VFAST Transactions on Software Engineering
https://vfast.org/journals/index.php/VTSE@ 2024, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 12, Number 2, April-June 2024 *pp: 01-13*

Enhancing IoT Security through Machine Learning-Driven Anomaly Detection

Usama Tahir [1]*, **Muhammad Kamran Abid** [1], **Muhammad fuzail** [1], **Naeem aslam** [1]

[1]NFC Institute of Engineering and Technology, Multan, Pakistan

Keywords: *IOT, Machine Learning, Deep Learning, IOT Anomaly Detection.*

Journal Info:
Submitted:
February 20, 2024
Accepted:
April 26, 2024
Published:
May 16, 2024

Abstract

This is study emphasizes the growing cybersecurity situations arising from the increasing use of Internet of Things (IoT) devices. Paying the main attention to the development of IoT security, the work here deploys the machine learning-based anomaly detection and adaptive defense mechanisms as proactive methods to counteract existing plus future cyber threat sources. The visual serves to expound the rapid development of the Internet of Things, and it also highlights the importance of infrastructures with robust safety features to secure the connected devices. IoT security statement brings out the hidden threat and vulnerabilities of the IoT, in this context advanced security measures are for the rescue. The objectives concentrate on improving security of IoT via machine learning detection of anomalies, and bring introduction of defense mechanisms that are adaptive. We specify the data sources, preprocessing tasks, and Random Forest, Decision Tree, SVM, and Gradient Boosting algorithms selected for anomaly detection in the methodology section. The abnormity negotiation function and the self-adaptive defense procedures are combined in order to strengthen the information technology ecosystems which are capable of dynamic simplification. The results and discussion part hotelates the effectiveness of machine learning models selected, and indicates about accuracy, precision, and recall metrics. To state in the most significant matter, Gradient Boosting brings the greater precision of 89.34%. Table 3 below indicates the various models' effectiveness. It is proven that Gradient Boosting is the most powerful model among all. The discourse unfolds with account of the results, acknowledgment of the limitations, and discussion crucial obstacles encountered in the realization of the research. The conclusion reaffirms the importance of machine learning in IoT security implementation, thus building a robust system that can evolve to fight the ever-emerging cyber-attacks, keeping up with the progressive direction for securing IoT through the connected world.

***Correspondence author email address:** usama007tahir@gmail.com
DOI: 10.21015/vtse.v12i2.1766

1 Introduction

The power of the Internet of Things (IoT) technology enabling IoT devices to communicate and share data unquestionably has had a pronounced impact on how we navigate through our daily lives. Nevertheless, this upsurge of cyber threats has twofold negative outcome – a complicated security environment and the presence of vulnerabilities.

Particularly in these circumstances, therefore, this study aims to solve the acute problem of the enhancement of IoT safety with the help of Machine Learning application. Attack prevention and cybersecurity reinforcement via anomaly detection and adaptive defenses will be among our critical areas of focus in the quest to turn the IoT ecosystem into a strong infrastructure impenetrable of emerging cyber threats.

Machine learning (ML) and deep learning (DL) are becoming indispensable techniques for resisting security risks as the Internet of Things (IoT) grows in popularity.

This paper investigates the inner workings of various machine learning techniques, such as Convolutional Neural Networks (CNNs), Random Forests, and Support Vector Machines (SVMs). In the context of the Internet of Things, anomaly detection refers to the identification of unusual patterns or behaviors in data that depart from the expected norm. In the healthcare industry, anomaly detection is crucial since errors can have serious effects. Precision is essential [1] [2]

The Internet of Things (IoT) is a concept with which the Internet will be expanded to become the medium through which interrelationships between people can occur.This web of connections encompassing traditional computing devices such as mobile phones, and further extending to items like household appliances, industrial equipment, and wearable gadgets has become so complex that it can be believed to be an extended heat transfer matrix. IoT have witnessed an accelerated growth, in tandem with their expanding functionality in diverse sectors such as transport, home automation, agriculture, and healthcare implied a new paradigm of connected and intelligent homes.

As lots of the IoT applications may evolve every day, it is both imperative and urgent to solve the increasing IoT security issues, since the whole IoT system gets more and more large-scaled and diversely connected [3], [4].

The profound nature of security in the Internet of Things (IoT) paradigm cannot be ignored as its surfaces every device to exchange information through the web. With the influx of internet of things devices in a lot of areas of our life and in the critical infrastructures, they become a desirable prey for criminals who penetrate systems and steal data; or for those who are disrupting the operation of the infrastructure [5].

Due to the nature of the IoT that the related information is sensitive and also the fact that the IoT software could be compromised by the threat of intrusion, it is important to put in place reliable measures to protect the system. Security lying at the core of the IoT system has become the most critical factor behind users' confidence level, deepening trust, and finally ensuring the dominance of this technology about everything related.

1.1 Problem Statement

There has been a profound rise in the generation of data directly due to the ongoing IoT devices explosion but also in complexity in the connection of these devices. On the positive side, the multitude of devices and technologies are reducing the barriers for many individuals to engage in this new trend; but the fast growth also leads to a wide range of security challenges identified as a major threat to the secure and reliable functioning of the IoT ecosystems. Only conventional cybersecurity solutions sometimes cope with the fast and changing of cyber risks that emerge with the use of IoT devices.

In the long run, there are many shortcomings, such as the imperfect ways of authentication, the lack of encryption mechanisms, and inadequate compliance to standardized security practices, which the case remains that unauthorized individuals exploit perceived weaknesses to initiate cyberattacks and thus causing disruption. Due to this reason, the security system of IoT devices needs improvement in order to well counter these issues and aid the interconnected

devices to be resilient in the IoT world.

1.2 Challenges and vulnerabilities in IoT security

The Internet of Things (IoT) security space is full of difficulties and perils that can only be dealt with through punctual responses. On the one hand, the broad band of devices created as part of IoT technology, coupled with their different specifications and functionalities, represents quite a tricky task when it comes to developing standardized security measures. Several IoT devices operate based on the limitations of available resources making it hard to implement security protocols that can be characterized as being strong. Also, the right to such smart devices opens the floodgate of admitted data, which at once becomes a vulnerable gateway for hackers, launching worries about data privacy and integrity [5], [6], [7].

The interconnectivity of IoT devices introduces a wide attackable surface, thus it can be targeted by various malicious attacks, e.g., unauthorized access, data pollution and denial of service. Moreover, currently security and standards regulation are not that vigorous and do not have uniform requirements, which negatively affects the whole level of the ecosystem security of the IoT. Resolving these multiple issues in balance should become a priority to stabilize the security ecosystem of IoT and maximize its advantages while minimizing dangers.

1.3 Need for advanced security measures

The rise of IoT, now its sophistication with the changing threat environment, emphasizes the necessity for robust measures of security. The limitations of traditional security protocols are evidenced by the fact that they keep giving way to the weaponry of the new type of cyber threats that focus on intelligent cyber-physical systems. Growing need of these security means is because of the situation where these measures should be proactive to fight against new threat vectors examples are advanced malwares, ransomwares or zero-day vulnerabilities.

On the other side, as the Hence, with the new and innovative IoT applications that come after all of these, the RoI (Return on investment), in the healthcare, smart cities and industrial automation becomes much higher than before. Comprehensive security mechanisms, such as using sophisticated encryption procedures, hunting for unusual behavior and self-adaptability, are vital elements of a strong IoT infrastructure which would ensure the privacy of private data and the integrity of the canals of the nation [7], [8]. The advancement of security protocols also should be a foundation point for building at the IoT network continuing operations.

Machine learning enables enhanced Internet of Things (IoT) security and is also one of the big influences in diminishing the modern threat posing by the connected devices. Machine learning system that leverages its competency of discovering the patterns and abnormalities in the large databases provides a better and pre-emptive defense mechanism against the changing face of the cyber threats. Utilization of these algorithms will allow for automation in terms of generation of responses and adaptation to different sensory inputs which in return initiates real time response and reaction to imminent threats. As this method is not merely conventional rule-based security practices but rather adapts itself on the basis of the situation, it can offer a better protection, which is not only efficient, but also fairly advanced. Incorporating machine learning into the architecture of IoT security is not only effective for IOT detection but also makes it possible for the development of predictive model that allow pre-emptive measures to be taken in advance to prevent security breaches which in turn increases the resilience of the IoT network.

Identifying and defending from anomalies with the help of smart tools like anomaly detection and adaptive defense mechanisms is the surefire formula for thwarting IoT attacks. Anomaly detection is about finding deviations from known patterns in the behavior of any device or system within an IoT environment, enabling to report security breaches or any abnormal activities as soon as they happen. The opposite approach is an anticipatory one that allows for early intervention and threat management before a situation goes out of control making it a very crucial measure. Other Strate-

gies

Against these changes, the adaptive algorithms are to build new rules of the communication security based on "hot" information about threats on the go. This reactivity permits the self-optimization of the IoT systems, which use the countermeasures in a bid to regulate the vulnerabilities in a bid to develop mechanisms that will deal with the emerging threats instantly [8]. Whereas static security measures of IoT could become stagnant, static measures features of anomaly detection and adaptive defense mechanisms work synergistically together, evolving IoT security beyond conventional static protection, offering a formidable and swift first line of defense against ever-changing and sophisticated threats of IoT devices in the interconnected world.

2 Literature Review

Insight into IoT is the security coverage summing up challenges and solutions in the domain. The platform for the security of the Internet of things can be referred to as the multi-faceted approach secured for the inter-connected sensors and the data they generate. The principal elements are authentication, encryption, and entrance control which serve the purpose of preserving confidentiality and integrity of data. Moreover, giving priority to secure device administration procedures along with timely software updates are also among the most effective methods to stay away from facing security risks [8]. Being aware of the security aspects specific to huge multitudes of devices, with limited resources, a holistic knowledge of controlling methods to suit the diverse environment should be obtained. This portion is going to help in understanding the very basic things concerning IoT security and put the reader on a steady footing ahead of a detailed exploration on machine learning powered anomaly detection and adaptive defense mechanisms [9].

In particular, the web of IoT is filled with many types of threats and attacks that may lead to potentially huge security issues for the devices that communicate with each other. Context is something that a hacker might compromise, data breaches, denial of service attacks, man-in-the-middle attacks, eavesdropping and physical tampering, are among the most common vectors of attacks that exploit the vulnerabilities in the Internet of Things. Creating the breach in the appliances, the falsification of data, and the disturbances in the communication are most dangerous, so it is all up to the organization to be equipped with the relevant security systems. Tackling these obstacles is critical in providing the proper support for IoT systems and the suggestions further in the paper will discuss how machine learning-based anomaly detection and adaptive defense mechanisms may help in resolving these issues and ensure the overall security of connected world [8].

At the present state, efforts to improve security within the Internet of Things (IoT) are usually in the form of different existing measures aimed at preventing the multiplicity of possible threats. Security related technologies like cryptographic keys and biometrics are the very strong authentication protocols that widely used in the realm of the Internet of Things (IoT) for the identification of connected devices and users in its ecosystem. Data confidentiality and integrity are protected by encryption techniques which are used to encrypt data at crucial intervals such as while in transit and while at rest. Access control mechanisms erect barriers to entry of unauthorized access to IoT devices on the other hand, the managed device protocols make device configuration reliable and properly monitor them [8]. Routinely, software basic updates and patch management are being designed to tackle zero-day vulnerabilities and further strengthen the ability of Internet of Things systems to defend from potential cyber-attacks [10].

However, these mitigation controls are still facing variegated nature of IoT devices, and the changing technology landscape making dynamic threat landscape highlighting the need for emerging security alternatives. The upcoming part will plunge into the capacities of anomaly detection, as well as adaptive defensive mechanisms in order to develop and fortify already existing security structure within the sphere of IoT [9]. ML shows more and more evidences that it is a useful technology to strengthen security of the

information realm of the IoT.

Hereunder, the paper presents the crucial ML function in strengthening IoT security and shows its significance in this field of studies with the help of illustrative examples. Through ML algorithms that can scan and identify various patterns from the datasets of millions of sizes, proactive and adaptive mechanisms can be provided to prevent cyber threats. Applying both supervised and unsupervised learning techniques will assist the ML models to identify anomalies, breaches or deviations rapidly in real-time which would greatly help them to detect emerging and sophisticated attacks [11]. Besides that, the ML is able to generate predictive models which help as a means for detecting and preventing security threats before they occur. This part will investigate the multiple faces of ML in IoT security that supersede the aforesaid challenges entailed in cyber threats as IoT world gets more interconnected and complicated every day [6].

The Machine Learning (ML) technology plays a central role in cybersecurity, having an array of use cases that span across threat detection and containment systems. Anomaly detection by ML proves effective in a way that it identifies an unusual and diverse character thus, improves behavioral analysis and recognizes the potential of the security breach. Continues to learn and gains new skills in detecting malware, preventing invasion and phishing identification through pattern and characteristics recognition [11]. By vulnerability management in ML, we ensure timely identification of the systems' weaknesses and the addition of the user and entity behavior detection gives us more reliable answers on insider threats. As well, ML enhancing SIEM technologies, these will conduct the processing of massive data comes to be threat intelligence accurately. This section will explain how MLs abilities can be rendered to let down a security barrier of IoT, bearing in mind anomaly detection and adaptative mechanism [10].

Existing research on embedding ML into Internet of Things (IoT) security is rapidly moving at picking the challenges of the continuing machine to machine ecosystems. Earlier investigations have been devoted to ML approaches in detections of anomalies, highlighting their strengths in revealing unusual structures and security dangers. Researchers too have discovered the usage of artificial intelligence as part of predicting and preventing cyberattacks on DTs [12]. This espouses the versatility of artificial intelligence in dynamic and intricate threat spaces. Consequently, endeavors towards the utilization of ML in Internet of Things devices on low-resource devices are the major elements of real-world deployment. In addition, research efforts have been concentrating on the creation of hybrid models that use ML-driven methods in order to compensate for cybersecurity deficiencies in traditional models and reinforce overall information security [13]. This section will analyze the results obtained so far enriching them with the view points of the experts in the field of cyber security and thus is a call for novel approach of integration which can use ML for anomaly detection and adaptive cyber defense mechanisms. Therefore, this section also contributes to the on-going discussion of IoT security robustness enhancement [14].

Detection of departure from typical IoT does mainly refer to things like identifying and handling of abnormal patterns or behaviors among the millions of gadgets that are all connected among them. The whole section will be devoted to the topic, identifying the importance of the anomaly detection, how it is applied and what the challenges are for this activity besides the pre-existent methods in the IOT environment. In the IoT domain, Abnormal detection (AD) achieved through the application Machine Learning (ML) algorithms detects deviation from assumed norms in a proactive manner. Prior studies covered a span of ML techniques, amongst them supervised and unsupervised learning [13], [15]. These methods proved to be efficient in detecting anomalies in IoT data that are conditioned in terms of variety and variability. Issues like the ability to scale, functional versatility for distinct IoT environments, and need for tracking of every happening in real-time, will be examined, and this preliminary discussion will form the basis for the next chapter which will highlight the ways in which these challenges can be resolved by means of adaptive protection mechanisms [10].

The potential of anomaly detection in regards to the Internet of Things (IoT) is actually crucial for the current cybersecurity system because the situation varies depending on the context of proactive security. Identifying anomalies that occur in the complex and interchained network of IoT devices is an important task used to discover situations that differ from the ordinary ones. This is where machine learning (ML) goes into action [16], [17]. The technology is used to determine the anomalies. When something is out of place or a security threat is at hand, the technology spots the issue in time for the timely detection and settlement. This becomes more crucial in the environment provided by the IoT which is upgraded and variable, where traditional security interests may be not handy. The value is in taking actions to discover earlier the danger, unauthorized access, or weird data patterns, which is the core activity that strengthens the security condition of IoT ecosystems. With IoT that continues to interact with a variety of industries, it begins to stand out as a critical solution to protecting and securely delivering data as well as connected objects [18].

A range of different anomaly detection methods will be used in the Internet of Things (IoT) to detect and cancel possible security threats. While these methods include the statistical modeling or correlation study of event occurrences, time series analysis or signature-based detection, modern methods are now additionally identified as machine learning algorithms like clustering, neural networks and behavioral analysis. Time series analysis is a cornerstone of IoT systems that tackles temporal nature of an IoT environment. Meanwhile, signature-based detection approach is data-driven and uses known attack patterns. Behavioral analysis is about studying and interpreting device operations and user behaviors, and flow-based analysis is related to network flows that contain all the communication patterns in the network. Moreover, the hardware- oriented techniques are specifically designed for the real-time monitoring functionality within the IoT devices [12]. Every solution is valuable and limited. Therefore, their effectiveness holds very weight and it is determined by the unique features of the IoT ecosystem. Thus, this part not only will break down principles of the outlined methods, but also focus their significance and development opportunities for anomaly detection for IoT security [14].

3 Proposed Methodology

Diving into the techniques of ML, we will take a look at Random Forest, which builds trees and combines their predictions using majority voting. Decision Tree, which is based on the If-Then rule, chooses the best split at a node, is used. Support Vector Machine (SVM), which determines the hyperplane that separates and classifies the data. And the Gradient boosting, which builds better models by improving on For instance, employing an algorithm to power an all-encompassing anomaly detection system in the internet of things ecosystem is a case in point. Random Forest can deal with complex datasets, yet it eliminates overfitting by combining different learning algorithms in parallel. The Decision Tree, widely known for its interpretability helps us understand the rationale supporting the categorization of anomalies. SVM enjoys remarkable classification properties that is, indeed, a powerful tool in differentiating patterns in the data of a very high level. Gradient Boosting performs an incremental correction, improving predictive accuracy and at the same time resolving complex interrelations hidden in the Data set of IoT.

3.1 Selection criteria for ML algorithms

When it comes to the implementation of ML algorithms in an IoT System with regard to the human factor, it is the issue of algorithm selection that is perhaps the most critical and hence, merits due attention. Several important criterions such as algorithm suitability, effectiveness of the approach, applicability, and efficiency help in judicial selection of algorithm for the research. First of all, the system must be able to properly account for the very high dimension and heterogeneity Iot multidimensional data. One among these algorithms, Random Forest, Decision Tree, Support Vector Machine (SVM), and Gradient Boosting (GB), for the reason that they are able to adapt themselves to work in the complex conditions of IoT, are rightly used for these IoT cases.

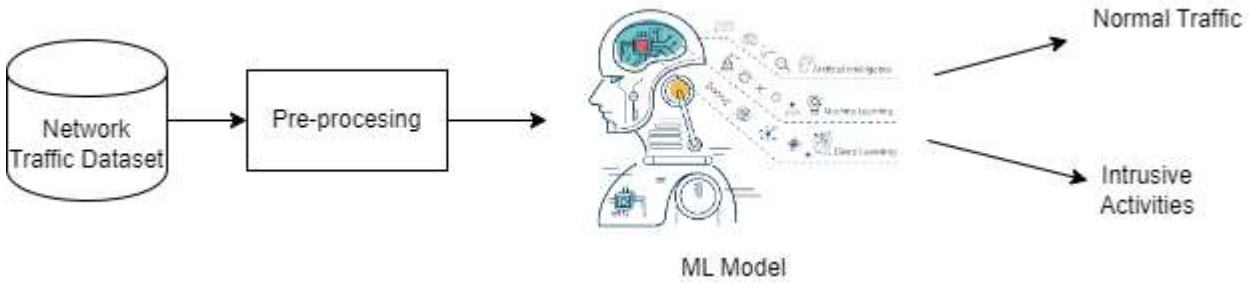

Figure 1. Anomaly detection using ML.

Another factor that needs consideration is scalability and efficiency. The proposed algorithms should be designed in such a way that they can transparently process and analyze the huge quantity of data generated by the multitude of IoT devices. Also, among other factors, the algorithms' interpretability is pretty important. With decision trees, the cause and effect of the anomaly is more apparent and easily grasped. The robustness of SVM in classification and the iterative improvement of Gradient Boosting Construction also come in handy when choosing ML algorithms. It ensures that the ML algorithms chosen fit the specific needs as well as for mutually enhancing incident security using sophisticated anomaly detection. This part will extensively demonstrate the reasons behind the choice of these algorithms and how each principle plays an important role in the fulfilment of the research by all means thereof.

3.2 Anomaly Detection and Adaptive Defense Mechanisms

The primary attention will be on conceiving and implementing anew abnormal detection mechanism that will exploit the advantages of these algorithms in order to detect irregular patterns and eventual security infringements in the real time. Adaptive defense mechanism will be addressed or alleviated in the context of how the system resiliently auto learning machine algorithms and intelligently adjust security measures to detect suspicious activities. The adaptive approach of this ensures that the resilience of IoT ecosystems grow hitherto more effective in counteracting any newly emerging cyber threats. This section will explore in detail the step-by-step process of anomaly detection, with an emphasis on how different ML algorithms fulfill respective tasks together to form an effective collaborative defense mechanism, and how those adaptive measures perfect the security of the IoT devices.

4 Experimental Results

This section describes the evaluation metrics and compare our proposed approach results with the state-of-the-art.

4.1 Random Forest

In Table 1, Random forest classifier report (10 iterations) – the machine learning algorithm used for classification tasks. The horizontal axes are those tube entries representing the traffic types which the model was trained on, and the vertical ones are the student marks of each class meeting.

Precision

(Positive Predictive Value) is one of the parameters of accuracy level of the model in identifying the certain class. Another example for this is the precision score which can be 1.00 for "Attack class" which implies that all the examples which was identified as 'Attack' by the model were actually attacks.

Sensitivity or recall

TPR (True Positive Rate) is the measure of how appropriate the model is at recognizing all the relevant classes among the classes that it is supposed to find. A recall of 1.00 for "Attack" class indicates that all considering the attack cases in our data set have been

Table 1. Classification Report Table for Random Forest

	Precision%	Recall%	F1-score%	Support
Attack	1.00	1.00	1.00	756
Benign	0.83	0.82	0.83	5253
C&C	0.34	0.36	0.35	1111
DDoS	0.60	0.67	0.63	9
Okiru	0.95	0.65	0.77	31
PartOfAHorizontalPortScan	0.94	0.94	0.94	2441
Accuracy			0.81	9601
Macro Avg	0.78	0.74	0.75	9601
Weighted Avg	0.82	0.81	0.81	9601

Table 2. Classification Report table for Support Vector Machine

	Precision%	Recall%	F1-score%	Support
Attack	0.00	0.00	0.00	756
Benign	0.55	1.00	0.71	5253
C&C	0.00	0.00	0.00	1111
DDoS	0.57	0.44	0.50	9
Okiru	0.00	0.00	0.00	31
PartOfAHorizontalPortScan	0.00	0.00	0.00	2441
Accuracy			0.55	9601
Macro Avg	0.19	0.24	0.20	9601
Weighted Avg	0.30	0.55	0.39	9601

recognized by the model. F1-Score is the harmonic mean of precision and recall and is applied to appraise the general sentiment of a model posing the right class. Support is the number of the samples or instances of each class.

Attack" and "PartOfAHorizontalPortScan" traffic

The high precision (1.00) and recall (1.00) guard the model from the miss-identification for the above mentioned categories. "CC" and "DDoS" traffic: Nevertheless, the model shows bias against these types of classes with precision low (0.34 and 0.60 respectively) which indicates that very many examples of this traffic class was automatically invalidated as CC or DDOS traffic.

4.2 Overall Accuracy

The accuracy of our model is of 81% which means that the traffic samples have been correctly classified in a ratio of 81%. While it is critical to take into account the class imbalance in the accuracy assessment, it is also vital to analyze unfair bias. This classifier in particular deals with the issue where many more samples are labeled as "Benign" compared to the other types. Thus, a model may acquire a shot at yielding such high accuracy that it would classify everything as benign - even if worse at seeing other kinds of traffic.

Weighted Average metrics usually re-scales the class so the properly size of the class imbalance and appear to give model performance a true picture. The weighted macro average precision is 0.82, the weighted macro average recall is 0.81, and the weighted macro average F1-score is 0.81.

Table 3. Classification Report for Decision Tree

	Precision%	Recall%	F1-score%	Support
Attack	1.00	1.00	1.00	756
Benign	0.83	0.84	0.83	5253
C&C	0.35	0.35	0.35	1111
DDoS	0.60	0.67	0.63	9
Okiru	0.91	0.65	0.75	31
PartOfAHorizontalPortScan	0.94	0.94	0.94	2441
Accuracy			0.82	9601
Macro Avg	0.66	0.63	0.64	9601
Weighted Avg	0.82	0.82	0.82	9601

Table 4. Classification Report for Gradient Boosting

	Precision%	Recall%	F1-score%	Support
Attack	1.00	0.99	0.99	756
Benign	0.85	0.98	0.91	5253
C&C	0.99	0.31	0.47	1111
DDoS	0.50	0.44	0.47	9
Okiru	0.75	0.10	0.17	31
PartOfAHorizontalPortScan	0.96	0.95	0.96	2441
Accuracy			0.89	9601
Macro Avg	0.84	0.63	0.66	9601
Weighted Avg	0.91	0.89	0.87	9601

Table 2 presents the mark of a ML model in classifying the network traffic into 4 categories. This model scored 100% accuracy in classifying "Attack" and "PartOfAHorizontalPortScan" traffic but had only a 34% precision in "CC" and a 60% precision on arresting "DDoS". The model is 81% in general, but it favors the normal traffic (more than half of the topics dealt with the benign traffic). Considering this imbalance, weighted metrics provide a better picture: With precision score of 82%, recall score of 81% our general performance is satisfying but being specific about traffic types lowers our results to a great extent.

Support Vector Machine Accuracy: 0.5472

High SVM error performances are revealed in the table 2 alongside the reduced accuracies compared with the last model. It not even once distinguishes imagined and real traffic labels as 'Attack', 'CC', 'Okiru', and 'PartOfAHorizontalPortScan' (0% precision). Though the AUC is perfectly rounded, this might be the case because of the class imbalance (many samples fall in the positive side). The overall accuracy is a false figure, and weighted metrics, on the other hand, confirm the very underwhelming performance (precision 30%, recall 55%). Such underpinning needs more research or improvement.

Decision Tree Accuracy: 0.8165

In the Table 3, the decision tree also produces results similar to those of the random forest (as shown in the preceding example). It operationalizes the "Attack" and "PartOfAHorizontalPortScan" traffic very well (with a precision of 100%) and has a lower performance rate for the "CC" and "DDoS" attacks around 35%. This can be seen by the large overall accuracy (82%) but again due to the class imbalance the weighted metrics (82 %

precision and recall) provide fairer representation of the performance of this system, hence it can be good choice for network traffic classification.

Gradient Boosting Accuracy: 0.8934

Gradient Boosting mini model gives great results and gains quite a lot of advantage in comparison with other previous versions as given in Table 4. It beats in this detection of categories "Attack," "Benign," and "PartOfAHorizontalPortScan" (precision above 95%, and recall above 90%). In terms of "CC" and "DDoS" difficulties, the model fails to keep up (lower precision), but it is far more effective than the models with 99% and 50% precision, respectively. Considering the degree of imbalance in our classes, we achieve good results with the weighted metrics (91% precision, 89% recall) that capture the overall performance of our model above average across nearly all traffic types.

4.3 Performance Comparison

Table 5. Models comparison

Models	Accuracy
Random Forest	0.812415
Support Vector Machine	0.547235
Decision Tree	0.816477
Gradient Boosting	0.893449

Table 5 for five table summarizes the performance of four machine learning models to classify network traffic. Among the algorithms considered, Gradient Boosting managed to attain the most reproducible results (89.34%). With accuracy around 81-82%, Random forests and Decision Trees are the closest ones followed by (Naive Bayes (76%) and KNN-1 respectively). SVM, Support Vector Machine, markedly fails to deliver with 54.72% accuracy To sum up, GBM has most of all needed and delivers pretty good results considering the challenges the model faces like "CC" and "DDoS".

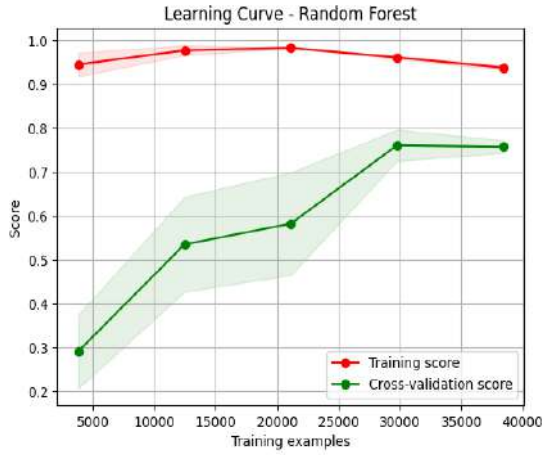

Figure 2. Learning Curve Random Forest.

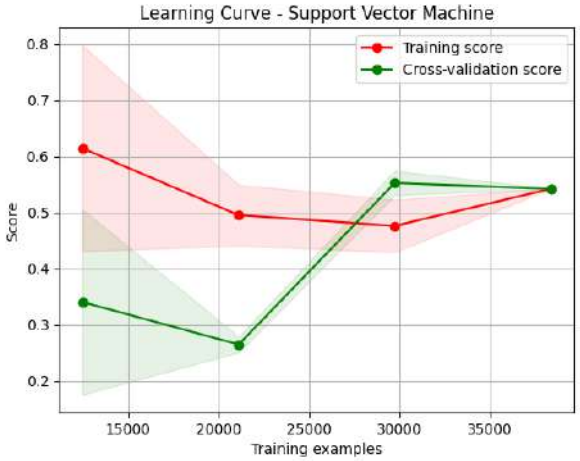

Figure 3. Learning Curve Support Vector Machine.

Figure 2 illustrate a learning curve of the random forest model. The learning curve is a representation of how a model's performance improves the more it is trained and as a result of more data. The x axis corresponds to the number of training examples, and the y axis represents the highlighted metric (for instance accuracy). The blue and the red line, in the graph, depict the training score and the cross-validation score. Score training shows how good the model is at the data it was trained on, while score cross-validation votes for the model's performance if the data has not been seen earlier. Optimal situation constitutes that the two lines lay closely to each other. This graph illustrates the dynamics of the training score growth with the passage of time as the model is trained by more data.

Figure 3 exemplifies the exploration curve of an SVM model classifying a dataset, where its performance increases at first and later stagnated. The

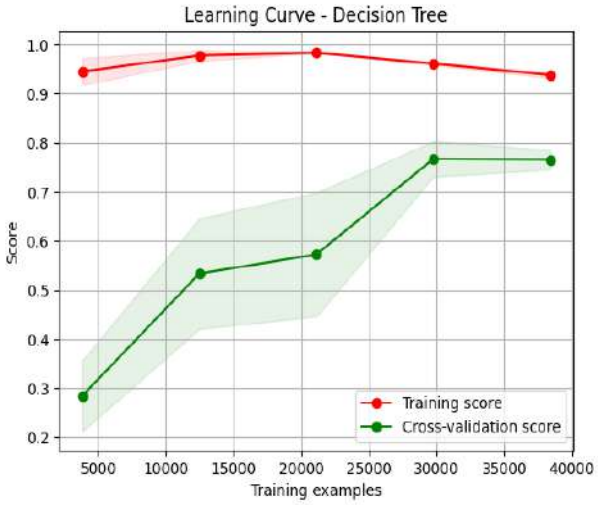

Figure 4. Learning Curve Decision Tree.

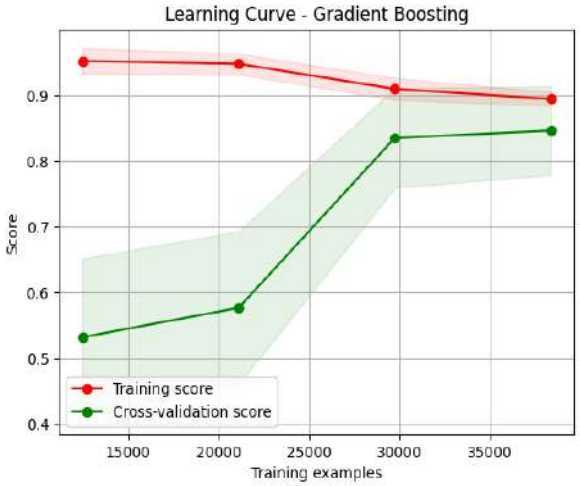

Figure 5. Learning Curve Gradient Boosting.

x-axis is the number of training events, and y-axis represents the model's score my likely be accuracy. A: The curves S1 and S2 indicate the training score and the cross-validation score. During the training phase, the training score indicates how accurately the model can do on the data it was trained by, on the other hand, the cross-validation score shows how well the model can perform on unseen data.

In Figure 4 there is a chance that as there will be unseen data the model will perform less.

In figure 5,There are two plots on the graph (the training score and the cross-validation score). And where the training score points to how well the model operates on the provided training data, the cross-validation score shows how well model operates on unseen dataset. both the train score and the cross-validation score are successfully measured to be higher as the model is trained with more data. The training score rises faster than the cross-validation rate, but do not panic. This implies that the model might be somewhat overfitting the data, or, in other words, not generalizing it for all unseen data.

5 Conclusions and Future Work

In this section, we conclude our whole work and give future directions for it.

5.1 Conclusions

The present work resewed into the efficacy of several machine learning methods for anomaly detection and adaptive security mechanisms in Internet of Things (IoT) networks. We focused on prevalent security issues in IoT devices and the fact that the traditional security solutions are not sufficient anymore. The research commissioned to compare the execution of Random Forest, Support Vector Machine (SVM), Decision Tree, and Gradient boosting models for network traffic classification was done.

Surprisingly, the Gradient Boosting approach gained a higher absolute accuracy (89.34%) in discrimination of network traffic classes, such as attacks, normal traffic, and more specific malicious traffic like Command-and-Control communication or DDoS attacks. Besides, the random forest and the decision tree models also showed good success with accuracy around 81% to 82%. On the one hand, SVM did worst and shows up with the lowest value of 54.72% accuracy, which was not so good.

Model behavior studies, which involve the learning curve analysis, added more colors to my understanding. However, even if the models had some degree of overfitting, Gradient Boosting performed superior in trade-off between data-driven training and capability for unseen data. On the decision tree chart, the lifeline exhibited a fast sharp response at the bottom fol-

lowed by a leveling that could be brought by overfitting if trained for too long.

The learning curve of SVM model indicated on the training process that it was producing exaggerated overfitting, this is the reason the model performs low. It is only evidence that Gradient Boosting can be effectively applied toward anomaly detection in that IoT is such powerful tool. The feature that possesses the competence to intensely learn complicated shapes and also to adapt to moving threats, makes it a precious component of the security systems in IoT ecosystems.

Nevertheless, those techniques still need to be explored to prevent the mentioned undesirable effect and implementation of different algorithms for creating even more robust and better adapted systems by using the advantages of diverse machine learning tools. It is also worth noting that designing physical-world implementations, comprising computational efficiency and scalability, is also a major area of necessity to make IoT TSN applicable in large size networks.

5.2 Future Work

Find methods for solving the problem with over-fitting in Gradient Boosting and other models. Inquiry ensemble learning methods that have the strong sides of multiple machine learning algorithms together and use of them to overcome their weaknesses.

The assessment is needed of how good the computational efficiency and scalability of the developed models are in terms of their ability to run in real-world large IoT networks. Perform research about adaptive defense tools which will be capable to automatically change security parameters when managed a threat.

Authors contributions

All authors contributed equally to accomplish this study. In addition, all authors read and approved the final manuscript.

Conflict of interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Data availability statement

The datasets will be available upon request to the corresponding author.

Code availability

The code will be available upon request to the corresponding author.

Compliance with Ethical Standards

It is declare that all authors don't have any conflict of interest. It is also declare that this article does not contain any studies with human participants or animals performed by any of the authors.

Furthermore, informed consent was obtained from all individual participants included in the study.

References

[1] V. V. Raje, S. Goel, S. V. Patil, M. D. Kokate, D. A. Mane, and S. Lavate, "Realtime anomaly detection in healthcare iot: A machine learning-driven security framework.," *Journal of Electrical Systems*, vol. 19, no. 3, 2023.

[2] S. Akbar, K. T. Ahmad, M. K. Abid, and N. Aslam, "Wheat disease detection for yield management using iot and deep learning techniques," *VFAST Transactions on Software Engineering*, vol. 10, no. 3, pp. 80–89, 2022.

[3] H. Bangui and B. Buhnova, "Recent advances in machine-learning driven intrusion detection in transportation: Survey," *Procedia Computer Science*, vol. 184, pp. 877–886, 2021.

[4] M. Ramzan, Z. U. R. Zia, M. K. Abid, N. Aslam, and M. Fuzail, "A review study on smart homes present challenges concerning awareness of security mechanism for internet of things (iot)," *Journal of Computing & Biomedical Informatics*, 2024.

[5] M. K. Abid, Z. U. R. Zia, and S. Farid, "Security and privacy for future healthcare iot," *Journal of Computing & Biomedical Informatics*, vol. 4, no. 01, pp. 132–140, 2022.

[6] M. Nankya, R. Chataut, and R. Akl, "Securing industrial control systems: Components, cyber threats, and machine learning-driven defense strategies," *Sensors*, vol. 23, no. 21, p. 8840, 2023.

[7] M. K. Abid, M. Qadir, S. Farid, and M. Alam, "Iot environment security and privacy for smart homes," *Journal of*

Information Communication Technologies and Robotic Applications, vol. 13, no. 1, pp. 15–22, 2022.

[8] D. Javeed, T. Gao, M. T. Khan, and I. Ahmad, "A hybrid deep learning-driven sdn enabled mechanism for secure communication in internet of things (iot)," *Sensors*, vol. 21, no. 14, p. 4884, 2021.

[9] M. A. Alsoufi, S. Razak, M. M. Siraj, I. Nafea, F. A. Ghaleb, F. Saeed, and M. Nasser, "Anomaly-based intrusion detection systems in iot using deep learning: A systematic literature review," *Applied sciences*, vol. 11, no. 18, p. 8383, 2021.

[10] S. Bharati and P. Podder, "Machine and deep learning for iot security and privacy: applications, challenges, and future directions," *Security and communication networks*, vol. 2022, pp. 1–41, 2022.

[11] E. Gyamfi and A. Jurcut, "Intrusion detection in internet of things systems: a review on design approaches leveraging multi-access edge computing, machine learning, and datasets," *Sensors*, vol. 22, no. 10, p. 3744, 2022.

[12] M. Aslam, D. Ye, A. Tariq, M. Asad, M. Hanif, D. Ndzi, S. A. Chelloug, M. A. Elaziz, M. A. Al-Qaness, and S. F. Jilani, "Adaptive machine learning based distributed denial-of-services attacks detection and mitigation system for sdn-enabled iot," *Sensors*, vol. 22, no. 7, p. 2697, 2022.

[13] S. K. Devineni, S. Kathiriya, and A. Shende, "Machine learning-powered anomaly detection: Enhancing data security and integrity," *Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-198. DOI: doi. org/10.47363/JAICC/2023 (2)*, vol. 184, pp. 2–9, 2023.

[14] I. Ullah, A. Ullah, and M. Sajjad, "Towards a hybrid deep learning model for anomalous activities detection in internet of things networks," *IoT*, vol. 2, no. 3, pp. 428–448, 2021.

[15] Y. N. Kunang, S. Nurmaini, D. Stiawan, and B. Y. Suprapto, "Attack classification of an intrusion detection system using deep learning and hyperparameter optimization," *Journal of Information Security and Applications*, vol. 58, p. 102804, 2021.

[16] B. Hussain, Q. Du, B. Sun, and Z. Han, "Deep learning-based ddos-attack detection for cyber–physical system over 5g network," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 2, pp. 860–870, 2020.

[17] A. Haider, M. Adnan Khan, A. Rehman, M. Rahman, and H. Seok Kim, "A real-time sequential deep extreme learning machine cybersecurity intrusion detection system," *Computers, Materials & Continua*, vol. 66, no. 2, pp. 1785–1798, 2021.

[18] M. Catillo, M. Rak, and U. Villano, "2l-zed-ids: A two-level anomaly detector for multiple attack classes," in *Web, Artificial Intelligence and Network Applications: Proceedings of the Workshops of the 34th International Conference on Advanced Information Networking and Applications (WAINA-2020)*, pp. 687–696, Springer, 2020.