Enhancing Internet of Things Security with Random Forest-Based Anomaly Detection

Ahmed Al Shihimi † , Muhammad R Ahmed†, Thirein Myo† and Badar Al Baroomi†

Military Technological College, Muscat, Oman

Abstract

The Internet of Things (IoT) has revolutionized communication and device operation, but it has also brought significant security challenges. IoT networks are structured into four levels: devices, networks, applications, and services, each with specific security considerations. Personal Area Networks (PANs), Local Area Networks (LANs), and Wide Area Networks (WANs) are the three types of IoT networks, each with unique security requirements. Communication protocols such as Wi-Fi and Bluetooth, commonly used in IoT networks, are susceptible to vulnerabilities and require additional security measures. Apart from physical security, authentication, encryption, software vulnerabilities, DoS attacks, data privacy, and supply chain security pose significant challenges. Ensuring the security of IoT devices and the data they exchange is crucial. This paper utilizes the Random Forest Algorithm from machine learning to detect anomalous data in IoT devices. The dataset consists of environmental data (temperature and humidity) collected from IoT sensors in Oman. The Random Forest Algorithm is implemented and trained using Python, and the accuracy and results of the model are discussed, demonstrating the effectiveness of Random Forest for detecting IoT device data anomalies.

Keywords:

Internet of Things, Challenges, Security, Random Forest

1. Introduction

Since the Internet of Things (IoT) has been introduced in the past few years, it has revolutionized the way devices and objects communicate and the way they operate as well by providing them with the ability to connect to one another[1], [2]. However, with this technological advancement comes significant security challenges that need to be effectively addressed[3]. IoT networks are composed of different layers, including devices, networks, applications, and services, each requiring specific security considerations[4]. Various types of IoT networks, such as PANs, LANs, and WANs, have distinct characteristics and specific security requirements[5], [6]. Communication protocols like Wi-Fi and Bluetooth, commonly used in IoT networks, can introduce vulnerabilities that necessitate additional security measures[7]. Security challenges in IoT encompass physical security, authentication, access control, encryption, software vulnerabilities, DoS attacks, data privacy, and supply chain security[8]. By implementing a multi-layered approach, including authentication mechanisms, encryption protocols, and security policies, we can enhance IoT network security and mitigate potential risks, ensuring trust in this transformative technology[9]. A generic Internet of Things (IoT) network with applications possibilities is shown in Fig 1.

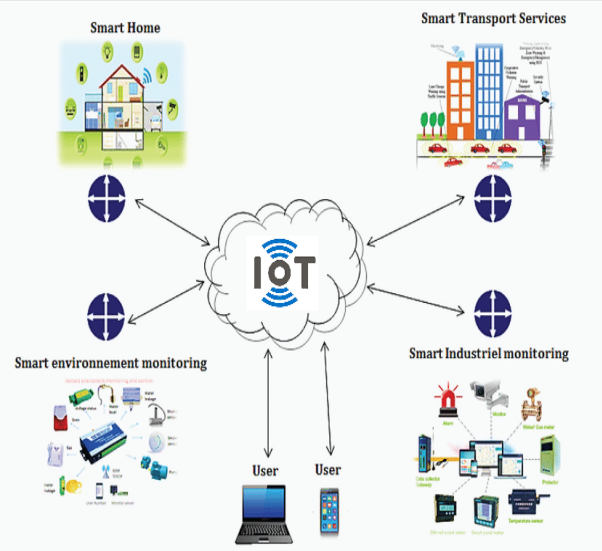

Fig 1. A generic Internet of Things (IoT) network

The four-layer model is a widely recognized framework for designing and implementing Internet of Things (IoT) systems[10]. The lowest layer of the system is the perception layer, where sensors and actuators collect data from the environment. The network layer establishes connectivity among devices using various networking technologies [11]. The middleware layer acts as a bridge between the perception and application layers, integrating devices and managing communication protocols. The application layer provides an interface for end-users to interact with the IoT system and offers applications for data acquisition, processing, and presentation. The four-layer architecture offers modularity and flexibility, allowing

Manuscript received June 5, 2024
Manuscript revised June 20, 2024
https://doi.org/10.22937/IJCSNS.2024.24.6.9

customization for specific requirements and enabling the development of new IoT applications[12]. However, challenges such as data security, privacy, and interoperability need to be addressed to fully leverage the potential of the four-layer IoT architecture. By overcoming these challenges, we can unlock the transformative power of IoT across diverse domains and applications. The four layer architecture are shown in the Fig. 2.

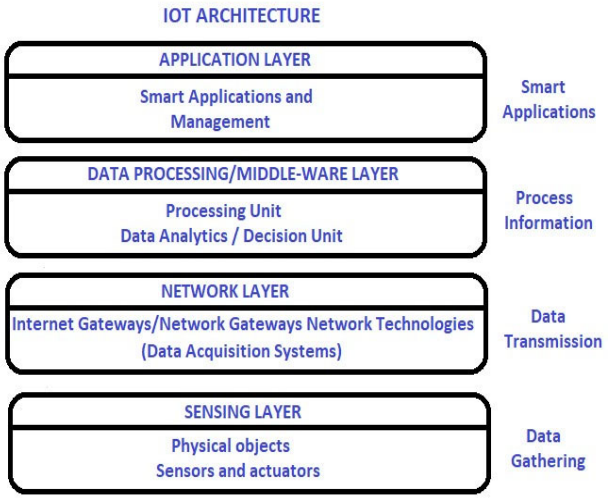

Fig 2. The IoT four layer Architecture

There are many types of networks emerging due to the rapid development of Internet of Things (IoT) networks which enable seamless communication and data exchange between devices. Each network type offers unique characteristics, advantages, and limitations, catering to specific applications. Wireless Personal Area Networks (WPANs), such as Bluetooth and Zigbee, enable short-range wireless communication, making them ideal for applications like home automation and wearable devices. Wireless Local Area Networks (WLANs) utilize Wi-Fi technology to cover larger areas like buildings and campuses, serving purposes in smart cities and industrial automation[13], [14]. Wireless Wide Area Networks (WWANs) span extensive geographical areas, connecting devices in transportation, logistics, and environmental monitoring. Low-Power Wide Area Networks (LPWANs) provide long-range connectivity with minimal power consumption, serving applications in agriculture, buildings, and cities. Satellite networks offer global coverage but are limited by their high cost and latency, making them suitable for niche applications such as maritime and aviation. Power Line Communication (PLC) networks leverage power lines for data transmission, commonly found in smart homes and grid applications. It is important to be aware that choosing an IoT network depends on factors such as the range of coverage, the bandwidth requirements, the latency, and the power consumption limitations. Technological advancements and an increasing demand for IoT applications will undoubtedly lead to IoT networks evolving to meet the needs of various industries and use cases.

In order for IoT networks to operate smoothly, the choice of a communication protocol will play a crucial role in enabling efficient data transfer between devices. A variety of communication protocols have been developed in order to satisfy the diverse demands of IoT applications in terms of bandwidth, latencies, and power consumption in order to cater to the diverse requirements of IoT applications. Some commonly used protocols in IoT networks include MQTT, a lightweight messaging protocol for low-bandwidth networks; CoAP, a lightweight protocol for memory-constrained devices; HTTP, enabling web-based communication with IoT devices; Zigbee, a low-power protocol for home automation and industrial control; LoRaWAN, providing long-range connectivity for IoT devices; BLE, designed for low-power devices like wearables and smart homes; and NFC, facilitating short-range data exchange for payment and access control systems[15], [16]. The selection of a communication protocol depends on factors like coverage range, data rate, and power constraints of the devices. Additionally, ensuring interoperability among devices and networks that employ different protocols is crucial for successful IoT system implementation.

Fig 3. Typical IoT Security Challenges

The development of an IoT network architecture presents numerous challenges. The challenges include scalability, interoperability, security, and privacy. In IoT networks, scalability is crucial as the number of connected

devices increases, requiring architectures that can handle increased traffic. Interoperability is another significant challenge, ensuring seamless communication between devices using different protocols. Security and privacy are paramount due to the vulnerability of IoT devices to cyber-attacks. Physical security measures can address risks such as tampering and theft, while robust authentication, access control, and encryption technologies safeguard against unauthorized access and data breaches. Regular software updates and patches mitigate software vulnerabilities, and measures like traffic filtering and connection reduction prevent DoS attacks. Protecting data privacy involves techniques like anonymization and data minimization. Ensuring supply chain security is essential to prevent unauthorized access or the introduction of malicious components. Implementing a multilayered approach involving physical, technical, and organizational measures is crucial. Additionally, the exponential growth of IoT devices introduces security challenges, given their constraints in terms of memory and computational power. Monitoring such a vast number of devices becomes increasingly challenging, potentially leaving gaps in security protocols that can be exploited. Data integrity is vital, and identifying anomalies in IoT data is crucial to maintain security[17]–[19]. Machine learning algorithms can play a significant role in real-time anomaly detection, but selecting the right algorithm and producing optimal models can be a complex task due to various dimensions involved.

In this paper we have used random forest algorithm to find the anomaly in the data to secure the Internet of Things Networks. The algorithm offers several advantages. It typically provides higher accuracy compared to a single decision tree since it reduces overfitting by averaging the predictions of multiple trees. In addition to that it can handle a large number of input variables and handle missing data effectively. Moreover, random forests can provide estimates of feature importance, allowing for better understanding of the underlying data.

Random forest algorithms have some disadvantages, such as the fact that they can be computationally expensive and may require a greater amount of resources due to the fact that they combine multiple decision trees. Training a large number of trees and searching for the best split at each node can be time-consuming. Another limitation is that random forests can struggle with imbalanced datasets where one class dominates the others.

There are several limitations to random forests, but they have become popular and widely used in a variety of domains, including finance, healthcare, and image recognition, due to their robustness, flexibility, and ability to handle complex problems.

2. Related works

The Internet of Things (IoT) has transformed device communication and operation, but it also brings significant security challenges. IoT networks consist of layers such as devices, networks, applications, and services, each with specific security considerations. PANs, LANs, and WANs are the three types of IoT networks with distinct security requirements. Communication protocols like Wi-Fi and Bluetooth have vulnerabilities that require additional security measures. Challenges include physical security, authentication, encryption, software vulnerabilities, DoS attacks, data privacy, and supply chain security. The four-layer IoT architecture and the choice of communication protocols are crucial for efficient and secure operation. Scalability, interoperability, security, and privacy pose challenges in IoT network development. Multilayered security measures and anomaly detection play vital roles in safeguarding IoT systems. several researchers have developed the algorithms to secure the IoT networks. We have discussed few major research conducted by the researchers.

Waheed et al. [20]propose the integration of Machine Learning (ML) techniques and Blockchain (BC) technology to enhance the security and privacy of IoT systems. They highlight the effectiveness of ML algorithms in detecting and predicting vulnerabilities, while emphasizing the relevance of BC in ensuring data integrity and secure communication within IoT networks. Secondly, the authors recognize the need for a comprehensive approach by combining ML and BC, as previous studies have focused on either one individually. However, one potential disadvantage of their approach is the increased complexity and computational overhead associated with implementing both ML and BC techniques simultaneously. This may pose challenges in terms of resource constraints and scalability, especially for resource-constrained IoT devices. Nevertheless, the authors' contributions shed light on the potential synergies between ML and BC in addressing IoT security and provide a foundation for further research in this domain.

This study done be bhabendu [21]makes significant contributions to the field of IoT security and privacy. It conducts a thorough review and identification of existing security and privacy issues within IoT systems, shedding light on the vulnerabilities that need to be addressed. Secondly, it explores the potential of blockchain technology as a solution to enhance security in IoT. The study provides detailed insights into the integration of blockchain with IoT, showcasing how this combination can mitigate security risks. Additionally, a case study is presented, implementing an Ethereum-based blockchain system in a smart IoT environment, which offers practical implications and real-world application. However, it is important to note that while blockchain technology can enhance security, it also

has its own limitations and challenges. The work in this research recognizes these drawbacks, confirming a complete understanding of the benefits and drawbacks of implementing blockchain in IoT systems. It provides valuable knowledge and recommendations that can contribute to the improvement of security and privacy aspects of IoT technology.

In a systematic review of the security requirements, attack vectors, and current security solutions for the Internet of Things done by Fatima et al [22], this lie was examined in its systematic review of IoT networks. Moreover, the identification of gaps in these solutions addressed through Machine Learning (ML) and Deep Learning (DL) approaches. In order to address a variety of security challenges, IoT devices and networks can benefit from embedded intelligence derived from ML and DL techniques. Furthermore, the paper discusses existing ML and DL solutions for different security problems in IoT networks, offering insights into their potential applications. However, it is important to note that ML and DL approaches also have their limitations, such as the need for significant computational resources and the possibility of adversarial attacks on the trained models. These disadvantages should be taken into consideration when implementing ML and DL solutions for IoT security. By highlighting the associated challenges and trade-offs of ML and DL, the research contributes to enhancing IoT security through ML and DL.

A multi-layer security approach has been proposed by Feroz et al[23] for addressing the security issues that arise in IoT networks, focusing in particular on jamming attacks. It introduces a detection mechanism for Distributed Denial of Service (DDoS) attacks, enhancing the protection of smart devices in IoT environments. The study also introduces a novel threshold-based countermeasure (TBC) to mitigate replay attacks at different layers. The proposed approach improves computational efficiency and energy consumption compared to existing schemes, offering a more scalable solution for securing IoT networks. One potential limitation of the study is that it primarily focuses on jamming attacks and replay attacks, while there may be other security concerns in IoT networks that are not fully addressed. Additionally, the paper does not extensively discuss the potential challenges or limitations of implementing the proposed multi-layer security approach.

The contributions of this paper by nizzi et al [24]include the proposal of a novel method called AShA (Address Shuffling Algorithm with HMAC) for performing network-wide address shuffling in IoT devices. AShA offers a simple implementation and minimal network overhead, making it suitable for resource-constrained devices. The theoretical analysis demonstrates how AShA parameters can be adjusted for different network sizes, while the simulations show its effectiveness in achieving collision-free address renewal in large networks. By constantly modifying the device footprint, AShA reduces the attack surface and enhances the security of IoT devices.

Ramya & Vijaya proposes a novel approach to address the security and privacy-preserving challenges in big data, specifically focusing on the healthcare industry[25]. In the A3DES algorithm, anonymization techniques and Triple DES encryption are integrated to protect sensitive data. Further, experimental results indicate that the proposed approach is superior to other related approaches, both in terms of performance and security. It is nevertheless important to acknowledge the potential disadvantages of the proposed approach as well as the advantages it offers. One of the potential limitation is the computational overhead introduced by the anonymization and encryption processes, which may impact the overall processing time and resource utilization. Additionally, the effectiveness of the approach may be influenced by the quality of anonymization and the choice of encryption algorithms.

Abiodon et al [26] have proposed a cryptography-based solution to address the security and assurance concerns in the context of big data generated by IoT devices. Specifically, the Triple Data Encryption Standard (3DES) algorithm is utilized to secure the IoT-generated data, ensuring its confidentiality and integrity. The performance evaluation of the proposed method demonstrates its effectiveness compared to other techniques, highlighting its potential for privacy and security in IoT data generation. Conversely, implementing cryptography-based technology introduce additional computational overhead and resource requirements. This affect the overall system performance.

Tewari et al [27] proposed a mutual authentication mechanism based on elliptic curve cryptography (ECC) for securing IoT devices. This mechanism addresses the growing concern of security in the IoT technology and offers advantages in terms of communication overhead and resistance against attacks. The proposed solution satisfies essential security requirements and has been validated through a series of security and performance analyses that have been performed. Cryptography method always makes the system slow.

A hybrid DNA-encoded elliptic curve cryptography (ECC) scheme for enhancing security in IoT devices was proposed by Durga [28]. The proposed scheme combines the strengths of ECC and DNA-based encryption in such a way that it offers multilevel security and enhanced stability. The unique way in which DNA sequences are selected and assigned, in combination with binary conversion and ECC encryption, provides double security. The paper demonstrates the feasibility of implementing this approach on IoT devices through various examples. One of the potential disadvantage of the proposed scheme is the increased computational complexity associated with DNA encoding and decoding processes. This may result in higher processing overhead and potentially impact the performance of resource-constrained IoT devices. Future

research and optimization efforts should be directed towards mitigating these computational challenges to ensure the practicality and efficiency of the proposed solution.

Prokash et al [29] proposed an improved elliptic curve digital signature algorithm (ECDSA) scheme, which enhances both security and efficiency compared to the original scheme. The main improvement lies in the elimination of the time-consuming finite field inversion process, resulting in faster computation speed and a reduced ratio of verification time to signature generation time. The proposed scheme offers practical significance in improving the efficiency of elliptic curve cryptography. Simulation results confirm the scheme's faster performance and higher efficiency in signature generation and verification without compromising security.

Mohammed et al [30] described an algorithm for securing IoT data using a three-stage encryption algorithm, which is specifically designed for greenhouse applications. There is no need for traditional RSA keys to be exchanged in this algorithm, thus resulting in a streamlined encryption process and a 30% reduction in the transmission time. The technology offers a high level of encryption, simplicity, and energy efficiency, which addresses the limitations of IoT nodes. Further evaluation is needed to validate its performance under different scenarios, and its effectiveness against advanced security threats should be examined.

In supervised machine learning, Random Forest uses an ensemble of decision trees. A large number of industries, such as finance, e-commerce, and health care, use machine learning for a variety of purposes, including predictive analysis and pattern recognition. Using the algorithm, multiple decision trees are created, each trained on a different set of data and features. There are many trees in the forest, so the algorithm gathers predictions from all of them and outputs the most frequent or average prediction out of all trees. Using this ensemble approach, the model is more accurate, generalizable, and resilient to overfitting, due to the fact that it is more resilient to overfitting. Furthermore, Random Forest is known for its ability to handle high-dimensional data as well as dealing with missing values, in addition to identifying important features. It is still important to keep in mind that decision trees may be computationally expensive and may also be difficult to interpret when compared to individual decision trees. However, the versatility and effectiveness of the tool make it useful for a variety of different machine learning tasks, as a consequence of its versatility

3. Methodology

Random Forest is a machine learning approach commonly used for regression and classification tasks. It employs ensemble learning, a technique that combines multiple classifiers to address complex problems effectively. The random forest algorithm consists of numerous decision trees. These decision trees are trained using bagging or bootstrap aggregating, a process that enhances the accuracy of machine learning algorithms. In this algorithm, the predictions of the decision trees are used to determine the final outcome[31]. The algorithm makes predictions by averaging or taking the mean of the outputs from multiple trees. As the number of trees increases, the precision of the predictions also improves.. The algorithm offers several key features which inclides[32]:

- Improved Accuracy: Compared to the decision tree algorithm, Random Forest tends to provide higher accuracy in predictions.
- Handling Missing Data: It offers an effective mechanism for handling missing data, allowing for robust analysis even with incomplete information.
- Reduced Hyper-parameter Tuning: Random Forest can generate reasonable predictions without extensive hyper-parameter tuning, simplifying the model development process.
- Overfitting Mitigation: It addresses the issue of overfitting commonly encountered in decision trees, resulting in more generalized and reliable models.
- Random Feature Selection: At each splitting point within each tree of the Random Forest, a subset of features is randomly chosen, enhancing diversity and reducing bias in the model's predictions.

Random Forest algorithms are built upon decision trees, making it essential to understand the fundamentals of decision trees. A decision tree is a tree-like structure used for decision support. It comprises three components: decision nodes, leaf nodes, and a root node. The decision tree is built by selecting appropriate attributes to partition the sample set into subsets, creating branch nodes until the samples in each node share the same type or meet a termination condition. This algorithm models structured thinking similar to how humans approach problems. For example, when examining financial data, one might prioritize assessing profitability. Decision tree construction involves two main steps: generating the decision tree using a training sample set and pruning the tree to refine its accuracy[33]. The decision tree construction algorithm follows a specific input sample set format, which is in equation (1)

$$I = \{(A_{00} \ldots, A_{0j} \ldots, A_{0n}, T_0) \ldots (A_{i0} \ldots, A_{ij} \ldots, A_{in}, T_i) \ldots\} \quad (1)$$

In the process of constructing a decision tree, we use a representation in which A_{ij} represents the value of the j-th

attribute in the set of the i-th sample, and T_i represents the mark type of the i-th sample. Then the output of the algorithm is the binary tree or a multi-branch tree. the binary trees are commonly utilized when the collection of data consists of attributes that can be evaluated using the judgments done by boolean logic. The decision tree method is shown in figure 4.

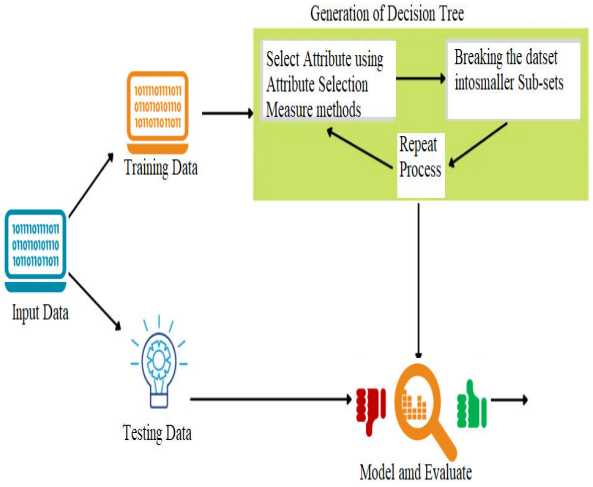

Fig 4. Decision tree method

Different decision tree classification algorithms employ various criteria to select split attributes. Among these criteria, information gain and information gain rate are two crucial ones. Information gain is used to determine the split attribute selection[34]. Let's consider a the sample set of training S and a set of attribute is in equation (2). Next, we calculate the proportion samples that belong to the category of j-th in the dataset as in equation (3). At this point, we compute the information entropy of the dataset samples S as in equation (4).

$$P = \{p_1 \dots p_i, \dots p_m\} \quad (2)$$

$$P(C_j) = \frac{|S_{ij}|}{|S|} \quad (3)$$

$$Entropy(S, p_i) = \sum_{j=1}^{n} -P(C_j) log_2 P(C_j) \quad (4)$$

Let's consider a sample dataset where the attribute pi has a value range $[a, b]$. We can define $S_{i(x)}$ as the samples subset in which the attribute p_i takes the value x. In this case, the information gain in the set of samples S for the attribute p_i can be determined using the following as in equation (5).

$$Gain(S, p_i) = Entropy(S, p_i) - \sum_{v \in v_i} \frac{|S_i(v)|}{|S|} Entropy(S, p_i) \quad (5)$$

Once the information gain in the set samples S is computed, the split information of S on the attribute p_i it is determined using the equation (6)

$$SplitGain(S, p_i) = -\sum_{v \in v_i} \frac{|S_i(v)|}{|S|} log_2 \frac{|S_i(v)|}{|S|} \quad (6)$$

The gain rate of information of the dataset sample S with respect to the attribute p_i is calculated as follows:

$$GainRatio(S, p_i) = \frac{Gain(S, p_i)}{SplitInfo(S, p_i)} \quad (7)$$

Random forest is a powerful ensemble classifier method that consists of decision trees. The combination of random forest and decision trees is illustrated in Figure 3. The samples reaching each internal node are divided into blocks based on this attribute. The end nodes, also known as leaf nodes, represent data collections with classification labels. The relationship between the root node and the leaf node of the tree is called a discriminant rule. The decision tree algorithm is based on a greedy approach used by top-down algorithms, where each internal node selects the attribute that yields the best classification result for the data partitioning. The process of classification continues until the tree is able to make an accurate classification. Among the key challenges that must be overcome in order to implement the decision tree algorithm is to select the optimal splitting attribute. Gini index used for attribute selection.

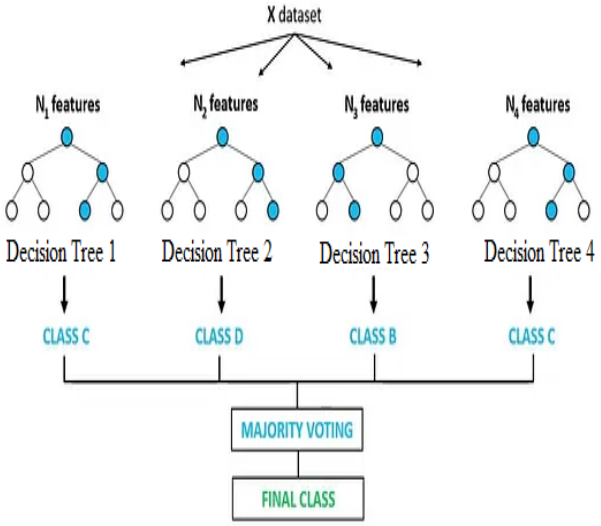

Fig 5. Random forest

In the algorithm ensemble-based machine learning that combines multiple decision trees to make predictions. Each decision tree in the random forest is trained on a different subset of the training data and uses a random subset of features for splitting at each node[35], [36]. The predictions from individual trees are then aggregated to obtain the final prediction.

4. Implementation of the Algorithm

The Random Forest found as most appropriate machine learning algorithm to identify the anomalies in metrological dataset containing temperature and humidity as features set, received from IoT devices. It also proposed an optimized model for detection of anomalies in metrological data (i.e. Temperature and Humidity) received from different IoT devices. It uses machine learning classifier technique by implementation and training of model in python language using Random Forest Algorithm. Figure 6 Implemented model for anomalies identification in data received from IoT Devices

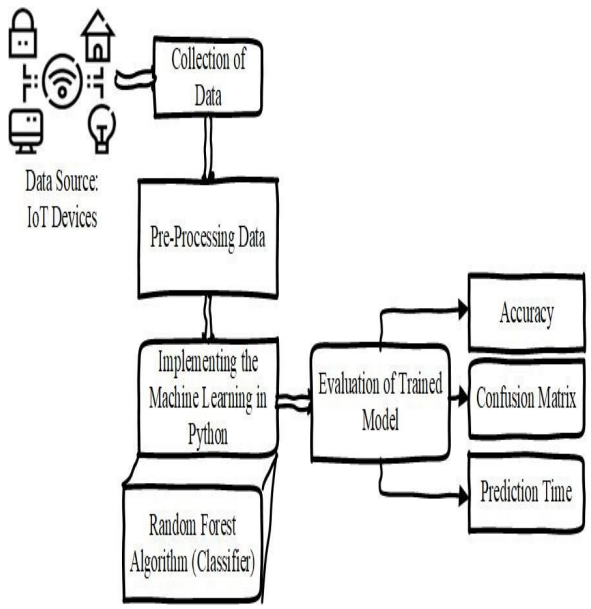

Fig 6. Implementation of random forest

In the implementation the sequential steps of the model for anomalies detection in dataset of IoT devices are as follows.

- Collection of data (Source: IoT Devices)
- Pre-processing of received data
- Implementing the classifier technique of machine using Random Forest (Classifier) Algorithm.
- Evaluation of trained model based on accuracy, confusion matrix and prediction time.

5. Results

We have implemented the algorithm in Python version 3.9.12 is with Jupyter Notebook IDE. Considering the average data in muscat from accuweather for April 2023 and May 2023 , we have created a hypothetical dataset for our scenario, therefore, to fit the criteria sample dataset containing 100 rows is being used. It contains two feature (temperature and humidity) and the target class labeled as anomaly. The normal ranges related to the environmental climate of Oman are 30 to 40 degree for temperature and 40 to 60 percent humidity, any values lower or higher than these ranges are considered to be anomaly. The Scatterplot and histogram of temperature and humidity of dataset is shown in the figure 7.

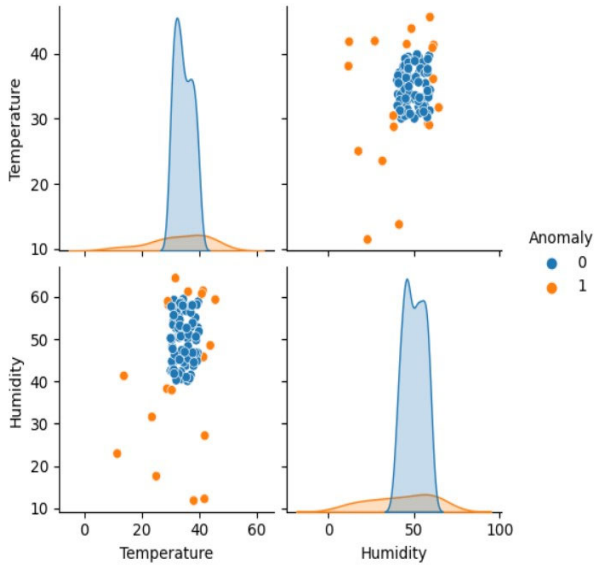

Fig 7. temperature and humidity of dataset

The data contains one hundred unique rows with two features labeled as temperature and humidity. Furthermore, the target class has two labels ‘0’ and ‘1’, where ‘0’ represents the normal/no anomaly and ‘1’ represents the abnormal/anomaly detection. The entire dataset was split into two portions, i.e. 80% for the training dataset and 20% for the testing the trained classifier model. In this process the random forest tree aggregation is shown in the figure. To illustrate the functioning of the random forest algorithm as an ensemble, a decision tree sample with an estimator value of 3 is selected. By utilizing the voting mechanism, which aggregates the individual predictions of multiple

decision trees, a collective prediction is generated as shown in figure 8.

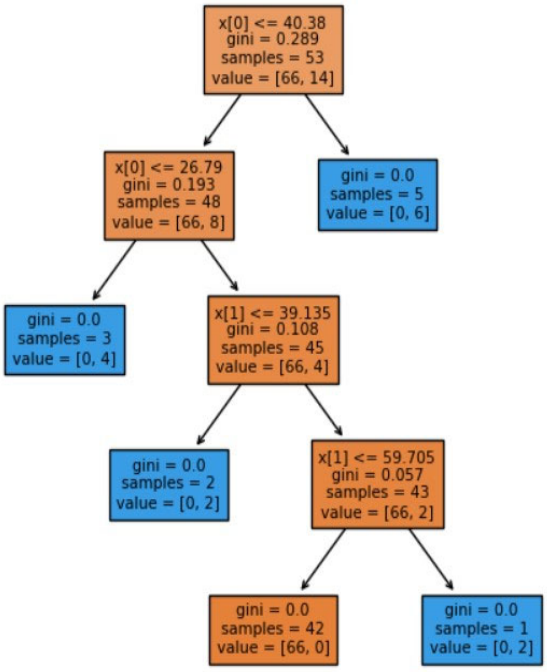

Fig 8. Individual predictions of multiple decision trees

The model training time is recorded up to: 4.000 secs and on average time of 2.87 secs, in the present environment explained in methodology. Furthermore, with current dataset the prediction time is up-to one second on average.

For classification models the most common metric is accuracy, it determines the correct prediction's fraction made by a model with respect to the total number of predictions that it has made. It gives a general overview of measuring metrics. Accuracy is calculated by following equation:

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Predictions} \quad (8)$$

Accuracy score is recorded as 0.95 which means it classification accuracy is up-to 95%, with the given training and test data.

The confusion matrix reflect the prediction's actual results into four categories i.e TP – True Positive, TN – True Negative, FP – False Positive and FN – False Negative. Confusion matrix is given as following box.

The labels, TP (True Positive) and TN (True Negative) shows the segment of correctly predicted from the dataset. Whereas, FP (False Positive) and FN (False Negative) indicate the segment of test data which is wrongly predicted by the classifier.

Confusion matrix with respect to the test i.e. 20% of the actual given data as shown in figure 9. Twenty rows tested in trained dataset, where true positive and true negative values sums up (16 + 3 = 19) where as no false positive and false negative = 1. Based on confusion matrix the accuracy rate calculated is 0.95.

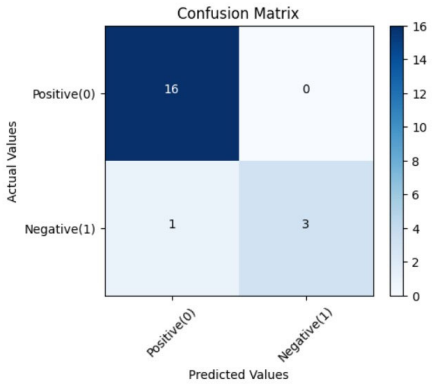

Fig 9. Confusion matrix

6. Conclusion

The rapid growth of the Internet of Things (IoT) has significant implications for our interactions with the environment. Reliable network architectures are crucial for supporting the large number of connected devices. IoT networks are classified into PAN, LAN, and WAN, utilizing communication protocols such as Wi-Fi, Bluetooth, Zigbee, and LoRaWAN. Security challenges include physical security, authentication, encryption, software vulnerabilities, DoS attacks, data privacy, and supply chain security. In this research we focused on anomaly-based security mechanisms and integrating the Random Forest Algorithm to find the anomaly and secure the IoT network. in our implementation we found that our detection accuracy reached 95%. In future, we would like to work on more data and with collection of real time data using ESP 32 dev kit .

References

[1] K. O. M. Salih, T. A. Rashid, D. Radovanovic, and N. Bacanin, "A Comprehensive Survey on the Internet of Things with the Industrial Marketplace," *Sensors*, vol. 22, no. 3, Art. no. 3, Jan. 2022, doi: 10.3390/s22030730.

[2] M. Ayaz, M. Ammad-Uddin, Z. Sharif, A. Mansour, and E.-H. M. Aggoune, "Internet-of-Things (IoT)-

Based Smart Agriculture: Toward Making the Fields Talk," *IEEE Access*, vol. 7, pp. 129551–129583, 2019, doi: 10.1109/ACCESS.2019.2932609.

[3] C. Maple, "Security and privacy in the internet of things," *Journal of Cyber Policy*, vol. 2, no. 2, pp. 155–184, May 2017, doi: 10.1080/23738871.2017.1366536.

[4] M. Wu, T.-J. Lu, F.-Y. Ling, J. Sun, and H.-Y. Du, "Research on the architecture of Internet of Things," in *2010 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE)*, Aug. 2010, pp. V5-484-V5-487. doi: 10.1109/ICACTE.2010.5579493.

[5] M. Saqlain, M. Piao, Y. Shim, and J. Y. Lee, "Framework of an IoT-based Industrial Data Management for Smart Manufacturing," *Journal of Sensor and Actuator Networks*, vol. 8, no. 2, Art. no. 2, Jun. 2019, doi: 10.3390/jsan8020025.

[6] G. Pau, C. Chaudet, D. Zhao, and M. Collotta, "Next Generation Wireless Technologies for Internet of Things," *Sensors*, vol. 18, no. 1, Art. no. 1, Jan. 2018, doi: 10.3390/s18010221.

[7] R. C. Braley, I. C. Gifford, and R. F. Heile, "Wireless personal area networks: an overview of the IEEE P802.15 working group," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 4, no. 1, pp. 26–33, Jan. 2000, doi: 10.1145/360449.360465.

[8] S. Kraijak and P. Tuwanut, "A survey on IoT architectures, protocols, applications, security, privacy, real-world implementation and future trends," in *11th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM 2015)*, Sep. 2015, pp. 1–6. doi: 10.1049/cp.2015.0714.

[9] M. R. Ahmed, A. Al Shihimi, T. Myo, B. Al Baroomi, and M. A. Aseeri, "Internet of Things Network Architecture and Security Challenges," presented at the Second International Conference on Advances in Software Engineering and Information Technology, Mumbai: Hinweis, Jun. 2023.

[10] M. A. Jabraeil Jamali, B. Bahrami, A. Heidari, P. Allahverdizadeh, and F. Norouzi, "IoT Architecture," in *Towards the Internet of Things: Architectures, Security, and Applications*, M. A. Jabraeil Jamali, B. Bahrami, A. Heidari, P. Allahverdizadeh, and F. Norouzi, Eds., in EAI/Springer Innovations in Communication and Computing. Cham: Springer International Publishing, 2020, pp. 9–31. doi: 10.1007/978-3-030-18468-1_2.

[11] C.-L. Zhong, Z. Zhu, and R.-G. Huang, "Study on the IOT Architecture and Gateway Technology," in *2015 14th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*, Aug. 2015, pp. 196–199. doi: 10.1109/DCABES.2015.56.

[12] S. Kumar, P. Tiwari, and M. Zymbler, "Internet of Things is a revolutionary approach for future technology enhancement: a review," *J Big Data*, vol. 6, no. 1, p. 111, Dec. 2019, doi: 10.1186/s40537-019-0268-2.

[13] S. Oza *et al.*, "IoT: The Future for Quality of Services," in *ICCCE 2019*, A. Kumar and S. Mozar, Eds., in Lecture Notes in Electrical Engineering. Singapore: Springer, 2020, pp. 291–301. doi: 10.1007/978-981-13-8715-9_35.

[14] D. Sehrawat and N. S. Gill, "Smart Sensors: Analysis of Different Types of IoT Sensors," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, Apr. 2019, pp. 523–528. doi: 10.1109/ICOEI.2019.8862778.

[15] S. Al-Sarawi, M. Anbar, K. Alieyan, and M. Alzubaidi, "Internet of Things (IoT) communication protocols: Review," in *2017 8th International Conference on Information Technology (ICIT)*, May 2017, pp. 685–690. doi: 10.1109/ICITECH.2017.8079928.

[16] I. Heđi, I. Špeh, and A. Šarabok, "IoT network protocols comparison for the purpose of IoT constrained networks," in *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2017, pp. 501–505. doi: 10.23919/MIPRO.2017.7973477.

[17] K. Shaukat, T. M. Alam, I. A. Hameed, W. A. Khan, N. Abbas, and S. Luo, "A Review on Security Challenges in Internet of Things (IoT)," in *2021 26th International Conference on Automation and Computing (ICAC)*, Sep. 2021, pp. 1–6. doi: 10.23919/ICAC50006.2021.9594183.

[18] K. Kimani, V. Oduol, and K. Langat, "Cyber security challenges for IoT-based smart grid networks," *International Journal of Critical Infrastructure Protection*, vol. 25, pp. 36–49, Jun. 2019, doi: 10.1016/j.ijcip.2019.01.001.

[19] H. Lin and N. W. Bergmann, "IoT Privacy and Security Challenges for Smart Home Environments," *Information*, vol. 7, no. 3, Art. no. 3, Sep. 2016, doi: 10.3390/info7030044.

[20] N. Waheed, X. He, M. Ikram, M. Usman, S. S. Hashmi, and M. Usman, "Security and Privacy in IoT Using Machine Learning and Blockchain: Threats and Countermeasures," *ACM Comput. Surv.*, vol. 53, no. 6, p. 122:1-122:37, Dec. 2020, doi: 10.1145/3417987.

[21] B. K. Mohanta, D. Jena, S. Ramasubbareddy, M. Daneshmand, and A. H. Gandomi, "Addressing Security and Privacy Issues of IoT Using Blockchain Technology," *IEEE Internet of Things Journal*, vol. 8, no. 2, pp. 881–888, Jan. 2021, doi: 10.1109/JIOT.2020.3008906.

[22] F. Hussain, R. Hussain, S. A. Hassan, and E. Hossain, "Machine Learning in IoT Security: Current Solutions

and Future Challenges," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1686–1721, 2020, doi: 10.1109/COMST.2020.2986444.

[23] A. B. Feroz Khan and A. G, "A Multi-layer Security approach for DDoS detection in Internet of Things," *International Journal of Intelligent Unmanned Systems*, vol. 9, no. 3, pp. 178–191, Jan. 2020, doi: 10.1108/IJIUS-06-2019-0029.

[24] F. Nizzi, T. Pecorella, F. Esposito, L. Pierucci, and R. Fantacci, "IoT Security via Address Shuffling: The Easy Way," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3764–3774, Apr. 2019, doi: 10.1109/JIOT.2019.2892003.

[25] R. Ramya Devi and V. Vijaya Chamundeeswari, "Triple DES: Privacy Preserving in Big Data Healthcare," *Int J Parallel Prog*, vol. 48, no. 3, pp. 515–533, Jun. 2020, doi: 10.1007/s10766-018-0592-8.

[26] M. K. Abiodun, J. B. Awotunde, R. O. Ogundokun, E. A. Adeniyi, and M. O. Arowolo, "Security and Information Assurance for IoT-Based Big Data," in *Artificial Intelligence for Cyber Security: Methods, Issues and Possible Horizons or Opportunities*, S. Misra and A. Kumar Tyagi, Eds., in Studies in Computational Intelligence. Cham: Springer International Publishing, 2021, pp. 189–211. doi: 10.1007/978-3-030-72236-4_8.

[27] A. Tewari and B. b. Gupta, "A lightweight mutual authentication protocol based on elliptic curve cryptography for IoT devices," *International Journal of Advanced Intelligence Paradigms*, vol. 9, no. 2–3, pp. 111–121, Jan. 2017, doi: 10.1504/IJAIP.2017.082962.

[28] H. D. Tiwari and J. H. Kim, "Novel Method for DNA-Based Elliptic Curve Cryptography for IoT Devices," *ETRI Journal*, vol. 40, no. 3, pp. 396–409, 2018, doi: 10.4218/etrij.2017-0220.

[29] P. Barman and B. Saha, "DNA Encoded Elliptic Curve Cryptography System for IoT Security." Rochester, NY, Mar. 19, 2019. Accessed: Jun. 18, 2023. [Online]. Available: https://papers.ssrn.com/abstract=3355530

[30] M. Aledhari, R. M. Parizi, A. Dehghantanha, and K.-K. R. Choo, "A Hybrid RSA Algorithm in Support of IoT Greenhouse Applications," in *2019 IEEE International Conference on Industrial Internet (ICII)*, Nov. 2019, pp. 233–240. doi: 10.1109/ICII.2019.00049.

[31] W. Sullivan, *Decision Tree and Random Forest: Machine Learning and Algorithms: The Future Is Here!* CreateSpace Independent Publishing Platform, 2018.

[32] A. Panesar, *Machine Learning and AI for Healthcare: Big Data for Improved Health Outcomes*. Apress, 2019.

[33] C. Strobl, *Statistical Issues in Machine Learning: Towards Reliable Split Selection and Variable Importance Measures*. Cuvillier Verlag, 2008.

[34] M. R. Ahmed, T. Myo, B. Al Baroomi, M. H. Marhaban, M. S. Kaiser, and M. Mahmud, "A Novel Framework to Detect Anomalous Nodes to Secure Wireless Sensor Networks," in *Applied Intelligence and Informatics*, M. Mahmud, C. Ieracitano, M. S. Kaiser, N. Mammone, and F. C. Morabito, Eds., in Communications in Computer and Information Science. Cham: Springer Nature Switzerland, 2022, pp. 499–510. doi: 10.1007/978-3-031-24801-6_35.

[35] S. Wan and H. Yang, "Comparison among Methods of Ensemble Learning," in *2013 International Symposium on Biometrics and Security Technologies*, Jul. 2013, pp. 286–290. doi: 10.1109/ISBAST.2013.50.

[36] M. W. Ahmad, M. Mourshed, and Y. Rezgui, "Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression," *Energy*, vol. 164, pp. 465–474, Dec. 2018, doi: 10.1016/j.energy.2018.08.207.