-words and word embeddings in conjunction with feature selection methods utilizing both deep learning (DL) and conventional machine learning (CML) methodologies to represent clinical records. The study aims to determine whether patients are suffering from single or multiple morbidity conditions by analyzing their previous clinical record.

In this methodology, the author details the dataset used in the study, the n2c2 dataset containing clinical record. They employ classical machine learning and deep learning approaches for multi-classification of clinical record using bag of words TF-IDF and word embedding feature representation methods. The author compares the performance of CML classifiers and DL models. They have created a model by combining DL model and CML classifiers using a voting strategy to improve prediction and mitigate biased behavior. The results highlighted from the study are that the dataset size plays a main role in the performance of the Dl model, especially when the training dataset is unbalanced. Classical machine learning performs better than Dl model when used with word embedding representations. Word embeddings significantly outperform TF-IDF representations and feature selection algorithms in DL approaches. Model created by combining Dl model and CML classifiers, it proves beneficial for small datasets. The author highlights the importance of dataset size in improving DL model performance, future work is suggested to explore techniques like data augmentation and advanced word embeddings representations. The research paper [3] by Adil Khan and Ishu Sharma focuses on the need for effective security steps in healthcare organizations to protect patient data from ransomware attacks. The risk of being exposed to cyber-attacks has also increased tremendously due to the rapid growth of internet usage and smart devices. Ransomware also poses a critical threat to all the applications, and systems. The paper emphasizes the most on deploying machine learning models to detect and prevent ransomware attacks, more on Android devices. They do so by introducing a semantic-based method using Application Programming Interface (APIs) weighted contextual dependency graphs, [3] aim to detect zero-day malwares and categorize Android

malware effectively. The methodology in [3] involves the deployment of different machine learning algorithms, like Logistic Regression, Support Vector Machine (SVM) and Random Forest Classifier (RFC), to identify ransomware attacks in the healthcare domain. For the early detection of malware and ransomware, data training is done to build the Random Forest model. The study emphasizes that by implementing machine learning algorithms for improving network security, we can protect the critical information that is held within the healthcare organization which includes patient records and staff data. Confusion matrices are made for each model for evaluating the classification performances and making decisions regarding feature selection and model optimization, and to ensure the accuracy and effectiveness of the models. It is observed that Random Forest outperforms all the other models with an accuracy of 99.79%, and the Logistic Regression algorithm also shows high accuracy in identifying ransomware attacks. To assess the model's potential to comprehend in between the positive and negative classifications, we use ROC (Receiver Operating Characteristic) and Area Under the Curve (AUC). The overall research concludes that Random Forest Classifier is the best algorithm for identifying ransomware while deploying it in the healthcare sector. The paper by Aryan Tuteja et.al [4] highlights the importance of intrusion detection systems (IDS) in ensuring security and confidentiality of healthcare systems and organizations. The potential outcomes of data breaches in patient data and the increasing risk of data to cyber-attacks is highlighted in [4]. To enhance the effectiveness of IDS in detecting and preventing cyber-attacks, the authors emphasize various machine learning techniques like logistic regression, etc. [4] also aims to spread awareness of implementing IDS in the healthcare domain to protect the sensitive data and mitigate risks. In [4], the authors tend to implement a logistic regression model for IDS in healthcare organizations. It analyzes the CSE-CIC-IDS2018 dataset and the utilization of different machine learning algorithms, including logistic regression, random forest, and multi-layer perceptron have also been described. The authors explain the process of data collection, analysis, and balancing to address data imbalance issues and ensure the reliability of the model. To ensure the accuracy of the dataset processes like data cleaning, data analysis and data balancing are also explained by the authors. They also outline the architecture, which includes model training, feature extraction and evaluation of the model. To assess the performance and accuracy of the model, cross validation techniques like F1-score, recall, and precision are used. The results of the intrusion detection system show that after data pre-processing a mean cross validation accuracy of 98% which includes 68 features have been reported. The logistic regression model demonstrates high precision, recall, and F1-score, indicating its effectiveness in classifying and predicting cyber-attacks.

2.2. IOT

The paper by Abdallah Ghourabi [5] highlights the importance of cybersecurity in the healthcare organizations and systems which arise because of the sensitive nature of the patient data and due to the increasing numbers of the cyberattacks targeted towards the IoT devices and the networks. The author also emphasizes the limited availability of the datasets which specifically are focused on the different attacks on medical devices, and on further research ECU-IoHT and ICE are chosen as the primary datasets for this domain. However, the author still feels that these datasets are also not enough for developing robust IDS systems. Responding to these gaps, the author provides a security model which aims to protect healthcare organizations against cyber-attacks. The proposed work is presented in two keyways: an IDS (Intrusion Detection System) to monitor the IoMT systems and a malware detection system which looks after the computers used by medical staff. By combining these two ways, the goal is clear, which is to provide security irrespective of the type of device which can be targeted by hackers. The proposed work focuses mainly on machine learning algorithms, like LightGBM and BERT-based transformer models which enables detection of both attacks which are known and unknown. To enhance the detection of different types of attacks, the author highlights the advantages of using machine learning algorithms and the effectiveness of their hybrid security system. To achieve a high-level accuracy in identifying and mitigating cyber threats in healthcare systems, the model is trained on various

datasets, which include ECU-IoHT, ToN-IoT, Edge_IIoTset, and EMBER. In [5], the author emphasizes protecting the healthcare organizations from different cyberattacks and to do so, the author describes the detailed approach of his security model based on LightGBM and Transformer. The paper also emphasizes the importance of using diverse datasets to train the model effectively. Four different datasets are chosen for training the models: ECU-IoHT for analysis of different attacks in medical devices, ToN-IoT and Edge_IIoTset for IOT attacks and EMBER for tracking malware on Windows. The selection is done so, to ensure that the models have different characteristics and patterns of different cyberattacks. T For handling the textual data, a transformer model is trained and analyzed using the ECU-IoHT dataset while LightGBM, is utilized for categorical and numerical data which comes from the ToN-IoT, Edge_IIoTset, and EMBER datasets. To analyze the performance of the trained models, they are evaluated using various metrics such as ROC, AUC, accuracy, and classification results on different datasets. The author aims to achieve high accuracy rates (close to 100%) in detecting and categorizing different cyber threats in healthcare systems. The ECU-IoHT dataset achieves close to 100% accuracy, while the EMBER dataset which uses the LightGBM model yields an accuracy of 97.96% and a ROC AUC of 99.68%. Therefore, [5] confirms that for analyzing textual data, the Transformer model is useful and for categorical and numerical data, the LightGBM model can be used.

This paper by Panagiotis Radoglou-Grammatikis et.al [6] digs deep into the evolving landscape with the arrival of the Internet of Medical Things (IoMT). It emphasizes the numerous advantages that are brought by IoMT devices, such as real time monitoring, remote medical assistance and prevalent control. However, the increasing digitization raises cybersecurity and privacy concerns within the healthcare domain. Mainly, the focus is driven towards the IEC 60 870-5-104 protocol which is a widely adopted standard in the healthcare industry. The potential threats and vulnerabilities associated with IEC 60 870-5-104 are emphasized, pointing out the risks posed by cyberattacks like unauthorized access and DoS attacks. Furthermore, it is identified that cyberattacks targeting the IEC 60 870-5-104 protocol in healthcare systems can have a steep effect on the other critical infrastructures, increasing the impact of such security breaches. The methodology proposed in [6] involves the development of a quantitative threat model for evaluating the severe impact of cyberattacks on the IEC 60 870-5-104 protocol. The proposed model combines Common Vulnerability Scoring System v3.1 and Attack Defence Trees. Using machine learning and software defined networking technologies, an IDPS (Intrusion Detection and Prevention System) is introduced. With the help of Thompson Sampling (TS) method, the multi-armed bandit (MAB) is solved where SDN-based mitigation is framed while the IDPS utilizes a CART classifier for ID based on network flow statistics and payload flow statistics. The model shows promising results in terms of detecting intrusion and automated mitigation capabilities. For enhancing cybersecurity in industrial healthcare systems, the results show the potential of combining machine learning and SDN technologies. As the introduction of the research study by Andreou Andreas [7], enhanced accessibility and security are critical factors for IoT networks in the healthcare field. The author highlights the challenges of network privacy and accessibility specifically in the context of covid-19 period. The pandemic has brought attention to preparation gaps in medical and healthcare fields. The research proposes a robust encryption strategy as a solution to meet the increasing demand for data security in the healthcare field. By integrating cryptography in healthcare computer systems and advancing IoT frameworks, the research aims to provide a secure mechanism for transferring text information securely. The use of circulant matrices and the integration of the Gematria alphanumeric method are introduced as innovative techniques to facilitate confidential transmission of text messages. The suggested encryption architecture is described in full in this [7] approach. Real-time data is collected by IoT medical devices and sensors as part of the framework's architecture, then it is encrypted and stored securely. The cloud server gets the encrypted data. This data can be accessed by medical professionals and can be used for prescription writing and for tracking purposes. This

methodology also tells how to improve the secrecy for data flow between IOT devices by using circulant matrices and the Gematria alphanumeric method. The research section is focused on the performance evaluation and analysis of proposed algorithms for encryption and decryption. The research aims to minimize or optimize the time interval required to encrypt or decrypt data to get the best performance. The study shows how efficient it is to construct secure software for encryption. Future research directions include further enhancing the encryption process using fundamental mathematics, such as the generalized discrete Fourier transform, to implement symmetrical convolution for improved data security in healthcare ecosystems. The results highlight the potential of the proposed cryptographic method to enhance confidentiality and security in healthcare data sharing environments. Amir Djenna et al.'s work [8] provides a thorough examination of the growing threat that cyberattacks—especially Distributed Denial of Service (DDoS) attacks—pose to vital infrastructures, including medical, aviation, energy, and defense systems. The University Constantine2 in Algeria's Amir Djenna, Djamel Eddine Saidouni, and Wafia Abada wrote this paper, which emphasizes how urgent it is to implement cutting-edge cybersecurity measures in order to thwart cyber enemies' ever-evolving strategies. IoT (Internet of Things) has advanced to the forefront of technical breakthroughs in recent years due to society's growing reliance on technology. IoT offers a new era of networked devices and systems across numerous industries, including home automation, healthcare, transportation, energy, and defense. It does this by seamlessly integrating the physical and digital realms. But the quick spread of IoT devices has also exposed vulnerabilities, making them prime targets for cyber threats. The introduction of the paper highlights the need for cybersecurity strategies to overcome IoT-related cybersecurity. The author highlights the significance of early detection of threats and early mitigation of risk within the IoT network. By using artificial intelligence models and advanced detection mechanisms, the author proposed a methodological framework for identifying and stopping DDOS attacks targeting an IoT system. The integration of AI-driven solutions aims to enhance the flexibility and security posture of IoT infrastructures, enabling proactive threat mitigation and rapid response to cyberattacks. The result shows the efficiency of the proposed model and its accuracy in identifying the threat and its response to the attack, particularly in DDOS attack within the IoT network the model has high accuracy in detecting various type of attack such as including DDoS, TCP-SYN Flood, UDP Flood, and ICMP Flood. The methodology's ability to distinguish between normal network traffic and malicious activities is highlighted through metrics such as precision, accuracy, recall, F1-score and false alarm rate. Due to this robust model the detection and mitigating the different types of cybers attack has helped to enhance the security of IoT framework. The paper by Marshal R et.al [9] dives deep into the topic of cybersecurity in the world of IoT-based smart healthcare networks. It highlights the increasing importance of Internet of Things (IoT) technology across various domains and the different vulnerabilities it introduces. In [9], emphasis is laid upon the widespread adoption of IoT devices and their integration into smart healthcare applications, defining the different roles these technologies can play in modern healthcare organizations. The COVID-19 acts as a stimulus that highlights the importance of smart healthcare solutions, mainly in terms of remote monitoring and data analysis. Moreover, [9] ponders on the challenges faced by the healthcare sector due to the limited security features that the medical devices have, in turn posing life threatening consequences of cyber-attacks on these devices. The need for robust cybersecurity measures have also been emphasized, looking upon the critical data that the healthcare organizations hold. The methodology in [9] focuses on enhancing security of smart healthcare networks through different key strategies for mitigating security challenges in IoT-based smart healthcare environments include Deploying cyber security experts: To define the gap between IT (Information Technology) and OT (Operational Technology) in smart healthcare environments, an adequate number of cyber security experts can be deployed in hospitals who can monitor, update and protect the connected devices regularly. Inventory Maintenance: To continuously monitor the device behavior, traffic and connections, the development of a vulnerability database which is updated with the latest vulnerability reports is essential for identifying and addressing security risks and so that helps in maintaining an inventory of all the devices connected to the IoT network

which is important. Network Segmentation: To isolate critical devices from unauthorized access, micro segmentation can be implemented. This measure aims to ensure that devices are loosely coupled which minimizes the impact of a single device failure on overall network performance. Data Integrity: Highlights the importance of ensuring that the data is stored in storage mediums which are accessible only to authorized users. Also, emphasis is laid upon collecting only eh information, which is necessary, and that the data should be periodically backed up to mitigate the potential attacks.

Security Audit: - The idea of conducting third party audits periodically is proposed to assess the vulnerabilities in smart healthcare networks. The devices which cannot be updated with the latest patches should be replaced with new ones to enhance network security. By implementing these proactive measures, organizations can strengthen the security posture of IoT-based smart healthcare networks, which helps safeguard the critical patient data and ensure the reliability of critical healthcare services. Along with an integrated keyword search, this paper by Shanthi M.B. et.al [10] introduces a distinct method for public key cryptographic function. The research aims to generate cryptographic keys shared by both sender and receiver in healthcare IoT environments, inspired by the Asymmetric cryptographic key exchange protocol. The proposed method mainly focuses on securing private information in IoT healthcare to preserve the patient's private data, laying emphasis on the method based on searchable encryption and attribute-based encryption. Compared to existing methods, the proposed method aims to provide enhanced security, improved search efficiency, and increased privacy than the previous ones. To test the proposed solution, real time data from different healthcare IoT (Internet of Things) devices was used and it demonstrated superior performance in terms of privacy, security and search efficiency. [10] uses efficient public key authenticated searchable encryption involving many steps, which is used to protect the critical data in healthcare IoT. The paper by Sabrina Ahmed et.al [11] gives an outline of all the challenges and solutions related to protecting data in the context of emerging technologies like 5G and IoT. In [11], the importance of protecting the sensitive healthcare data is highlighted, as the information travels from IoT sensor devices to healthcare providers. To safeguard data at the sensor level and throughout the entire process, there is a need for cryptographic security mechanisms. The proposed solution in [10], aims to confirm the safe transmission of healthcare data in 5G Edge Computing Networks. Initiating Cryptographic Security at IoT Sensor Device Level: To ensure that the data is secured right from the start and that the overall security transmission process is enhanced, the sensitive healthcare data is directly encrypted at IoT sensor device before transmission. Encryption: Now, the sensitive data is encrypted using a cryptographic key-code matrix which forms a cryptographic data packet. This process is repeated to add an additional layer of security, making it more resistant to interception or unauthorized access. Transmission via Software Defined Networking (SDN) Routers in 5G Edge Computing: Now, the encrypted packets are securely routed through Software-Defined Networking (SDN) routers within 5G Edge cloud. This process ensures that the data is protected as it travels to its destination. E.g. Doctors' office. Decryption: As soon as the packet reaches the destination, using the inverse of the original key-code matrix the encrypted message is decrypted. This process allows the user to retrieve his original data back, ensuring integrity and confidentiality. Testing and Analysis: A prototype with two-layer encrypted communication is lab-tested to validate the effectiveness of cryptographic security solutions. The results provide understanding of the performance and security of the proposed work. Optimization and Scalability: Based on how critical the patient data is, the proposed work also considers optimization of encryption key levels. By adjusting the number of encryption key levels, healthcare industries can also optimize costs while maintaining data security. The test results of the two-layer encrypted communication setup were promising for achieving end-to-end security in healthcare data transmission within 5G Edge Computing networks and beyond. The analysis of the results indicated the feasibility and effectiveness of the proposed cryptographic security solution in safeguarding patient information during transmission. Overall, the results of the study validate the efficacy of the cryptographic security solution initiated at the IoT sensor device and

routed through SDN routers in ensuring secure and confidential healthcare data transmission in modern network environments.

2.3. Cybersecurity and Healthcare

The paper by Khadija Abu Ali et.al [12] points out how the transformation from traditional information to digital systems, like electronic health records (EHR), has impacted the security environment of the healthcare industry. This change has affected the healthcare organizations as now they are being exposed to a range of cyber threats and attacks as there are not many strong solutions formed in defending all these cyber-attacks. In [12], it is highlighted that these healthcare systems mainly deal with highly sensitive data which includes patient information, payment information and the private information which makes them a major target for the hackers. In [12], it is informed that cyberattacks performed on the healthcare systems result in major damage to both the organizations and the patients, highlighting the need to enhance security protocols in the healthcare domain. Moreover, it is mentioned in [12], how cybersecurity has become an integral part of various industries, including education, healthcare and government industries. It is noted that data breaches, security risks and cybersecurity awareness are the key areas of concerns in the healthcare industry.

3. ARCHITECTURE DIAGRAM

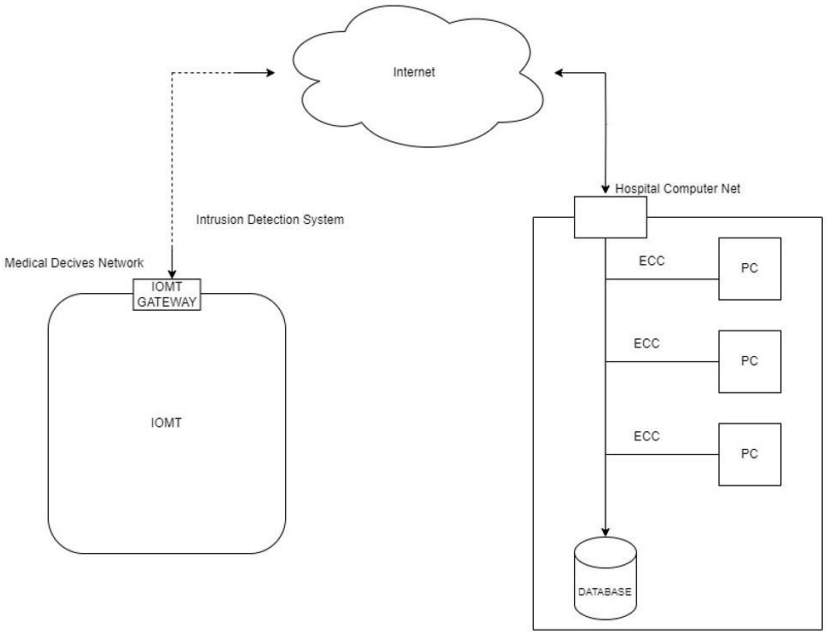

Figure 1. Architecture Diagram of the model

In the context where ECC stands for Elliptic Curve Cryptography, the connections in the diagram can be explained as follows Internet to PC: This connection represents the communication link between the external Internet network and a personal computer (PC) within the network

environment. Elliptic Curve Cryptography (ECC) may be utilized to secure data transmission between the Internet and the PC through encryption and decryption processes.

PC to DATABASE: This connection signifies the data exchange between a personal computer (PC) and a centralized database using Elliptic Curve Cryptography (ECC) method for secure data storage. ECC can be employed to protect sensitive information stored in the database from unauthorized access. DATABASE to ECC: This connection indicates the interaction between the centralized database and the Elliptic Curve Cryptography (ECC) system for securing data integrity and confidentiality within the database. ECC algorithms may be applied to encrypt and decrypt data stored in the database. ECC to Intrusion Detection System: This connection shows the integration between the ECC system and an IDS for enhancing network security. ECC can be used to encrypt data monitored by the Intrusion Detection System to prevent unauthorized access or tampering. Intrusion Detection System to ECC: This connection represents the bidirectional communication between the Intrusion Detection System and the Elliptic Curve Cryptography (ECC) system. ECC algorithms may assist in securing data integrity and confidentiality within the network monitored by the Intrusion Detection System. ECC to IOMT: This connection signifies the link between the Elliptic Curve Cryptography (ECC) system and the IOMT devices for securing data communication. ECC can be implemented to encrypt data exchanged between the ECC system and the interconnected medical devices within the IOT ecosystem. IOMT to GATEWAY: This connection illustrates the secure data transfer between the Internet of Medical Things (IOMT) devices and a gateway device using Elliptic Curve Cryptography (ECC) for encryption. ECC algorithms can ensure data confidentiality and integrity during communication between the IOMT devices and the gateway. GATEWAY to Medical Devices Network: This connection represents the secure communication between the gateway device and the network managing medical devices within a healthcare facility using Elliptic Curve Cryptography (ECC) for data protection. ECC can be employed to encrypt data transmitted between the gateway and the medical devices network. Medical Devices Network to Hospital Computer Network: This connection signifies the secure data exchange between the network housing medical devices and the hospital computer network infrastructure using Elliptic Curve Cryptography (ECC) for encryption. ECC algorithms can safeguard data transmission between medical devices and hospital systems within the network. By incorporating Elliptic Curve Cryptography into the network connections, data confidentiality, integrity, and security can be enhanced throughout the healthcare network environment.

4. PROPOSED METHODOLOGY

For our proposed work, we have integrated machine learning and cryptography to enhance security in the IoMT devices. We have divided our system into 2 parts: The first part (4.1) will explain about training the machine learning model i.e. Random Forest Classifier into the IoMT devices for malware detection. The second part (4.2) will explain about ECC (Elliptic Curve Cryptography) to protect the patient's sensitive data in hospital databases.

4.1. Machine Learning Model Development:

For our proposed work, we used the Random Forest Classifier model to train our CICIDS2017 dataset.

Random Forest Classifier is one the most popular machine learning algorithms which can be used for classification and regression tasks. We use the ensemble method to train the model i.e. it builds multiple decision trees during the training of the model

and outputs the prediction of the individual trees.

- Random Sampling:
    - Randomly select n samples from the training set with replacement to form a subset S= {(x1 ,y1),(x2 ,y2),...,(xn ,yn)}, where xi is the feature vector and yi is the corresponding label.
- Feature Randomness:
    - Let m be the number of features. At each node of the decision tree, a subset of k features is selected randomly, where k<<m.
    - Typically, k= √m for classification and k=m/3 for regression.
- Decision Tree Learning:
    - Grow a decision tree from the subset S using a standard algorithm like CART (Classification and Regression Trees).
- Voting (Classification) / Averaging (Regression):
    - For classification, each and every tree in the forest "votes" for a labeled class. The class with the most votes is the final prediction.

Random Forest is a powerful model for our intrusion detection system as it can handle high-dimensional data, its ability to deal with imbalanced datasets, and provide good generalization performance. It helps in doing so due to its: -

- Feature Importance: Random Forests rank features based on their importance in classification which helps in detecting intrusions.
- High Accuracy: While minimizing FP (false positives) and FN (false negatives), maintaining high accuracy in classifying instances is also necessary which is done through Random Forest Classification.
- Handling Imbalanced Data: Random Forest can easily handle imbalanced datasets well and intrusion detection datasets often have imbalanced class distributions.
- Robustness to Overfitting: Random Forest, with its reduced susceptibility to overfitting in contrast to standard decision trees, emerges as a potent tool for intrusion detection, as it necessitates robust generalization to unfamiliar data.
- Parallel Training: As Random Forest can be trained in parallel, they are efficient for larger datasets, which is very common in intrusion detection datasets.
- Anomaly Detection: For anomaly detection, Random Forests can be used by treating normal instances as one as anomalous instance as another class which

helps in detecting novel attacks which are not seen during training.

Feature Selection and Engineering: -

- Extracting relevant features:
    - Capturing network traffic patterns: - The features such as flow duration, number of packets per flow, packet size and protocol type can be extracted to capture characteristics of network traffic which in turn help in detecting various types of attacks.
    - Monitoring device behavior: - Monitoring the features can provide information into attacks that are potentially possible. These features include frequency of accessing certain files, amount of data transferred, etc.
    - Communication Protocols: - The features describing the usage of communication protocols like TCP and UDP, like the amount of data transmitted and the number of connections can be informative in determining attacks.

Applying different techniques for feature selection:

- For correlation analysis: - The calculation of correlation between each pair of features and identifying the highly correlated features that may be redundant is necessary as it helps in reducing overfitting and improving model interpretability. In our model, we have applied a correlation heatmap to identify the same.
- For Dimensionality reduction: - In our model, we have used techniques like PCA to reduce the dimensionality of the feature space while preserving the variance as much as it is possible which can help in visualizing high dimensional data and identify clusters and patterns.
- Engineering new features:
    - Time-based features: Creating features that capture the time between successive events (e.g., time between login attempts) or the frequency of events within specific time windows can help in detecting patterns of malicious behavior that occur over time.
    - Aggregated features: Aggregating features over different time intervals (e.g., average packet size per minute) or across different dimensions (e.g., total number of connections per source IP) can provide a higher-level view of the data and potentially reveal important patterns.
    - Interaction features: Creating interaction features between pairs or groups of features (e.g., the product of two features) can help capture

nonlinear relationships and interactions between different aspects of the data.

4.2. Elliptic Curve Cryptography (ECC):

Elliptic Curve Cryptography (ECC) is an asymmetric type of cryptographic technique based on the algebraic structure of an elliptical curve over a finite field. Elliptical curve is a matrix defined by an equation $y^2 = x^3 + Ax + B$ where A and B are constants and x and y are variables. Elliptic Curve Cryptography (ECC) is one of the most powerful and widely used methods in public key cryptography. It provides strong security even though it has a small key size as compared to RSA and DSA. Figure.2 shows the graphical representation of ECC.

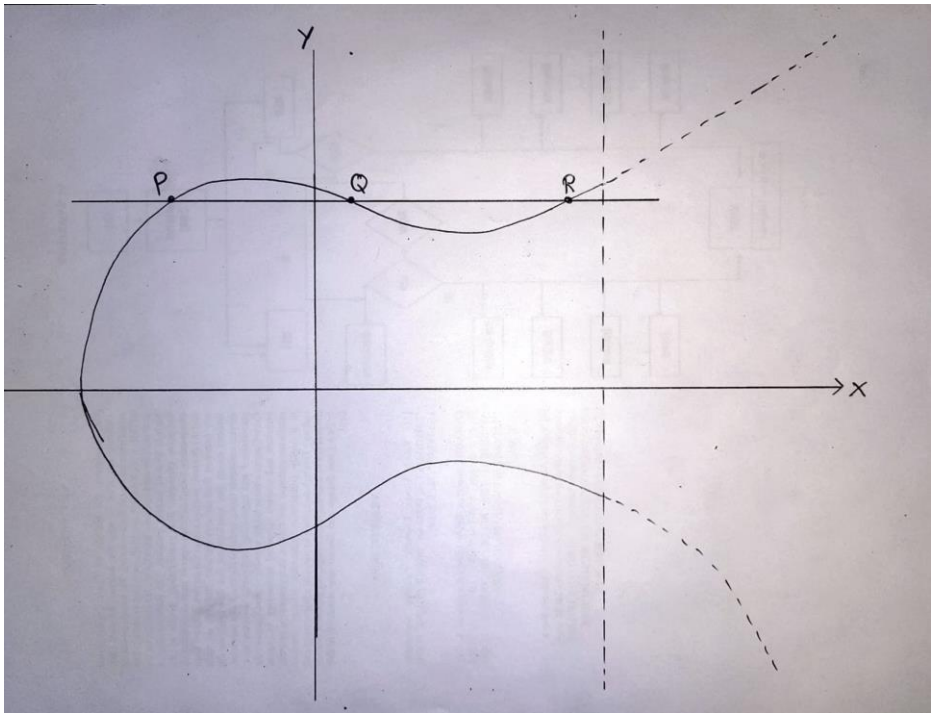

Figure 2. A graphical representation of ECC

As we can see an elliptic curve in Fig.2 it goes to infinity but we are limiting the curve up to value 'n' or line n. We can also see the curve above the x-axis is symmetric to the curve below the X-axis and if we draw a line on the curve, it cuts the curve maximum at 3 points. As we can see in the fig.1 a line cuts a curve at 3 points (P, Q, R). Now we will see the Elliptic Curve Cryptography (ECC) key exchange and key generation. For this we need global public elements i.e. G and Eq (a, b). In Eq (a, b), E is an elliptic curve. a, b and q are parameters and q is either a prime number or an integer in the form of 2^m. G is a point on the elliptic curve. Now the key generation for user 'A' first step is to select a private key (n_a), based on the private key public key will be calculated. Select the private key (n_a) where n_a will be less than n ($n_a < n$). Here is a point where we are limiting the curve. The public key for users will be (P_a). The value of public key for user 'A' will be equal to (P_a)

= (n_a x G). Here G is a point on the elliptic curve and (n_a) is a private of user ‘A’. Similarly, key generation for user ‘B’ first step is to select a private key (nb), based on the private key public key will be calculated. Select the private key (nb) where nb will be less than n (nb < n). Here is a point where we are limiting the curve. The public key for users will be (Pb). The value of public key for user ‘B’ will be equal to (Pb) = (nb x G). Here G is a point on the elliptic curve and (nb) is a private of user ‘B’. After calculating the public and private key of user ‘A’ and user ‘B’, we are going to calculate the secret key shared by user ‘A’ and by using private and public key of user ‘A’ and ‘B’. The secret key (K) by user ‘A’ will be K= na x Pb. where na is the private key of user ‘A’ and Pb is the public key of user ‘B’. similarly, we are going to calculate the secret key shared by user ‘B’ and by using private and public key of user ‘A’ and ‘B’. The secret key (K) by user ‘B’ will be K= nb x Pa. where nb is the private key of user ‘B’ and Pa is the public key of user ‘A’. Let’s see the Encryption in Elliptic Curve Cryptography (ECC), let us consider message M as plain text. The first test is to encode message M into a point on the elliptic curve, let the point be Pm.For encryption, it's necessary to select a positive integer at random let consider it as ‘K’. based on the value of the K we will calculate cipher point. These cipher point will be sent to receiver i.e. user ‘B’. let cipher denote by cm = {KG, Pm +KPb } , here K is a random positive integer, Pm is a plain text point on elliptic curve, G will be a point on the elliptic curve and Pb is a public key of user B. X coordinate is KG and Y coordinate is Pm + KPb . In this way we are calculating cipher points. Now for Decryption in Elliptic Curve Cryptography (ECC), after receiving the cipher point i.e. cm. based on cipher point receiver will calculate plain text point i.e. Pm. First multiply the x coordinate of the cipher point with the receiver secret key (KG x nb). After this, subtract (KG x nb) from the coordinate of the cipher point that will be equal to Pm +KPb - (KG x nb). we know that Pb is equal to (nb x G). That's why Pm +KPb - KPb = Pm. so the receiver got Pm i.e. plain text point. In this way decryption is done in Elliptic Curve Cryptography (ECC).

In Table.1 we can see different asymmetric key algorithms with their key sizes in bits. As we know the key size in Elliptic Curve Cryptography (ECC) is less as compared to another asymmetric key algorithm such as Digital Signature Algorithm (DSA) and RSA. Where DSA and RSA key size are 152360 and key size of ECC is 512. Although the key size of ECC is less, it provides the same level of protection. It is a major advantage of elliptic curve cryptography (ECC).

DSA	RSA	ECC
1024	1024	160
2048	2048	224
3072	3072	256
7680	7680	384
15360	15360	512

Table 1. Public Key Size (Bits) Of Different Asymmetric Key Cryptography

5. RESULTS AND CONCLUSION

Performance Metrics

On evaluating our intrusion detection system using Random Forest with the CICIDS2017 dataset, the model achieved an impressive accuracy of 99.63%, providing effective results in classifying network traffic into normal and malicious categories.

Classification Report

For further analysis of the performance of our model, we used a detailed classification report which provides an overview of the model's performance across different metrics which include precision, recall, F1-score and support, for both normal and malicious categories.

Precision*:* It helps to measure the accuracy of positive predictions.

a. Precision=TP/ (TP + FP)

Recall (Sensitivity): It helps to measure the ability of the model to find all the positive instances.

b. Recall=TP/ (TP+FN)

F1-score*:* It helps to provide a balance between precision and recall where a higher F1-score indicates to us that the model is performing better.

c. F1-score=2× (Precision*Recall/Precision + Recall)

Support: It represents the count of actual instances of the class within the dataset, indicating the number of true occurrences for each class.

In conclusion, our research highlights the importance and effectiveness of integrating a Random Forest Classifier based Intrusion Detection System and using ECC (Elliptic Curve Cryptography) to enhance the security of IoMT (Internet of Medical Things) devices. The system shows high accuracy in detecting various types of cyber-attacks such as DoS attacks and Man in the Middle (MITM) attacks. Our work contributes to strengthening cybersecurity measures in healthcare, emphasizing the importance of robust intrusion detection and encryption techniques to protect the sensitive data in IoMT networks.

REFERENCES

[1] Ahmad A. Alzahrani, Using Artificial Intelligence and Cybersecurity in Medical and Healthcare Applications. Natural Sciences Publishing (NSP) 12(3), 1579–1590 (2023).

[2] V. Kumar, D. R. Recupero, D. Riboni and R. Helaoui, Ensembling Classical Machine Learning and Deep Learning Approaches for Morbidity Identification From Clinical Notes, in IEEE Access,9(), 7107–7126 (2021)

[3] A. Khan and I. Sharma, Machine Learning-Based Methodology for Preventing Ransomware Attacks on Healthcare Sector, International Conference on Research

Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India, 1–5 (2023)

[4] A. Tuteja, P. Matta, S. Sharma, K. Nandan and P. Gautam, Intrusion Detection in Health Care System: A logistic Regression Approach, 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 1794–1799, (2022)

[5] A. Ghourabi, A Security Model Based on LightGBM and Transformer to Protect Healthcare Systems From Cyber Attacks, in IEEE Access, 10(), 48890-48903, (2022)

[6] P. Radoglou-Grammatikis et al., Modeling, Detecting, and Mitigating Threats Against Industrial Healthcare Systems: A Combined Software Defined Networking and Reinforcement Learning Approach, in IEEE Transactions on Industrial Informatics,18(3) , 2041–2052 (2022)

[7] A. Andreas et al., ”Robust Encryption to Enhance IoT Confidentiality for Healthcare Ecosystems,” 2021 IEEE 26th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), Porto, Portugal, 1–6 (2021)

[8] A. Djenna, D. E. Saidouni and W. Abada, A Pragmatic Cybersecurity Strategies for Combating IoT-Cyberattacks, International Symposium on Networks, Computers and Communications (ISNCC), Montreal, QC, Canada, 1-6, (2020)

[9] M. R, G. K and V. V. Rao, ”Proactive Measures to Mitigate Cyber Security challenges in IoT based Smart Healthcare Networks,” 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, 1–4 (2021).

[10] B, Shanthi, M, Navaneetha, R, Kartheek and B D, Parameshachari , An Effective Public-key Authenticated Searchable Encryption for Protecting Sensitive data in Healthcare IoT, (2023)

[11] S. Ahmed, Z. Subah and M. Z. Ali, Cryptographic Data Security for IoT Healthcare in 5G and Beyond Networks, IEEE Sensors, Dallas, TX, USA, 1–4, (2022)

[12] K. Abu Ali and S. Alyounis, CyberSecurity in Healthcare Industry,International

Conference on Information Technology (ICIT), Amman, Jordan, 695–701, (2021)

Authors

Arnav Deshmukh