ACCURATE AND EFFICIENT SECURITY AUTHENTICATION OF IOT DEVICES USING MACHINE LEARNING ALGORITHMS

IlhamAlghamdi[1] and Mohammad Alzahrani[2]

[1,2]Department of Computer Science, Faculty of Computing & Information
Al-Baha University, Al-Baha, Saudi Arabia.

ABSTRACT

The rapid proliferation of Internet of Things (IoT) devices has led to an increase in botnet attacks targeting these devices. A botnet attack is a cyber-attack in which a network of compromised devices, referred to as "bots" or "zombies," is utilized to execute a synchronized attack. These attacks can result in substantial harm to both the devices and the network to which they are connected. This study investigates the deployment of security authentication protocols to verify the identity of IoT devices prior to network connection. The study also evaluates the classification accuracy of four distinct supervised machine learning algorithms: Random Forest (RF), Naïve Bayes (NB), DecisionTree (DT), and eXtreme Gradient Boosting (XGBoost). It was foundXGBoost was the best performing classifier among the various machine learning algorithms tested, in terms of detecting botnet attacks in IoT networks using the Bot-IoT dataset.

KEYWORDS

Cyber Security, Authentication, Internet Of Things, Supervised Machine Learning, Botnet Attack.

1. INTRODUCTION

The Internet of Things (IoT) enables the connection between the physical world and computer networks. Privacy and security measures are crucial for upcoming IoT systems, especially in applications like infrastructure management and environmental monitoring [1-3] The Industry Internet of Things, or IIoT, has emerged as the most rapidly developing breakthrough technology in recent years. It can digitise and combine a vast array of industries, opening up enormous economic opportunities and contributing to global Gross Domestic Product, or GDP, increase. IIoT applications include supply chain management, interconnected cars, smart towns, electric power systems, logistics, and transportation [4].

Even while the IIoT has tremendous opportunities for the growth of several industrial applications, it needs stronger security requirements because it is susceptible to assaults. Hackers from all over the world are interested in the vast amount of information generated by the numerous gadgets such as sensors, controllers, and actuators employed in the IIoT network to make wise business decisions. Conversely, IIoT-enabled sensors and devices are considered constrained by resources, having limited capacity for memory, power, and communications.

Thus, to connect sensor and cloud servers, edge devices such laptops, desktops, routers, micro servers, handheld devices, and mobile phones are used. These gadgets gather sensor data and transfer it to servers in the area after doing any necessary pre-processing. However, as the industry's IoT edge devices count has increased quickly, a number of security and privacy concerns have

David C. Wyld et al. (Eds): NBIoT, MLCL, NMCO, ARIN, CSITA, ISPR, NATAP-2024
pp. 79-90, 2024. CS & IT - CSCP 2024 DOI: 10.5121/csit.2024.140507

emerged, presenting a serious risk to the security and reliability of IIoT [5]. These edge devices could provide advantage to intruders. This could lead to financial and damage to reputation, as well as inefficiency in operations. Making sure the IIoT is secure is one of the main problems in today's industrial settings. It includes communications and personal security protection, blocking unwanted access, and virus defence. The security, protection, and unwavering quality of the IIoT might be expanded by carrying out careful and exceptional security conventions. One of the most crucial security elements for IoT/IIoT security is an Intrusion Detection Device (IDS), which keeps an eye on networks regarding unauthorised activity or policy violations.

The foundation of IDS may be the identification of abnormalities or signatures. Signature-based detection techniques are highly effective in identifying attacks against previously identified patterns when combined with current criteria. On the other hand, assaults without clear patterns or unknown attacks are identified using anomaly-based identification. Recent studies have shown that machine learning techniques can prevent many security vulnerabilities and enhance the effectiveness of anomaly-based techniques for detection [6].When it comes to high-security authentications, cloud integrations are formed based on many factors such as data volume, size, and manner of transmission, which affect a great deal of data transmitted via wireless applications. To safeguard data, most IoT programmes, however, rely on specific unidentified components, which leaves some room for error in data delivery and receipt.

An important and quickly expanding area, the Internet of Things (IoT) has the potential to revolutionize our daily lives and the way we do business. The potential of IoT devices to improve efficiency, productivity, and quality of life in various applications, such as smart homes and cities, healthcare, manufacturing, and transportation, is driving this exponential growth. New security threats have emerged, though, due to the exponential growth of the Internet of Things. For the benefit of people, the Internet of Things (IoT) connects various systems and devices online so that they can share data. Theft and infringement of personal information is one way in which data exchange between devices can impact people's privacy. To protect user data, unlock the full potential of the Internet of Things (IoT), and mitigate threats, strong security measures are urgently needed to address the security issues related to the IoT. Cybercriminals have an enormous attack surface created by the vast number of interconnected devices. The proposed technique is introduced with the goal of enhancing the confidentiality and safety of data during both the sending and receiving phases, where different parametric assessments for data determinations are carried out. IoT processing techniques become more difficult when more black-box design functionalities are developed using a given set of data. Customers can handle various data difficulties [7], including denial-of-service assaults, cyberattacks, etc., by switching to an external source where the expenditure for monitoring is higher.

The initial lines of defence in Internet of Things environments are Authentication and Authorise (AA), which prohibit actions and operations. By imposing restrictions on users, AA seeks to prevent breaches that could expose vital resources to adversaries they do not want to face. Authentication, along with protection mechanisms against outside intrusions like man-in-the-middle and eavesdropping attacks, are usually tailored to address various risks at specific network conditions. Malicious activity, however, is erratic and cannot be prevented ahead of every attack [8].

1.1. Challenges in Real-World IoT Environments

Implementing machine learning to detect cyber threats in the IoT environment presents numerous substantial challenges.

Data Complexity: The primary challenge stems from the characteristics of data produced by IoT devices. The data is frequently vast, diverse, and rapidly changing, presenting a significant obstacle

to conventional machine learning methods. Hence, there is an urgent requirement for effective and scalable algorithms that can manage this level of complexity.

Device Heterogeneity : In IoT refers to the variety of devices with different architectures, protocols, and operating systems, which presents a significant challenge. The diversity among devices complicates the creation of universally applicable solutions, as each device may need a customized approach.

Dynamic Nature : The dynamic nature of cyber threats requires constant learning and adaptation in threat detection methods. These systems often need manual intervention due to their dynamic nature, which can hinder their efficiency[9] .

1.2. Define IoT

IoT devices are systems and appliances that can send and receive data with other systems and devices over the Internet or other forms of communication. They have software, sensors, processing power, and other technologies built in. Electronics, communication, and computer science are all parts of the Internet of things. The phrase "internet of things" is seen as misleading because devices only need to be able to connect to a network and be able to talk to each other, not the whole internet.
The "Internet of Things" (IoT) is a network of real-world objects, like cars, appliances, and other machines, that are equipped with software, sensors, and a network connection so that they can send and receive data.

1.3. Applications of IoT

A network of gadgets that send data into a platform that allows for automation control and communication is known as the Internet of Things (IoT). It links digital interfaces to tangible devices. Here is a list of the top ten IoT applications.

1.1.1 Improving business solutions

Large organizations utilize specialised IT workers who develop, manage, and keep an eye on their technological infrastructure.

1.1.2 Implementing smarter home automation

A smart reside is the most obvious use of the Internet of Things. Sensors are used in a smart home's resource management, lighting, and security systems. A smart house is a smaller and self-sufficient form of a smart city.

1.1.3 Agricultural innovation

The Internet of Things has enormous prospective benefits for the agriculture business. According to estimates, there will be around 10 billion people on the planet by 2050. Governments have given the growth of agricultural systems top priority as a result. Due to this and climate change, farmers are integrating technology into their methods of agriculture.

1.1.4 Creating smarter urban areas

An urban area that employs wireless or cellular technology and sensors installed in common locations like antennas and lamp posts is known as a "smart city."

1.1.5 Managing the supply chain better

The procedure of Supply Chain Management (SCM) is designed to optimise the movement ofproducts and services from the acquisition of raw materials to the end users. Scheduled maintenance, vendor connections, fleet management, and inventory management were all involved.
And there are some applications such as Reimagining medical services, Developing up intelligent grids, Changing the game for wearable's, Connected factories integrated etc [10] .

2. OBJECTIVES OF THE STUDY

- Used effectiveness of machine learning methods on Bot-IoT dataset to evaluate four machine learning classifiers: RF, NB, DT, and XGBoost
- To protect IoT device communications and stop unwanted access, include authentication and encryption methods.

3. LITERATURE SURVEY

Over the past 10 years, the Internet of Things (IoT) has given the Internet a new dimension; yet, security remains a major problem in IoT, especially with regard to assaults on authentication. Most research projects take into account external assaults that come from networks outside of the Internet of Things. At the beginning of a session, users are authenticated using their authentication mechanisms. However, because they are more accessible than an outside attacker, a device or user outside the network's boundaries can pose a greater threat[11].

In order to precisely anticipate diamond prices, this study undertook a thorough review of several models for supervised machine learning, repressors, and classifiers. Because of the non-linear correlations between important qualities including carat, cut, understanding, table, and depth, valuing diamonds is a difficult undertaking. The goal of the investigation was to create an accurate forecasting model by applying both classification and regression techniques. Overall, the study shows that the Random Forest (RF) performs better than the other systems in terms of precision and ability to predict, as shown by its perfect classification performance, lowest RMSE, and highest R2 score[12].

This study compares the performance of two approaches for projecting Bangladesh's annual rice production (1961–2020): ARIMA (Autoregressive Integrated Moving Average) and the extreme Gradient Booster (XGBoost). The data prompted the selection of a significant ARIMA (0, 1, and 1) framework with drift based on the lowest Corrected Akaike Informative Criteria (AICc) values. The value of the drift parameter indicates a favourable trend upward in rice output. It was discovered that the ARIMA (0, 1, and 1) hypothesis with drift was significant. However, the XGBoost model for data from time series was created by regularly adjusting the tuning parameters to achieve the best results[13].

One of the most popular technologies of the modern day is the Internet of Things (IoT), which has a big impact on our lives in many different ways, namely social, commercial, and monetary ones. IoT innovations, both current and future, have huge potential for upgrading the nature of human life generally speaking through mechanization, efficiency, and customer solace across an expansive assortment of use areas, from instruction to savvy urban communities. Notwithstanding, in the IoT setting, shrewd applications are vigorously affected by gambles and cyberattacks. Given the

ongoing security issues and the complex multiplication of many sorts of assaults and dangers, the ordinary strategies for IoT security are lacking[14].
An Internet of Things (IoT)-assisted Wireless Sensor Network (WSN) is a cooperative system in which WSN systems and IoT networks cooperate to share, collect, and manage data. Improving automation and data analysis to enable better decision-making is the main goal of this partnership. Protective measures must be put in place to ensure the safety and dependability of the linked WSN, an or IoT components in order to secure IoT with WSN's help. By combining the power of machine learning with the Firefly Algorithm, this study dramatically increases the state of current practice in IoT and wireless sensor networks security[15].

The Internet of Things (IoT) and Machine Learning (ML) are two of the most popular areas for study. ML and IoT are used in the implementation of "smart x" systems, which include Early Warning Systems (EWSs), Smart Homes, Smarter Cars, Smart Campuses (SCs), etc. These systems will change the way numerous entities in the world talk to one other. The important functions that IoT plays in SS are highlighted in this research. Additionally, this emphasises the value of ML in IoT-based SS. Additionally, a summary of smarts and IoT is provided[16].

In this work, a layered architecture for intelligent manufacturing application which combines Machine Learning (ML) with Blockchain Technology (BCT) is presented inside the Industrial Internet-of-Things (IIoT). The suggested architecture consists of five layers: application, complex services (i.e., BCT data, ML, and cloud), network/protocol, transportation controlled with BCT elements, and sensing. While ML brings its effectiveness in attack detection, such as DoS (Denial of Service), DDoS (Distributed Denial of Service), injections, Man in the Middle (MitM), brute force, Cross-Site Scripting (XSS), and scanning attempts by using classifiers separating among normal and malicious behaviour, BCT enables for the gathering of sensor access control information. To identify achievable advantages, our architecture's design is contrasted with comparable designs found in the literature[17].

The explosion of digital material has increased demand for computerised text classification techniques, particularly regarding Natural Language Processing (NLP)-based news classification. With an emphasis on Naive Bayes (NB) algorithms for categorising news headlines, this paper presents a Python-based news categorization system. The MLP Classifier demonstrated its efficacy by achieving the best accuracy, while Multinomial and Complementary Naive Bayes (NB) shown resilience in the categorization of news. Proper pre-processing of the data was essential to proper classification[18].

While still developing, machine learning's use in the intensive care unit is presently restricted to prognostic and diagnostic purposes. Sequential dynamic analysis of variables is provided via a straightforward and user-friendly machine learning technique named the Decision Tree (DT) algorithm. It is easy to use and might be a useful tool for bedside clinicians during COVID-19 to forecast ICU outcomes and support important choices like patient allocation in the case of restricted ICU bed capacity and end-of-life decisions[19].

One significant virtual network that enables distant users to access connected multimedia devices is the Internet of Things (IoT). Continuous research efforts are a result of the growth of IoT and its widespread applicability across several fields of daily life. Because security is so important to the adoption of any new technology, it is a perceptual issue for researchers working in the Internet of Things. To handle a specific scenario of protecting an IoT network, a great deal of research has been done focusing on the degree of security available on a certain mechanism, on particular applications, or on classifying vulnerabilities[20].

4. IMPLEMENTATION

The main goal of the paper is to implement security authentication measures to verify the identity of IoT devices before they connect to the network, as mentioned earlier in the previous sections. The study also evaluates the efficacy of machine learning algorithms in identifying IoT network attacks. This section outlines the process of device authentication, which involves verifying the device's identity before granting it access to the network, dataset, and machine learning algorithms. It also details the implementation steps.

4.1. Dataset

The Bot-IoT dataset was created by UNSW Canberra in 2018 and released in 2019 as a practical dataset that mainly centers on IoT infrastructure [21]. The Bot-IoT dataset is a publicly accessible dataset designed for machine learning-based research on detecting botnets, particularly targeting IoT devices. The dataset was developed to fill the gap in publicly accessible datasets for detecting botnets on IoT devices. The dataset contains network traffic data gathered from a diverse network of IoT devices, including cameras, routers, and printers. The data was gathered in a controlled laboratory setting, where different botnet attacks were initiated on the devices [22].

The Bot-IoT dataset is divided into two parts: the training set and the testing set. The training set consists of network traffic data from 10 IoT devices, including cameras, routers, and printers, which were infected with different types of botnets. The data was collected over a period of 20 days, with 10 days of normal traffic and 10 days of botnet traffic. The testing set consists of network traffic data from 9 different IoT devices, including cameras, smart home hubs, and smart TVs, which were also infected with various botnets. The data was collected over a period of 7 days, with 3 days of normal traffic and 4 days of botnet traffic. Splitting the dataset into training and testing sets enables the assessment and comparison of various machine learning models for botnet detection. The Bot-IoT dataset is utilized for binary classification to differentiate normal traffic from botnet traffic. The objective is to develop a machine learning model capable of accurately categorizing network traffic as either normal or botnet traffic.

4.2. Machine Learning Algorithms

We used the Bot-IoT dataset to evaluate four machine learning classifiers: RF, NB, DT, and XGBoost.

4.2.1. Random Forest (RF) and Decision Tree (DT):

These two methods, Random Forest and Decision Tree, both use rules from data features to build a model. An decision tree (DT) can be used for both sorting things into groups and figuring out what those groups mean. A lot of data may be needed to store Random Forest models, but they work well with them. If the data isn't balanced, this algorithm can still handle it. AI programs like XGBoost and random forest (RF) use a lot of different kinds of decision trees to make their more complex rules. At UNSW University's Cyber Range and IoT Lab, the newest datasets (TON-IOT) are all about testing different machine learning methods for hacking into IoT devices. They include Telemetry datasets for IoT and IIoT sensors, Operating systems datasets for Windows 7 and 10, Ubuntu 14 and 18 TLS, and Network traffic datasets, among other types of data. An experiment taught and tested DT, RF, XGBoost, ANN, and Multi-layer Perceptron algorithms to tell the dif-

ference between attacks and normal network traffic. The study used the Accuracy, Precision, Recall, F1-Score, and Confusion Matrix to rate how well the model did its job.

Random Forest (RF) is a widely used machine learning method suitable for classification and regression tasks. An ensemble method that enhances prediction accuracy and reduces variance by combining multiple decision trees. The Random Forest algorithm operates by generating a collection of decision trees using random samples of the training data and features. Each tree is constructed autonomously, without any communication between them, and the ultimate prediction is derived from the average of all the trees' predictions [23-24].

4.2.2. Naive Bayes (NB):

Naive Bayes is a classification algorithm that works well in both two-class and multi-class settings. It has been used to find anomalies and intrusions. This algorithm works really well with discrete data. The naive Bayes classification is used a lot in IDS because it is easy to understand and quick to compute. A specific set of detected traffic parameters, such as status flags, protocols, and latency, have been used to figure out how likely it is that network traffic is either abnormal or normal. It figures out how likely each feature is given a class and what the prior probability is for each class. Based on the features that have been seen, the probabilities are used to figure out the posterior probability of each class [25].

4.2.3. Extreme Gradient Boost (XGBoost):

Gradient Boosting is a common machine learning method for creating powerful models. It takes several weak predictive models and combines them into a single strong model. The method is famous for being able to make correct and dependable predictions, even on datasets that are complicated and have a lot of features. Gradient Boosting works on the same idea as boosting, where a set of weak learners are iteratively combined to create a strong model. We chose the XGBoost algorithm because is faster and show better performance than other classifiers [26].

4.3. Implementation Steps

Our method consists of five basic steps: device authentication, data analysis, feature selection, preprocessing, and classification.

4.3.1. Device Authentication:

The authentication of IoT devices functions as the primary defence line in IoT security, ensuring that only verified and trusted devices can access the network. This authentication process is integral to our holistic security approach, complementing the botnet attack detection system by preventing unauthorized access from the outset, consequently minimizing the potential attack surface for botnets. Having recognized the critical role of device authentication in bolstering the security of IoT networks, we present our comprehensive approach that combines the power of ML with robust device authentication mechanisms. While the classification of potential threats remains a pivotal step in securing IoT networks, it is imperative to underscore the significance of device authentication as the primary safeguard against unauthorized access and potential breaches. The security of IoT devices and networks is a critical concern in the current digital age.

In this digital age, keeping IoT devices and networks safe is very important. Strong authentication mechanisms are one of the most important parts of making sure this security. In the context of IoT, authentication is the process of making sure that a device is who it says it is before it can connect to

the network. The integrity of the IoT network is protected by this process, which stops unauthorized access and possible attacks [27].

Another library that we utilized in this study is the AWS IoT Device SDK for Python. This SDK, or Software Development Kit, provides easy-to-use APIs and handles the low-level details of communication with AWS IoT Core. AWS IoT Core is a managed cloud service that lets connected devices interact with cloud applications and other devices, providing secure and bidirectional communication between IoT devices and the Amazon Web Services (AWS) cloud [28]. In our implementation, we used MQTT with TLS client authentication for device security. TLS, or Transport Layer Security, is a cryptographic protocol designed to provide communications security over a computer network.

We used TLS to make sure that the communication between the IoT devices and the MQTT broker is safe and that each device's identity is checked before it can connect to the network. Adding authentication to IoT networks is a complicated process that needs to be thought through carefully, taking into account the network's needs and the devices' protocols.

Therefore, it is important to choose the appropriate libraries and methods for the specific context of the IoT network. In this study, we found that the combination of Python, Paho MQTT, and AWS IoT Device SDK provided a robust and flexible framework for implementing device authentication in our IoT network.

4.3.2. Data Analysis:

The Bot-IoT dataset comprises numerous dependent and independent features that can be utilized to train ML models. The dependent feature in the dataset is binary classification, indicating whether a network flow is benign or malicious. This feature is essential for training supervised learning models and is based on the network traffic captured by the network taps in the testbed environment. The independent features in the Bot-IoT dataset can be divided into two categories: traffic flow-based features and host-based features. The traffic flow-based features relate to the flow of network traffic between two hosts, while the host-based features relate to the characteristics of a single host. The traffic flow-based features include the number of packets and bytes transferred between hosts, the average packet size, and the duration of the flow. It also has information on the source and destination ports used for the connection, as well as the protocol used for the flow, such as TCP or UDP. The Bot-IoT dataset has host-based features that tell you about the device and operating system being used. This information includes the type of device, the operating system version, and the name of the maker and model of the device.

Additionally, the dataset includes information about the network interfaces and services being used by the device. These independent features can be used to train ML models to identify and classify malicious network traffic accurately. For example, host-based features like device type and operating system version can be used to identify vulnerabilities that are specific to certain devices or operating systems.

Additionally, the dataset includes information about the network interfaces and services being used by the device. These independent features can be used to train ML models to identify and classify malicious network traffic accurately. For example, host-based features like device type and operating system version can be used to identify vulnerabilities that are specific to certain devices or operating systems. In the same way, different types of network attacks can be found and categorized by looking at traffic flow-based factors such as the number of packets sent and the length of the flow. Overall, the combination of dependent and independent features in the Bot-IoT dataset

provides a comprehensive set of data that can be used to train and evaluate ML models for network security applications. In our research.

4.3.3. Feature Selection:

Feature selection is an important step in the data analysis process, as it helps identify the most relevant and informative features for building ML models. In this paper, we used top correlating features that were selected after conducting a correlation analysis between the features in the Bot-IoT dataset and the target variable, we identified a set of features that exhibited strong correlation.

4.3.4. Pre-Processing:

This is an important step in the early stages of machine learning. It prepares data for ML models by getting rid of unnecessary data that could affect how accurate the dataset is.

4.3.5. Classification:

Attacks or intrusions are distinguished from routine network events using the ML technique known as classification. The tests were all done in Python with the help of machine learning libraries for Python.

4.3.6. Evaluation Metrics

When assessing the performance of machine learning models, it is essential to establish performance metrics such as accuracy, precision, recall, and F1-score.

$$Precision = \frac{TP}{TP+FP}\ldots\ldots \quad (1)$$

$$Recall = \frac{TP}{TP+FN}\ldots\ldots \quad (2)$$

$$F1score = \frac{2*(precision * recall)}{precision + recall}\ldots\ldots \quad (3)$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}.. \quad (4)$$

5. RESULTS

Based on the results obtained from the various ML models, it can be concluded that the proposed intrusion detection system is effective in detecting and classifying network attacks. The XGBoost model, in particular, achieved an accuracy of 99.98% in detecting attacks and an accuracy of 99.99% in classifying attack types. The system also demonstrated high precision and recall rates across all attack types and subcategories.

The results indicate that the suggested intrusion detection system could be a valuable asset for network security experts in detecting and addressing potential attacks. The system can continuously monitor network traffic and promptly alert security personnel, enabling them to respond

effectively to mitigate the impact of attacks. Additional assessment of the system's performance on larger and more varied datasets is required to confirm its efficacy in real-world situations.

Table 1. Models accuracy using feature importance

No.	ML Algorithms	Attack	Category	Subcategory
1	Random Forest	0.99	0.91	0.98
2	Naive Bayes	0.99	0.71	0.98
3	Decision Tree (Information Gain)	0.99	0.90	0.97
4	Decision Tree (Gini Index)	0.99	0.91	0.98
5	Gradient Boost	0.99	0.99	0.99

Future researchers should explore various machine learning algorithms and techniques to optimize the prediction of network traffic anomalies. Moreover, enhancing the model by including additional features pertaining to network traffic, such as packet size and protocol, may enhance its accuracy. Continuously updating the dataset and utilizing larger datasets is crucial for accurately reflecting present network traffic patterns. It is advisable to take into account the ethical considerations of analyzing network traffic and put in place measures to safeguard user privacy.

6. CONCLUSIONS

In the next years, a wide range of assaults might target user-transmitted data because the majority of it is sent wirelessly. Furthermore, when wireless activities are integrated with Internet of Things applications like medical, surveillance, etc., greater data security is needed. However, limitation of system development and management currently have significantly lower security requirements for IoT-related processes, which results in users retaining poor integrity in IoT environmental situations.

The major challenges and limitations of this study are that using machine learning to find cyber threats in the IoT environment is hard for a number of reasons. For example, the data that IoT devices generate is very complicated, and IoT devices have a lot of different architectures, protocols, and operating systems, which is another big problem. Finally, cyber threats are always changing, so continuous learning is needed.

The intrusion detection system is effective in detecting and classifying network attacks based on the results from different machine learning models like Random Forest, Naive Bayes, Decision Tree, and Gradient Boost. This research is efficient as it achieved a 99.98% accuracy rate through a XGBoost straightforward implementation.

FUTURE WORK

In result, the work presents a path for further study that utilises ML-based AA to address IoT security issues in a coordinated and cooperative manner. Ultimately, it is envisaged that this work will act as a platform for further developments in cybersecurity analytics using effective neural networks and machine learning models on edge devices in the future.

REFERENCES

[1] X. Li, R. Lu, X. Liang, and X. Shen, “Smart community: An Internet of Things application,” IEEE Commun. Magazine, vol. 49, no. 11, Nov. 2011, pp. 68–75.
[2] Z. Sheng, S. Yang, Y. Yu, and A. Vasilakos, “A survey on the IETF protocol suite for the Internet of Things: Standards, challenges, and opportunities,” IEEE Wireless Commun., vol. 20, no. 6, Dec. 2013, pp. 91–98.
[3] X. Liu, M. Zhao, S. Li, F. Zhang, and W. Trappe, “A security framework for the Internet of Things in the future Internet architecture,” Future Internet, vol. 9, no. 3, pp. 1–28, Jun. 2017.
[4] H. Qiao, J. O. Blech, and H. Chen, "A Machine learning based intrusion detection approach for industrial networks," in 2020 IEEE International Conference on Industrial Technology (ICIT), 2020, pp. 265-270.
[5] Ghannadrad, "Machine learning-based DoS attacks detection for MQTT sensor networks," 2021.
[6] G. E. I. Selim, E. Hemdan, A. M. Shehata, and N. A. El-Fishawy, "Anomaly events classification and detection system in critical industrial internet of things infrastructure using machine learning algorithms," Multimedia Tools and Applications, vol. 80, pp. 12619-12640, 2021.
[7] G. Siaterlis, M. Franke, K. Klein, K. A. Hribernik, G. Papapanagiotakis, S. Palaiologos, et al., "An IIoT approach for edge intelligence in production environments using machine learning and knowledge graphs," Procedia CIRP, vol. 106, pp. 282- 287, 2022.
[8] Hussain, F.J.; Hussain, R.; Hassan, S.A.; Hossain, E. Machine learning in IoT security: Current solutions and future challenges. IEEE Commun. Surv. Tutorials 2020, 22, pp 1686–1721.
[9] Fatima Alwahedi, et.al. , “Machine Learning Techniques for IoT security : Current Research and Future vision with generative AI and Large Language models”, Internet of things and cyber physical systems, vol 4, 2024, pp 167-185.
[10] Shafiq, Muhammad; Gu, Zhaoquan; Cheikhrouhou, Omar; Alhakami, Wajdi; Hamam, Habib (3 August 2022). "The Rise of "Internet of Things": Review and Open Research Issues Related to Detection and Prevention of IoT-Based Security Attacks". Wireless Communications and Mobile Computing. 2022: e8669348
[11] Sahu, A. K., Sharma, S., & Raja, R. (2022). Deep learning-based continuous authentication for an IoT-enabled healthcare service. Computers and Electrical Engineering, 99, 107817.
[12] Kigo, S. N., Omondi, E. O., &Omolo, B. O. (2023). Assessing predictive performance of supervised machine learning algorithms for a diamond pricing model. Scientific Reports, 13(1), 17315.
[13] Noorunnahar, M., Chowdhury, A. H., & Mila, F. A. (2023). A tree based eXtreme Gradient Boosting (XGBoost) machine learning model to forecast the annual rice production in Bangladesh. PloS one, 18(3), e0283452.
[14] Sarker, I. H., Khan, A. I., Abushark, Y. B., &Alsolami, F. (2023). Internet of things (iot) security intelligence: a comprehensive overview, machine learning solutions and research directions. Mobile Networks and Applications, 28(1), pp 296-312.
[15] Karthikeyan, M., Manimegalai, D., &RajaGopal, K. (2024). Firefly algorithm based WSN-IoT security enhancement with machine learning for intrusion detection. Scientific Reports, 14(1), 231.
[16] Abdalzaher, M. S., Fouda, M. M., Elsayed, H. A., & Salim, M. M. (2023). Toward Secured IoT-Based Smart Systems Using Machine Learning. IEEE Access, 11, pp 20827-20841.
[17] Mrabet, H., Alhomoud, A., Jemai, A., & Trentesaux, D. (2022). A secured industrial Internet-of-things architecture based on blockchain technology and machine learning for sensor access control systems in smart manufacturing. Applied Sciences, 12(9), 4641.
[18] Veziroğlu, M., Eziroğlu, E., &Bucak, İ. Ö. (2024). Performance Comparison between Naive Bayes and Machine Learning Algorithms for News Classification. In Bayesian Inference-Recent Trends. Intech Open.
[19] Elhazmi, A., Al-Omari, A., Sallam, H., Mufti, H. N., Rabie, A. A., Alshahrani, M. ... & Arabi, Y. M. (2022). Machine learning decision tree algorithm role for predicting mortality in critically ill adult COVID-19 patients admitted to the ICU. Journal of infection and public health, 15(7), pp 826-834.
[20] Cherbal, S., Zier, A., Hebal, S., Louail, L., &Annane, B. (2024). Security in internet of things: a review on approaches based on blockchain, machine learning, cryptography, and quantum computing. The Journal of Supercomputing, 80(3), 3738-3816.
[21] “The bot-IOT dataset,” The Bot-IoT Dataset | UNSW Research, https://research.unsw.edu.au/projects/bot-iot-dataset (accessed Nov. 3, 2023).
[22] 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,” Future Generation Computer Systems, vol. 100, 2019, pp. 779–796.

[23] M. S. Alam and S. T. Vuong. S.L., “Random forest classification for detecting android malware,” .: in Proc. IEEE Int. Conf. Green Comput.Commun. IEEE Internet Things IEEE Cyber, Phys. Soc. Comput.,, Aug. 2013,. pp. 663–669.
[24] Experimental Analysis of Classification for Different Internet of Things (IoT) Network Attacks Using Machine Learning and Deep learning. Tasnim, Anika, et al. Chiangrai, Thailand : IEEE, 2022.
[25] L. Deng, D. Li, X. Yao, D. Cox, and H. Wang,. s.l., “Mobile network intrusion detection for IoT system based on transfer learning algorithm,”. : Clust. Comput., vol. 22, Jan. 2018, pp. 9889–9904.
[26] T. Chen and C. Guestrin, “XGBoost,” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016. doi:10.1145/2939672.2939785.
[27] P. Perkis and C. Mozenter, “AR,” Amazon, https://aws.amazon.com/ar/blogs/iot/how-to-implement-mqtt-with-tls-client-authentication-on-port-443-from-client-devices-python/ (accessed Nov. 3, 2023).
[28] D. Tao, “How to use MQTT in python with Paho Client,” www.emqx.com, https://www.emqx.com/en/blog/how-to-use-mqtt-in-python (accessed Nov. 3, 2023).