MJSAT

Malaysian Journal of Science and Advanced Technology

journal homepage: https://mjsat.com.my/

A Comparative Analysis of Medical IoT Device Attacks Using Machine Learning Models

Mubashir Mohsin[1], and Akinul Islam Jony* [1]

[1] *Department of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh.*

KEYWORDS

CICIoMT2024 Dataset
Cybersecurity
Intrusion Detection
IoMT
Machine Learning

ARTICLE HISTORY

Received 7 May 2024
Received in revised form 8 September 2024
Accepted 10 September 2024
Available online 29 September 2024

ABSTRACT

The Internet of Medical Things (IoMT) is revolutionizing healthcare by providing remarkable possibilities for remote patient monitoring, instantaneous data analysis, and customized healthcare delivery. However, the widespread use of interconnected medical devices has exposed vulnerabilities to cyber threats, posing significant challenges to the security, privacy, and accessibility of healthcare data and services. The CICIoMT2024 dataset is a crucial resource in IoMT security, offering a wide range of cyber-attacks targeting IoMT devices. This paper uses data balancing techniques like SMOTE and advanced machine learning (ML) models to analyze cyber threats on IoMT devices, aiming to improve healthcare system safety by identifying and mitigating cyberattacks. By conducting extensive experiments, the paper has determined the most effective ML models for three different levels of classification of the dataset: binary, multiclass, and multitype. Employing ML techniques like AdaBoost, Random Forest, kNN, and XGBoost proves to be extremely powerful in accurately categorizing various types of attacks. This study emphasizes the importance of proactive cybersecurity measures in IoMT ecosystems, as well as the effectiveness of ML techniques in protecting healthcare systems from evolving cyber threats.

1. INTRODUCTION

The Internet of Things (IoT), which has emerged as a disruptive force across numerous sectors, is profoundly reshaping interaction with both the digital and physical realms. IoT is a network that consists of physical devices, automobiles, appliances, and other items that are equipped with sensors, software, and network connectivity that enables the gathering and exchange of data [1]. IoT applications provide never-before-seen possibilities for productivity, efficiency, and convenience by allowing objects to exchange data and interact via the internet. IoT technology has brought forth notable progress, especially in the fields of healthcare, urban planning, and automation in industries. Moreover, IoT devices produce enormous volumes of data, which makes them more fascinating since it makes it easier to analyze the data and make better decisions for the industry. Nevertheless, the growing utilization of IoT devices also exposes security risks. Worldwide, there were more than 112 million incidents of cyber assaults on the IoT in 2022 [2]. Multiple industries that heavily rely on IoT devices have recognized a significant number of attacks originating within the IoT network. Heavy industry-level IoT and control system applications are especially vulnerable to threats including active-passive eavesdropping, Man-in-the-Middle (MitM), masquerade, DoS and DDoS, spoofing, phishing, viruses, ransomware, protocol attacks, reconnaissance, and supply chain attacks [3, 4]. These risks include the possibility of data breaches, illegal accessibility, and denial-of-service attacks, which jeopardize user safety and privacy while also putting the integrity of critical systems in jeopardy. Hence, it is crucial to prioritize the resolution of these security concerns in order to fully use the capabilities of IoT applications.

The healthcare sector is one that depends more and more on technology innovation than other industries. Its medical applications and control systems have a noticeable IoT integration. The Internet of Medical Things (IoMT) is a

*Corresponding author:
E-mail address: Akinul Islam Jony <akinul@aiub.edu>.
https://doi.org/10.56532/mjsat.v4i4.318

distinct subset of the broader IoT that is specifically dedicated to healthcare and medical applications. Devices for remote patient monitoring, wearable fitness trackers, and advanced diagnostic systems are just a few examples of the wide range of technologies that make up IoMT devices. With the use of these tools, patients and healthcare professionals may communicate directly and continuously, allowing for real-time health monitoring, data collection, and analysis. Pradhan et al. [5] classify healthcare IoT technologies into three key domains: identification, communication, and location technologies. This three-part categorization, which seeks to expand the predominance of smart technology-enabled advanced healthcare systems, includes large databases, servers, cloud integration, network streams, and the inclusion of different service devices and control systems. Ensuring a reliable and effective smart healthcare framework is contingent upon the proper distribution, administration, and control mechanisms of these technologies. Therefore, it is crucial to prioritize the necessity for recognition and mitigation of the risks posed by the cyber-attacks associated with IoT in this regard. Protecting patients and everyone involved in the healthcare system depends on promptly detecting and eliminating such risks.

In the field of IoMT security, the CICIoMT2024 dataset [6] is a trailblazing benchmark that represents a coordinated effort to enhance the development and validation of security solutions tailored for healthcare systems. This dataset contains the outcomes of 18 well-planned attacks against an IoMT testbed, including 40 real and simulated devices. The dataset captures the diversity and intricacy of healthcare's digital infrastructure. The dataset incorporates multiple protocols such as Bluetooth, MQTT, and Wi-Fi, underscoring its diversity and alignment with real healthcare communication standards. We methodically categorize the attacks into five major categories: DDoS, DoS, Recon, MQTT, and Spoofing, enabling an organized approach to analysis and mitigation. The primary objective of the CICIoMT2024 dataset is to improve the security of healthcare systems, making it a very important resource for both researchers and practitioners in the field. The seminal work by Dadkhah et al. [7] delineates the CICIoMT dataset, which stands as a crucial resource curated from diverse IoT devices, providing a comprehensive overview of the methodology employed. The rigorous planning, implementation, and data collection represent significant advancements in the field of medical IoT. Dadkhah et al. [7] emphasize the importance of the CICIoMT2024 dataset, which was an important contribution to the IoMT dataset because it provided a thorough collection of real-time attacks on IoMT devices as well as extensive IoMT profiling. This groundbreaking endeavor greatly improves the current state of the IoMT dataset landscape. This underscores the urgent need for robust and adaptable security solutions to safeguard the confidentiality and precision of medical records and services, particularly as our world becomes increasingly interconnected. Our primary aim is to employ various machine learning models for dataset analysis, effectively detecting diverse attack classes, thereby contributing to enhancing the security of IoMT devices within healthcare facilities.

2. LITERATURE REVIEW

The application of IoT technology in healthcare offers a substantial array of devices for both patients and healthcare workers. These devices gather tremendous amounts of data during different phases of their functioning and preserve confidential patient information, presenting significant security risks. In keeping with the many uses and intricacies of medical IoT, the produced datasets cover a broad range of categories. IoT device cyberattacks take advantage of flaws in operational controls and communication protocols, which expose critical industries to significant risk and could lead to catastrophic effects including loss of data, service interruption, or even complete data destruction [4]. Hussain et al. [8] commonly employ traditional security measures either at the network or host level. While host-level security is often stronger, the limited resources and processing capabilities of IoT devices frequently make it hard to implement. This means that, as network-based cyberattacks are the most common danger to the security of healthcare IoT data, network-based security techniques are preferred for safeguarding IoT items.

Over an extensive period of time, researchers have aggregated network traffic data from diverse IoT devices, meticulously capturing, organizing, and analyzing various forms of attacks to construct datasets tailored for security purposes and intrusion detection systems [9]. Prominent examples of such attack datasets include the TON IoT Datasets [10], the IoT Network Intrusion Dataset [11], the RT IoT 2022 dataset [12], the Bot-IoT dataset [13], and the CoAP-DoS dataset [14]. The WUSTL EHMS 2020 Dataset [15], the ECU-IoHT dataset [16], BlueTack [17], the ICU dataset [8], the IEC dataset [18], CIC-IoT2023 [19], and the CIC IoMT 2024 dataset [7] are some of the most important IoMT attack datasets. The Canadian Institute of Cybersecurity [6] manages the CICIoMT 2024 dataset, which stands out as a recent and extensive IoMT dataset [7]. The authors of the study stress how important the CICIoMT2024 dataset is, pointing out that it has a lot of real IoMT devices that can be attacked in different ways and that it could be used to make full IoMT pro- files. The dataset comprises network traffic data from 40 medical IoT devices, consisting of 25 authentic and 15 simulated devices. The dataset encompasses 18 distinct types of IoT attacks, categorized into five distinct attack categories: DoS, DDoS, MQTT, spoofing, and reconnaissance. The collection also contains device profiling, allowing the identification of issues with specific devices at various points in their lifespan throughout the healthcare network.

In their study, Dadkhah et al. [7] discuss the significant contributions of their research, which include the creation of a large IoMT attack dataset, the utilization of novel methods to simulate attacks on the IoMT, the development of IoMT life-cycle profiles to enhance comprehension, and the application of several models to assess the dataset from various perspectives. However, they see the possibility of future improvement, especially by delving further into machine learning (ML) algorithms and approaches. By utilizing state-of-the-art ML algorithms and evaluation metrics, our research aims to enhance this assessment. Specifically, we focus on the six-category classification present in the CICIoMT2024 dataset: Benign, DoS, DDoS, MQTT, Recon, and Spoofing. Our work aims to aid in the efficient identification of various attack classes within the IoMT environment by utilizing cutting-edge ML algorithms with an optimal selection of parameters and extensive assessment metrics.

Anwer et al. [20] compiled a thorough analysis of the methods used to identify ML attacks on the IoT. The analysis

provides six main approaches, notably supervised learning, unsupervised learning, ensemble learning, semi-supervised user learning, reinforcement learning, and active learning. For our study, we've decided to use supervised learning in our ML methods to detect threats in the IoMT space. The choice is based on the meticulous organization of the CICIoMT dataset, which provides separate CSV files that outline various types of attacks. Many algorithms for ML have been extensively used for the detection of IoT attacks. These algorithms include Random Forest [7,20,21,22,23], Logistic Regression [7,22,23], k-Nearest Neighbors [21,23,24], Naive Bayes [21,23], Support Vector Machines (SVM) [20,21,24], Gradient Boosting [7,20,22], LSTM-based model [19], and Neural Networks [7,21,22,24]. The objective of our research is to comprehensively evaluate these ML algorithms by employing them to detect IoMT attacks. Subsequently, we will meticulously evaluate the performance of several ML models through rigorous testing utilizing diverse techniques in order to choose the most effective one.

3. METHODOLOGY

3.1 Dataset Information

The CICIoMT 2024 dataset [6] is a complete benchmark for evaluating the security of IoMT devices used in healthcare facility scenarios. The dataset comprises simulated instances of actual attacks on a testbed consisting of 40 IoMT devices. This ensures the seamless integration of these devices into vital healthcare infrastructure. This dataset uses attack simulation to evaluate the widely used Bluetooth Low Energy (BLE), WiFi, and MQTT protocols in healthcare. Creating a practical benchmark dataset was the primary objective of that project in order to facilitate the creation and evaluation of IoMT security solutions.

The CICIoMT2024 dataset, a tabular dataset in CSV formats, contains information relevant to cybersecurity incidents classified into three levels: binary (based on benign and attacks), categorical (based on six distinct classes), and attacks (based on 19 attack types). The zipped dataset we obtained from the official website of the dataset [6] was structured with two folders, 'train' and 'test', each containing 51 and 21 CSV files, respectively. Each entry in the dataset corresponds to a singular occurrence of a cybersecurity instance, and there are 45 columns that include various features linked to each instance. The study of Dadkhah et. al. [7] extensively covers the specific characteristics and statistics of the features. Our study focused only on the Wi- Fi/MQTT data in the dataset, conducting a variety of attacks against MQTT-simulated devices and Wi-Fi-equipped IoMT devices.

In terms of class categorization, the dataset can be divided into three classification levels into distinct classes. The binary categorization divides the data into benign (non-attack) and attack categories. The multiclass classification classifies the data into six specific categories: DDoS (Distributed Denial of Service), DoS (Denial of Service), MQTT (Message Queuing Telemetry Transport), benign (non-attacks), recon (Reconnaissance), and spoofing. The categorization of the data into multiple attack types, which produced a total of 19 classes for the multitype classification, was one of the most meticulous contributions to the dataset. These classifications are the most common types of cybersecurity risks encountered in IoT-based network infrastructures.

Upon closer examination of the dataset and its data for each classification level, a notable imbalance between classes becomes apparent, as visually depicted in the supplementary bar plot in Figure 1. The dataset exhibits a significant class imbalance at each level of classification, with the attack classes (e.g., DDoS and DoS type classes) being the most prevalent. This indicates an unequal distribution of cybersecurity incidents across various classes at different classification levels. The presence of class imbalance presents difficulties in training and evaluating ML models since these models might exhibit biases towards the dominant class, resulting in poorer outcomes for the minority classes. The importance of the CICIoMT2024 dataset goes beyond its originality, as it possesses vast potential to tackle critical cybersecurity concerns in the healthcare domain. The dataset is a fundamental resource that researchers, as well as professionals, can use to build and assess strong security solutions that protect healthcare systems from emerging and forthcoming cyberattacks.

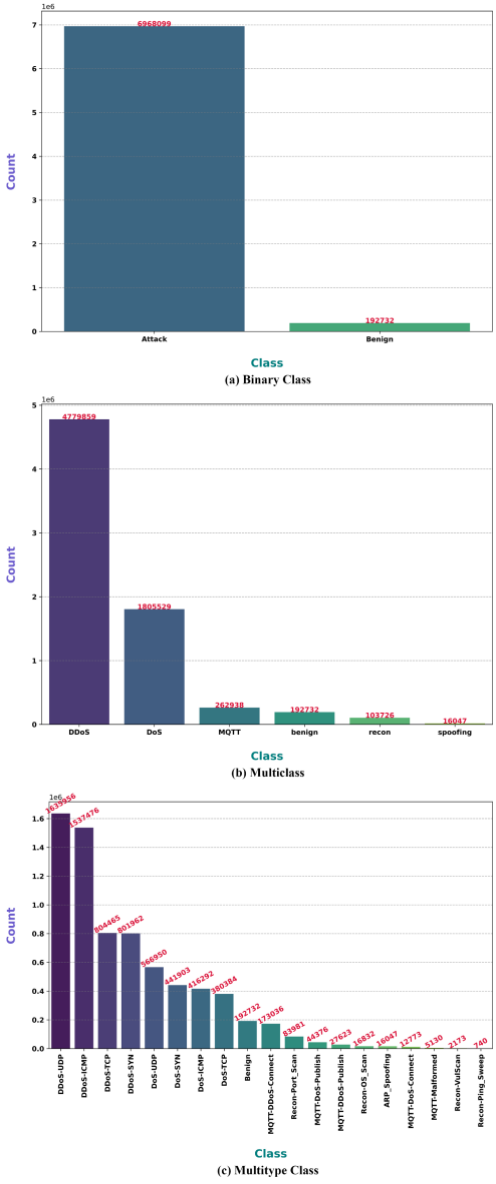

Fig. 1. Training data class imbalances across three classification levels

3.2 Data Preprocessing

Following the multitype classification level, the downloaded dataset comprised training and test data in several directories, with many CSV files based on 19 distinct class types inside each folder. After reviewing the CSV files and multiple class types, we discovered no feature mismatches. Additionally, we discovered that the dataset had no null or missing values, obviating the need for additional data cleansing. Following statistical analysis and visualization of features' unique value densities, we discovered that the feature 'Drate' had a value of zero in every dataset instance; hence, we removed it from the training data. Next, we implemented the data segregation procedure to classify the dataset into three distinct levels. Based on the requirements of the classification level, we separated the training and test data and created three levels of distinct databases: binary, multiclass, and multitype. Subsequently, we combined the training data into a single CSV file and the test data into a separate CSV file for each of the databases. In order to streamline our machine learning approach, we inserted a new column named 'class' into the combined dataset. We assigned the various class labels as values for each of the three categorization levels. Different classes from all three levels of classification are shown in Table 1. The combined training data for each categorization level exhibited a significant data imbalance, as seen in Figure 1. To address the severely unbalanced dataset, we next put the SMOTE balancing approach into practice, as covered in the next subsection 3.3. We meticulously organized the data and reduced the complexity of the class labels by encoding them to facilitate the assessment process after the machine learning algorithm was trained.

Table 1. Three level of Categorization of the CICIoMT2024 dataset

Binary Class	Multi Class	Multitype Class
Non-Attack	Benign	Benign
Attack	DoS	DoS TCP
		DoS ICMP
		DoS SYN
		DoS UDP
	DDoS	DDoS TCP
		DDoS ICMP
		DDoS SYN
		DDoS UDP
	Recon	Ping Sweep
		Recon VulScan
		OS Scan
		Port Scan
	MQTT	Malformed Data
		DoS Connect Flood
		DDoS Connect Flood
		DoS Publish Flood
		DDoS Publish Flood
	Spoofing	ARP Spoofing

3.3 Data Balancing

One of the most common challenges in machine learning, especially in classification tasks, is the uneven distribution of classes within datasets. In the CICIoMT2024 dataset, Figure 1 clearly illustrates the substantial class imbalance across the three categories: binary, multiclass, and multitype. ML models are severely hampered by this imbalance, mainly because of their innate bias in favor of the majority class. Consequently, training models on datasets with an uneven distribution often results in poor performance and erroneous predictions. An imbalanced distribution of classes in the CICIoMT2024 dataset might lead to biased model outputs while building and evaluating security solutions. This issue weakens the trustworthiness and effectiveness of security systems, thereby jeopardizing the precision and reliability of healthcare cybersecurity measures that depend on this type of dataset. To make sure that the security solutions derived from the CICIoMT2024 dataset work and are reliable, it is important to address the issue of class imbalance. We utilize the Synthetic Minority Over-sampling Technique (SMOTE) to rectify the imbalance in the CICIoMT2024 dataset. SMOTE is a widely used technique for addressing the issue of unbalanced datasets by creating artificial samples for minority classes. SMOTE addresses the issue of class imbalance by generating synthetic instances along the line segments that connect the nearest neighbors of k individuals from a minority class. This approach minimizes the emergence of bias while effectively resolving the problem. We expect that SMOTE will improve the robustness and generalization capacity of ML models trained on the CICIoMT2024 dataset by mitigating imbalances across binary, multiclass, and multitype categories.

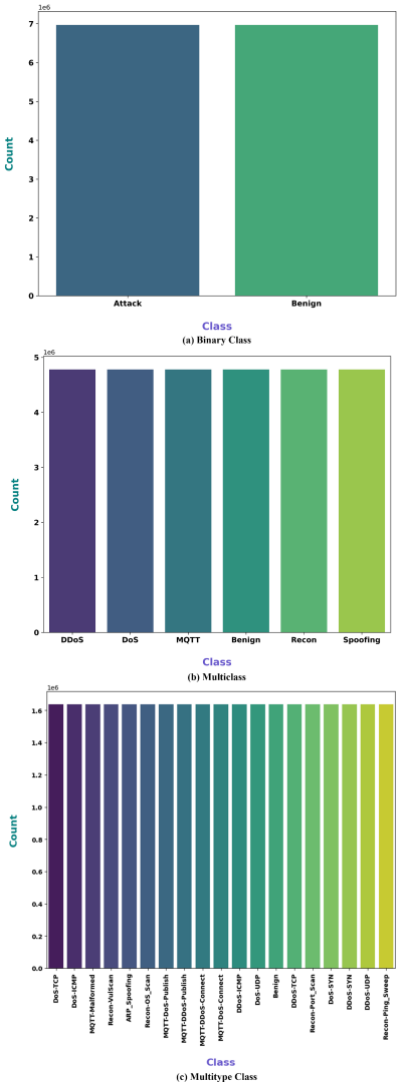

Fig. 2. Training data class balanced across three classification levels after applying SMOTE

The application of SMOTE to the CICIoMT2024 dataset results in a substantial reduction in class imbalance across all categories. Specifically, before SMOTE, the binary class exhibited a significant gap of over 6.7 million samples between the two classes. However, after applying SMOTE, both classes achieve balance with a total of 6.9 million samples each. Similarly, SMOTE balances all six classes in the multiclass category, which had class im- balances ranging from 3 to 4.7 million samples. Additionally, in the multitype class, which originally had imbalances of up to 1.6 million samples between classes, SMOTE achieves balance across all 19 classes, with each class containing a total of 1.6 million samples. Figure 2 shows the evenly distributed classes across three different class categories after applying SMOTE. This transformation not only addresses the inherent biases in the dataset but also enhances its suitability for training robust and reliable machine learning models for IoMT security applications. In conclusion, the application of the SMOTE technique effectively mitigates class imbalance within the CICIoMT2024 dataset, thereby improving the dataset's suitability for training machine learning models.

3.4 Machine Learning Models

In this study, we attempted to apply the most widely used and appropriate machine learning algorithms in many research studies (such as [25] and [26]) to evaluate the dataset at three distinct categorization levels. We thus picked seven machine learning algorithms: AdaBoost, k-Nearest Neighbors (k-NN), Logistic Regression (LR), Naive Bayes, Random Forest (RF), ANN-based Support Vector Machine (SVM-ANN), and XGBoost. To help understand our work better, we have included a brief discussion of the algorithm explanations below.

AdaBoost (Adaptive Boosting): AdaBoost is an ensemble learning technique that combines numerous weak learners to produce a powerful classifier. The way it operates is that weak learners are successively fitted to learners that are modified frequently, and the sum of all the weak learners yields the final prediction. AdaBoost prioritizes the most challenging situations for upcoming learners by giving more weights to occurrences that are erroneously identified. A weighted total of weak learners, with each learner's contribution determined by accuracy, makes up the final model. The AdaBoost algorithm can be understood by the equation below:

$$F(x) = \sum_{t=1}^{T} \alpha_t f_t(x) \tag{1}$$

Where F (x) is the final prediction function, t is the weight assigned to weak learner ft(x), and T is the number of weak learners.

k-NN (k-Nearest Neighbors): An algorithm for non-parametric classification that divides instances into groups according to the majority class of their *k* nearest neighbors. The process begins with calculating the distance between each training instance and the test instance, after which the k nearest neighbors are determined. The class label of the test instance is determined by a majority vote among its neighbors.

Logistic Regression: It is a linear classification approach that uses a logistic function to estimate the likelihood of a binary result. It calculates the likelihood that a specific example is a member of a specific category based on its characteristics. Logistic regression employs a logistic function to estimate the likelihood and generates predictions by applying a threshold. The following equation provides a thorough understanding of the logistic regression algorithm:

$$p(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}} \tag{2}$$

Where *p(y = 1x)* is the probability that the target variable *y* is 1 given the input *x, 0, 1, ..., n* are the coefficients of the features *x1, x2, ..., xn*, and *e* is the base of the natural logarithm.

Naive Bayes: It is a probabilistic approach to classification that relies on the notion of feature independence and is based on Bayes' theorem. The algorithm computes the likelihood of each category based on a given set of characteristics and chooses the category with the greatest likelihood. The following equation provides a thorough understanding of the algorithm:

$$P(C_k|x) = \frac{P(x|C_k)P(C_k)}{P(x)} \tag{3}$$

Where *P (Ck|x)* is the posterior probability of class *Ck* given the features *x, P (x|Ck)* is the likelihood of the features given class *Ck, P (Ck)* is the prior probability of class *Ck*, and *P (x)* is the probability of the features.

Random Forest: It constructs many decision trees throughout the training process and produces the most common class or numerical prediction where decision trees serve as the fundamental building blocks by recursively splitting the dataset based on feature values and constructing the tree, aiming to create homogeneous subsets of data at each node. The purpose is to generate sets of data that are similar at each node. During the training process, multiple decision trees are built, with each tree employing a random subset of the training data and characteristics.

SVM-ANN (Support Vector Machine with Artificial Neural Networks): It is a hybrid model that combines the strengths of Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs). In our implementation, the SVM component is represented by a single fully connected layer defined within the SVM class. This layer is initialized with a linear transformation from the input features to the number of classes, leveraging the margin maximization property inherent to SVMs. This hybrid approach aims to capture both the non-linear relationships in the data, enabled by ANNs, and the margin-based classification characteristic of SVMs, thereby enhancing the model's capability for effective classification.

XGBoost (Extreme Gradient Boosting): This enhanced gradient boosting algorithm creates numerous decision trees progressively to repair prior errors. With a more regularized model formulation than gradient boosting approaches, XGBoost controls overfitting and ensures model generalization to unseen data. Adding penalty terms to the loss function penalizes complex models, promoting simpler ones. The advanced capabilities of XGBoost include tree pruning and column subsampling. These methods improve model efficiency by reducing computation time and memory usage without affecting performance.

3.5 Evaluation Matrices

Accuracy: It is the ratio of accurately predicted instances to the total number of instances. It provides a general summary of model performance but may not work for imbalanced datasets.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Predictions}} \quad (4)$$

Precision: It shows the percentage of all expected positive cases that were accurately predicted. It emphasizes the model's capacity to prevent false positives, which is crucial in situations where false positives might be expensive.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (5)$$

Recall (Sensitivity): It quantifies the accuracy of accurately predicting positive instances relative to the total number of actual positive **cases**. The model's capacity to accurately identify all positive occurrences is brought out, which is of utmost importance in scenarios where overlooking a positive example can result in substantial consequences.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (6)$$

F1-Score: It provides equilibrium between the two measures and is the harmonic mean of recall and precision. Considering both false positives and false negatives is particularly beneficial when working with unbalanced datasets.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

4. RESULTS

4.1 Binary Class Classification Results

Seven different machine learning models were evaluated to classify instances as either "Attack" or "Benign" in the binary class classification results. The confusion matrices in Figure 5 provide a visual representation of the performance of each model. A higher concentration along the diagonal indicates more accurate predictions. With an astounding accuracy of 99.81%, AdaBoost was one of the best-performing models; kNN came in next with an equally outstanding accuracy of 99.66%. The SVM-ANN model demonstrated strong performance, achieving an accuracy of 97.31%.

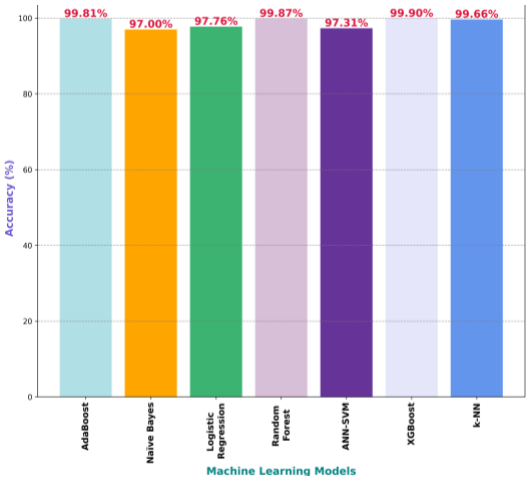

Fig. 3. Test accuracy of Binary classification in models

This highlights its effectiveness in accurately classifying instances. Alternatively, Logistic Regression showed a respectable performance, although with a slightly lower accuracy of 97.76%— almost on par with Naive Bayes' 97%. In addition, Random Forest showed an impressive accuracy of 99.87%, which further solidifies its dependability in binary classification tasks. Surprisingly, XG Boost outperformed all the other models, achieving an exceptional accuracy rate of 99.90%. The test accuracy attained by the models can be seen clearly in Figure 3.

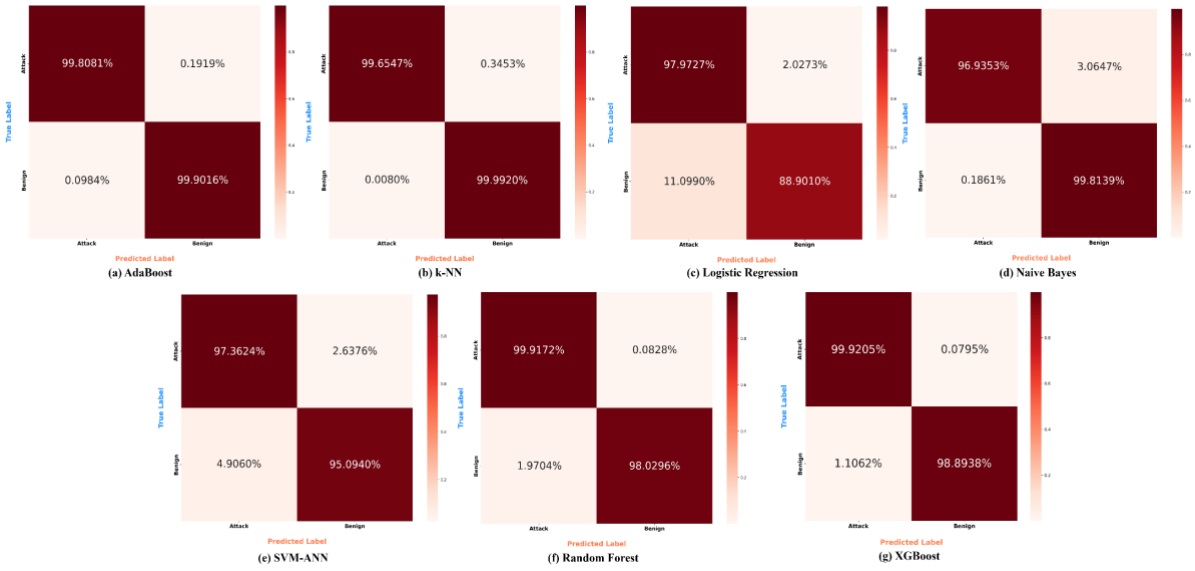

Fig. 5. Confusion Matrices of all seven ML models for Binary classification

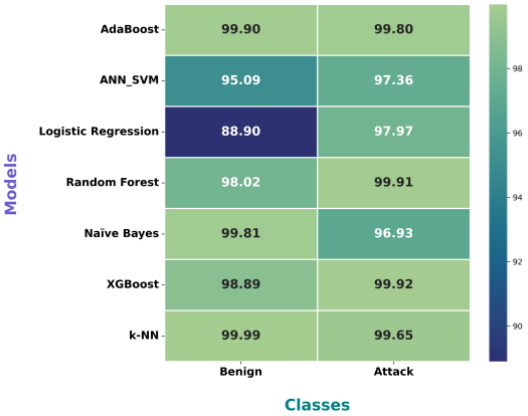

Fig. 4. Model accuracy for each of the binary classes

The analysis of accuracy for each model by class is depicted in Figure 4. Most samples in the benign class were correctly identified using kNN, whereas the majority of samples in the malignant class were correctly diagnosed using Random Forest and XGBoost models. These comprehensive evaluations help in making well-informed decisions when choosing the most appropriate algorithm for future binary classification endeavors. The results highlight the effectiveness of the machine learning models in accurately distinguishing between "Attack" and "Benign" instances. AdaBoost, Random Forest, and XGBoost stand out as the top performers in terms of overall accuracy and performance.

4.2 Multiclass Classification Results

We evaluated the performance of seven machine learning models across six distinct classes: Benign, DDoS, DoS, Recon, MQTT, and Spoofing, for the multiclass classification results. Normalized confusion matrices in Figure 8 were used to test how well each model could predict, which showed how accurate each class's classification was. The AdaBoost model exhibited strong performance across all classes, achieving an overall accuracy of 99.66%. Notably, it demonstrated exceptional accuracy in predicting Benign, DoS, and DDoS instances, with accuracy rates of 99.68%, 99.82%, and 99.88%, respectively, which can be seen in Figure 7. However, it showed slightly lower accuracy for the Recon class at 94.96% and poor accuracy in Spoofing class at 49.94% only. On the other hand, the ANN-SVM model was very poor at classifying instances, getting them wrong at zero percent accuracy for most of the classes. Logistic Regression demonstrated moderate performance, achieving accuracies ranging from 10.84% for the MQTT class to 99.73% for the DDoS class and zero accuracy for the DoS and Recon classes. Random Forest emerged as one of the top-performing models, achieving high accuracy across all classes, with notable performance in predicting DDoS and MQTT instances at 99.99% and 99.53%, respectively. Naive Bayes demonstrated relatively lower accuracy for the MQTT and Spoofing classes at 4.49% and 53.55%, respectively. XGBoost exhibited robust performance, particularly in predicting Benign and DoS instances, with accuracies of 97.83% and 99.84%, respectively. Additionally, k-NN achieved high accuracy across all classes, surpassing 89% accuracy for each class except Spoofing, with notable performance in predicting Recon instances at 93.86%. Among the seven ML models, AdaBoost, Random Forest, kNN, and XGboost performed very well, with 99.6%, 99.85%, 99.79%, and 99.33%, respectively, as shown in Figure 6. These results underscore the varying effectiveness of ML models in classifying multiclass data, with some models outperforming others across different classes.

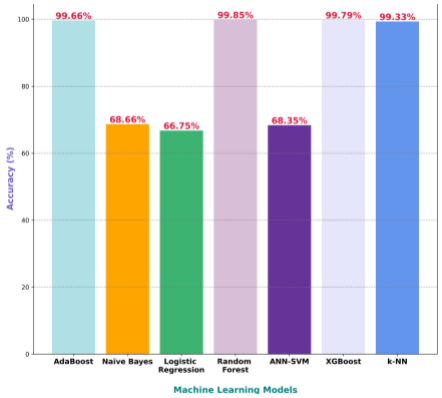

Fig. 6. Test accuracy of Multiclass Classification in models

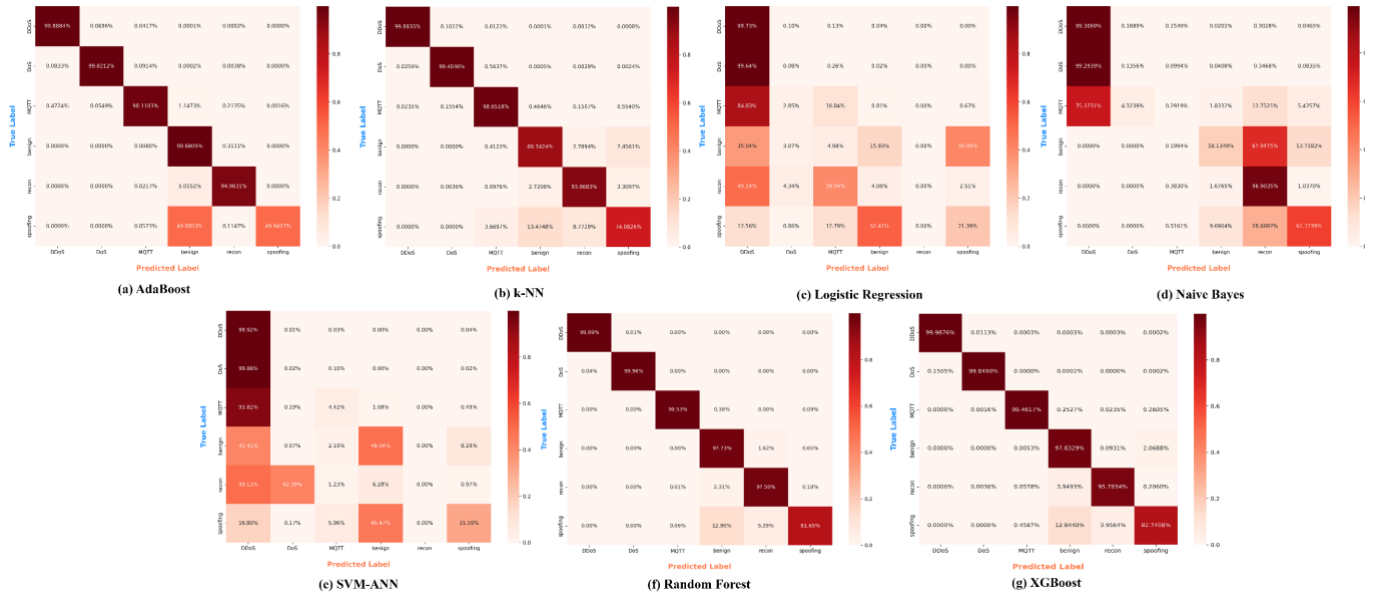

Fig. 8. Confusion Matrices of all seven ML models for Multiclass classification

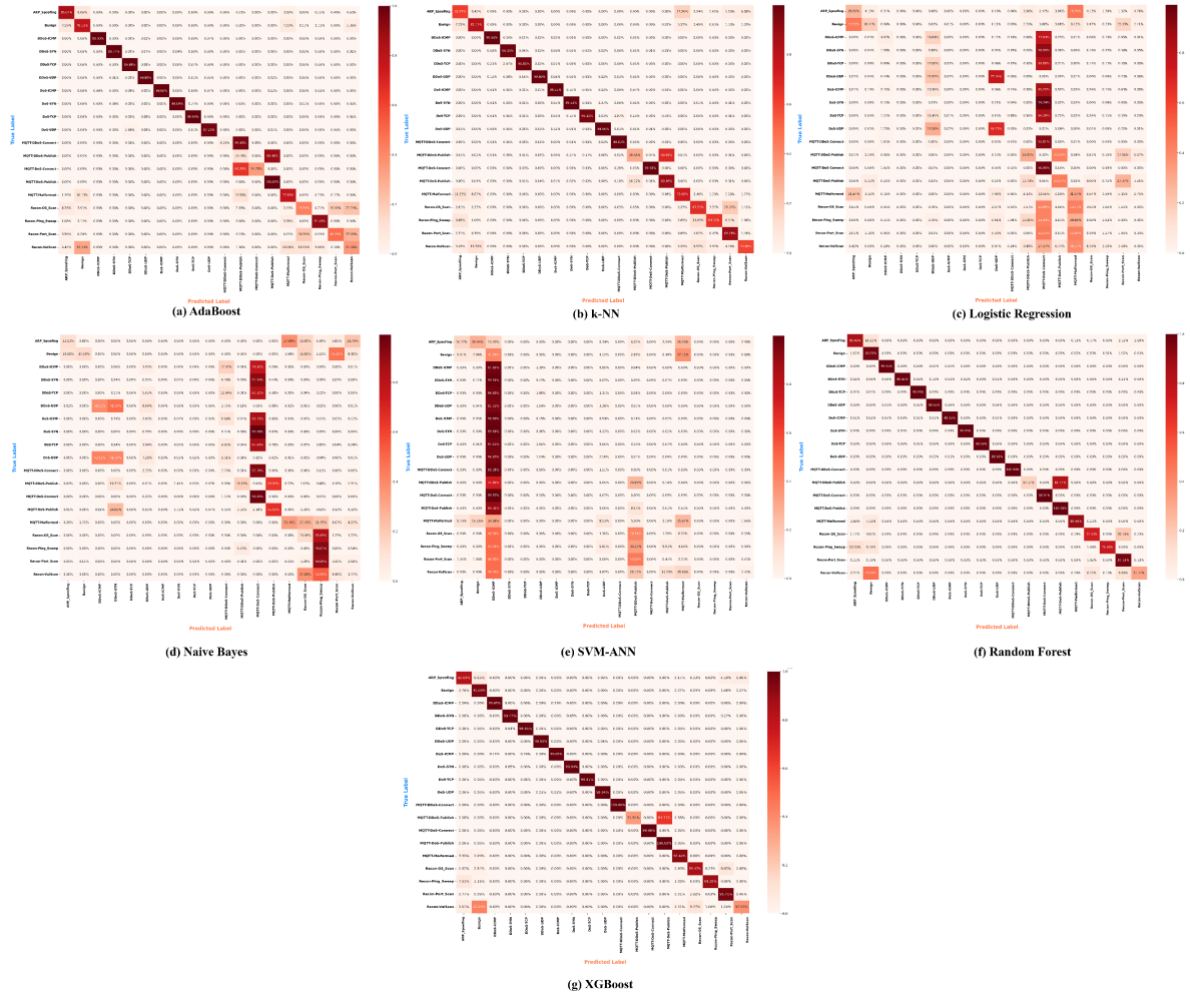

Fig. 9. Confusion Matrices of all seven ML models for Multitype classification

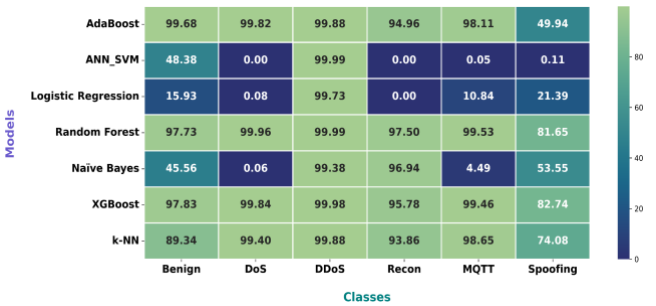

Fig. 7. Model accuracy for each of the multiclass classes

4.3 Multitype Class Classification Results

We assessed the performance of seven machine learning models across a diverse set of 19 distinct classes for the multitype classification results. Each model's predictive capabilities were evaluated using normalized confusion matrices in Figure 9, providing detailed insights into the classification accuracy for each class. The AdaBoost model demonstrated strong performance overall, achieving an accuracy of 95.06%. Notably, it was very good at predicting certain types of attacks, like DoS and DDoS-based attacks, with average 99.90% accuracy rates, respectively, and it was also very good at classifying MQTT-DoS-Publish instances with 100% accuracy. Conversely, the ANN-SVM model showed limited effectiveness, particularly for classes like Recon and MQTT, with accuracies of nearly zero percent accuracy. Logistic Regression had the poorest performance across classes, with accuracies ranging from maximum at zero to 68.75% for DoS-UDP instances, as we can see in Figure 11. Random Forest emerged as one of the top-performing models along with XGBoost, achieving high accuracy across most classes, with notable performance in predicting DDoS and MQTT instances with average 99.9% and 99.99% accuracy, respectively. Naïve Bayes demonstrated relatively lower accuracy than ANN-SVM for classes such as Recon and MQTT-based attack types. k-NN also had great accuracy across all classes, beating 60% accuracy for all but MQTT-DDoS-Publish. It did especially well at predicting benign, DoS, and DDoS instances. Figure 10 exhibits that Random Forest and XGBoost are the best at classifying multitype classes. These results underscore the varying effectiveness of machine learning models in classifying multitype data, with certain models demonstrating superior performance across different classes.

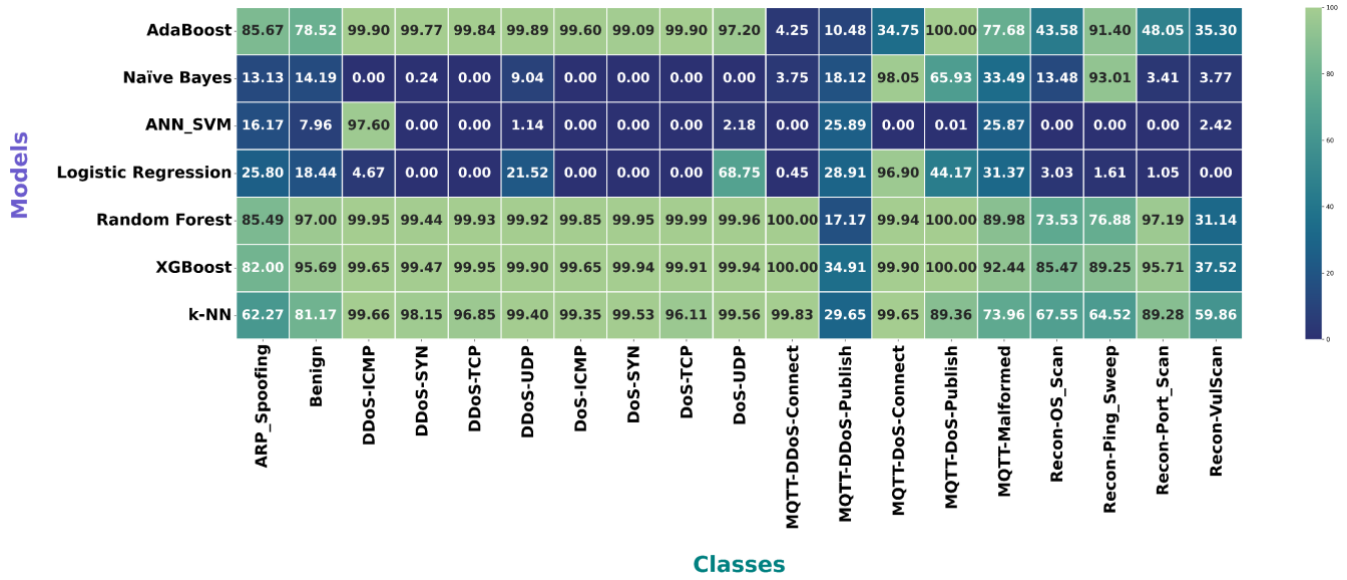

Fig. 11. Model accuracy for each of the multitype classes.

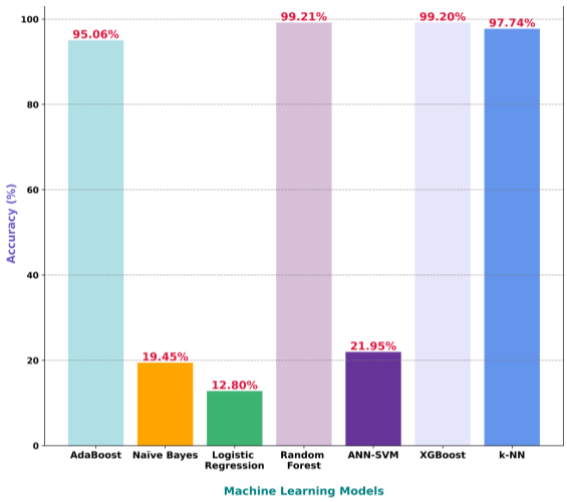

Fig. 10. Test Accuracy of Multitype Classification in Models

5. DISCUSSION

The models displayed different levels of performance at the binary classification level. AdaBoost, Random Forest, and XGBoost were the top-performing models, with accuracy rates of 99.81%, 99.87%, and 99.90%, respectively. These models showed impressive abilities in differentiating between attacks and non-attack instances. ANN-SVM and Logistic Regression demonstrated slightly lower accuracies of 97.31% and 97.76%, respectively. However, they still performed well in effectively classifying binary data. k-NN achieved a high accuracy of 99.66%, showcasing their strong performance in the diverse landscape. Nevertheless, the accuracy of ANN-SVM was the lowest at 99.66%. Overall, all seven models performed very well in binary classification among attack and benign samples, with all of them surpassing 97% accuracy for this classification task. This shows the high resiliency and efficiency of the ML models in classifying binary class attributes in the CICIoMT2024 dataset.

For multiclass classification, KNN, AdaBoost, XGBoost, and Random Forest emerged as the top-performing models, with accuracy rates of 99.33%, 99.66%, 99.79%, and 99.85%, respectively. These models showed exceptional performance in accurately categorizing in- stances across various classes, showcasing their adaptability and efficiency in dealing with a wide range of data categories. The performance of Logistic Regression, ANN-SVM, and Naive Bayes was moderate, with accuracies ranging from 66.75% to 68.35%. Furthermore, these models have lower accuracy; among the six classes, such as DoS and MQTT, they were unable to classify relatively few class instances. It is also clear that not every ML model was effective in classifying the multiclass features in the dataset.

Random Forest and XGBoost have demonstrated their superiority as the most effective models in multitype classification, attaining an exceptional accuracy rate of 99.2%. This illustrates their remarkable ability to correctly categorize instances in each of the 19 classes. In addition, both kNN and AdaBoost demonstrated remarkable performance, attaining respective accuracies of 96.74% and 95.06%. The models proved to be versatile and efficient in handling complex situations with a variety of data categories. Accuracy was lower for Logistic Regression, Random Forest, and k-NN, indicating rather poor performance at this level of classification. Additionally, the analysis reveals that as the number of classes increased, the effectiveness of Naive Bayes, ANN-SVM, and Logistic Regression models in classifying classes decreased proportionally. This indicates that these models did not exhibit resilience across the three classification levels, unlike AdaBoost, Random Forest, kNN, and XGBoost models.

Overall, Random Forest, AdaBoost, kNN, and XGBoost outperformed all the other models in all three classification levels, positioning them as strong contenders for the top models. Their impressive performance emphasizes their ability to handle a wide range of classification tasks within the CICIoMT024 dataset, demonstrating their versatility and reliability with medical IoT device attack classification. Therefore, these four models could be used for further advancement to classify all of the classes efficiently by

implementing ensemble learning or hybrid ML model learning.

6. CONCLUSION

This study aims to thoroughly investigate the security aspects of medical IoT devices, namely in the healthcare industry, using the cutting-edge CICIoMT2024 dataset. The emergence of the IoT has culminated in significant and extensive improvements in numerous fields, offering exceptional efficiency and convenience while also giving rise to concerns over cybersecurity. Given the rapid increase in IoT usage, it is crucial to prioritize the protection of these interconnected devices from cyber threats. Our study makes a substantial contribution to this effort by specifically addressing the security concerns that are inherent in IoMT systems and presenting strong solutions that utilize ML techniques. Our work aimed to improve healthcare systems' security by identifying and reducing cyber threats using sophisticated ML models. After a thorough analysis of the CICIoMT2024 dataset, we have identified the most optimal ML models for three distinct classification tasks: binary, multiclass, and multitype. This dataset comprises a diverse array of genuine and simulated attacks on IoMT devices. The results of our study highlight the efficacy of ensemble learning techniques, such as AdaBoost, Random Forest, kNN, and XGBoost, in accurately categorizing occurrences across various attack types. These models have shown exceptional ability to recover quickly and adjust to new situations, highlighting their potential for use in real-world scenarios to strengthen the security of medical IoT ecosystems.

Although our study has been successful, it is important to acknowledge several limitations that should be considered for future research efforts. The use of a single dataset, which might not adequately capture the wide variety of cyberthreats present in IoMT environments, is one significant drawback. For example, we have gathered the WiFi and MQTT attack data from the CICIoMT website; nevertheless, our study does not classify all forms of attacks on IoMT devices because it does not include the Bluetooth traffic data. Furthermore, the effectiveness of ML models can vary depending on the dataset's unique attributes and types of attacks. In order to overcome these restrictions, future research could focus on creating more extensive datasets and investigating innovative ML techniques specifically designed to tackle the distinct issues presented by security in medical IoT. Overall, our study is a groundbreaking endeavor to enhance the security of IoMT devices in healthcare facilities. With the help of the CICIoMT2024 dataset and advanced ML methods, our research demonstrated that using hybrid learning can effectively protect medical IoT networks from cyberattacks and make them more resilient. Our research lays the foundation for future studies focused on developing robust security solutions that safeguard the privacy, accuracy, and accessibility of healthcare data and services in an increasingly linked world.

DATA AVAILABILITY STATEMENT

The CICIoMT2024 dataset employed in this study was obtained from the website of the Canadian Institute for Cybersecurity at the University of Brunswick. The dataset was viewed and downloaded on March 14, 2024. To access the dataset and learn more, follow this link: https://www.unb.ca/cic/datasets/iomt-dataset-2024.html.

ACKNOWLEDGEMENT

The authors wish to clarify that this research was carried out independently, without receiving any external funding or assistance. All materials utilized in the project, including hardware and software, were privately owned, and managed by the authors. The authors are grateful for the opportunity to undertake this research autonomously and for the resources available to them.

REFERENCES

[1] "What is the Internet of Things (IoT)? | IBM." [Online]. Available: https://www.ibm.com/topics/internet-of-things.

[2] "Global annual number of IoT cyber attacks 2018-2022," Statista, May 03, 2023. [Online]. Available: https://www.statista.com/statistics/1377569/worldwide-annual-internet-of-things-attacks/. [Accessed: Apr. 30, 2024]

[3] A. I. Jony and S. A. Hamim, "Navigating the Cyber Threat Landscape: A Comprehensive Analysis of Attacks and Security in the Digital Age," Journal of Information Technology and Cyber Security, vol. 1, no.2, pp. 53–67, 2023, doi: https://doi.org/10.30996/jitcs.9715.

[4] K. Tsiknas, D. Taketzis, K. Demertzis, and C. Skianis, "Cyber Threats to Industrial IoT: A Survey on Attacks and Countermeasures," IoT, vol. 2, no. 1, pp. 163–186, Mar. 2021, doi: 10.3390/iot2010009. [Online].

[5] B. Pradhan, S. Bhattacharyya, and K. Pal, "IoT-Based Applications in Healthcare Devices," Journal of Healthcare Engineering, vol. 2021, pp. 1–18, Mar. 2021, doi: 10.1155/2021/6632599. [Online].

[6] "IoMT Dataset 2024," Canadian Institute for Cybersecurity|UNB, Feb. 2024. [Online]. Available: https://www.unb.ca/cic/datasets/iomt-dataset-2024.html. [Accessed: Mar. 05, 2024]

[7] S. Dadkhah, E. Carlos Pinto Neto, R. Ferreira, R. Chukwuka Molokwu, S. Sadeghi, and A. Ghorbani, "CICIoMT2024: Attack Vectors in Healthcare devices-A Multi-Protocol Dataset for Assessing IoMT Device Security," Feb. 2024, doi: 10.20944/preprints202402.0898.v1. [Online].

[8] F. Hussain and S. G. Abbas and G. A. Shah and I. M. Pires and U. U. Fayyaz and F. Shahzad and N. M. Garcia and E. Zdravevski, "A Framework for Malicious Traffic Detection in IoT Healthcare Environment," Sensors, vol. 21, no. 9, p. 3025, Apr. 2021.[Online].Available: http://dx.doi.org/10.3390/s21093025

[9] W. Ma, L. Ma, K. Li, and J. Guo, "Few-shot IoT attack detection based on SSDSAE and adaptive loss weighted meta residual network," Information Fusion, vol. 98, p. 101853, Oct. 2023, doi: 10.1016/j.inffus.2023.101853. [Online].

[10] N. Moustafa, "A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets," Sustainable Cities and Society, vol. 72, p. 102994, Sep. 2021, doi: 10.1016/j.scs.2021.102994. [Online].

[11] H. K. Kim, "IoT network intrusion dataset," IEEE DataPort,Sep.27,2019.[Online].Available: https://ieee-dataport.org/open-access/iot-network-intrusion-dataset

[12] B. S. and R. Nagapadma, "RT-IoT2022," UC Irvine Machine Learning Repository, 2023. [Online]. Available: https://doi.org/10.24432/C5P338

[13] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, "Towards the Development of Realistic Botnet Dataset in the Internet of Things for Network Forensic Analytics: Bot-IoT Dataset," arXiv.org, Nov. 02, 2018. [Online].

[14] J. Mathews, P. Chatterjee, and S. Banik, "CoAP-DoS: An IoT Network Intrusion Data Set," 2022 6th International Conference on Cryptography, Security and Privacy (CSP), Jan. 2022, doi: 10.1109/csp55486.2022.00025. [Online].

[15] 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,doi: 10.1109/access.2020.3000421. [Online]. Available: http://dx.doi.org/10.1109/access.2020.3000421

[16] M. Ahmed, S. Byreddy, A. Nutakki, L. F. Sikos, and P. Haskell-Dowland, "ECU-IoHT: A dataset for analyzing cyberattacks in Internet of health things," Research Online. [Online].

[17] D. Unal, "BlueTack," IEEE DataPort, Jan. 09, 2021. [Online].Available:https://ieee-dataport.org/documents/bluetack

[18] P. Radoglou-Grammatikis et al., "Modeling, Detecting, and Mitigating Threats Against Industrial Healthcare Systems: A Combined Software Defined Networking and Reinforcement Learning Approach," IEEE Transactions on Industrial Informatics, vol. 18, no. 3, pp. 2041–2052, Mar. 2022, doi: 10.1109/tii.2021.3093905. [Online].

[19] A. I. Jony and A. K. B. Arnob, "A long short-term memory-based approach for detecting cyber-attacks in IoT using CIC-IoT2023 dataset," Journal of edge computing, vol. 3, no. 1, pp. 28-42, 2024.

[20] M. A. Anwer, S. M. Khan, M. Farooq, and Waseemullah, "Attack Detection in IoT using Machine Learning," Engineering, Technology and Applied science research/Engineering, Technology and Applied Science Research, Jun. 12, 2021. [Online].

[21] Churcher et al., "An Experimental Analysis of Attack Classification Using Machine Learning in IoT Networks," Sensors, Jan. 10, 2021. [Online]. Available:https://www.mdpi.com/1424-8220/21/2/446

[22] I. Vaccari, G. Chiola, M. Aiello, M. Mongelli, and E. Cambiaso, "MQTTset, a New Dataset for Machine Learning Techniques on MQTT," Sensors, vol. 20, no. 22, p. 6578, Nov. 2020, doi: 10.3390/s20226578.[Online].

[23] V. Tomer and S. Sharma, "Detecting IoT Attacks Using an Ensemble Machine Learning Model," Future Internet, vol. 14, no. 4, p. 102, Mar. 2022, doi: 10.3390/fi14040102.

[24] D. Unal, S. Bennbaia, and F. O. Catak, "Machine learning for the security of healthcare systems based on Internet of Things and edge computing," Cybersecurity and Cognitive Science, pp. 299–320, 2022, doi: 10.1016/b978-0-323-90570-1.00007-3. [Online].

[25] S. S. Shanto, Z. Ahmed, and A. I. Jony, "Binary vs. Multiclass Sentiment Classification for Bangla E-commerce Product Reviews: A Comparative Analysis of Machine Learning Models", International Journal of Information Engineering and Electronic Business (IJIEEB), vo. 15, no. 6, pp. 48-63, 2023.

[26] S. S. Shanto, Z. Ahmed, and A. I. Jony, "Mining User Opinions: A Balanced Bangla Sentiment Analysis Dataset for E-Commerce", Malaysian Journal of Science and Advanced Technology, vo. 3, no. 4, pp. 272-279, 2023.