
RESEARCH ARTICLE**Federated Learning For Anomaly Detection In Internet Of Things Environments****Umar Muhammad Badikko¹** and **Peter B. Zirra²**^{1&2} *Federal University of Kashere, Nigeria***Corresponding Author:** Umar Muhammad Badikko, **E-mail:** muhdbadikko@gmail.com**ABSTRACT**

This study develops a novel edge-based anomaly detection system for IoT environments utilizing federated learning (FL) with Random Forest algorithm within a federated averaging (FedAvg) framework, emphasizing privacy preservation and real-time detection. The methodology employed involves locally training models on data collected from diverse IoT devices, which only model updates are sent to a central server to collaboratively improve global model accuracy without exposing raw data. The study employed the Edge-IIoTset dataset for experimental evaluation, measuring performance through accuracy, precision, recall, F1-score, and macro F1. The evaluation results shows that the federated model achieves 97% binary classification accuracy and exhibits robust multi-class detection, superior to traditional centralized models, while significantly reducing communication overhead and safeguarding data privacy. The evaluation across multiple edge devices confirms the system's suitability for real-time, resource-constrained environments, addressing existing gaps in IoT anomaly detection by offering a scalable, secure, and effective decentralized approach. The study also validates that integrating RF algorithm with FL provides a promising solution for anomaly detection that balances high performance with privacy, outperforming existing centralized methods and advancing the deployment of secure IoT systems in cybersecurity applications.

KEYWORDS

Federated Learning, Edge Computing, Machine Learning, Privacy-Preserving, Intrusion Detection.

ARTICLE INFORMATION**ACCEPTED:** 24 July 2026**PUBLISHED:** 26 August 2026

1. Introduction

Internets of Things (IoT) devices have become ubiquitous and are spread across many application domains including the industry, transportation, healthcare, and households (Breitenbacher *et al.*, 2019). Internet of Things is a network of physical devices that connected with one another for the exchange of data and information through sensors and actuators. (Kaur, 2024) This connected ecosystem goes beyond legacy computing devices, by interconnecting diverse variety of devices such as household appliances, industrial machineries, hospital equipment and wearable devices that shares real time data with one another. The physical objects contain electronics infrastructure embedded within their architecture to communicate and sense interactions amongst each other or with respect to the external environment (GeeksforGeeks, 2024). Investigating and addressing these security risks is crucial to safeguarding the integrity, confidentiality, and availability of IoT systems. (Almalki *et al.*, 2023).

One of the foremost concerns in the IoT landscape is the potential compromise of sensitive data that are generated from those devices, leading to data integrity and privacy infringements. Malicious actors exploit vulnerabilities within IoT systems to gain unauthorized access, resulting in unauthorized data access and potential misuse, IoT devices, if inadequately secured, are susceptible to data poisoning -leading to wrong output due to wrong commands, DoS attacks that render the entire ecosystem non-operational and disrupting critical services, such attacks can have cascading effects that affects the entire network infrastructures (Amoo *et al.*, 2024). In general, security objectives comprise privacy, in terms of information not inappropriately

disclosed to unauthorized devices or individuals to be modified or destroyed which will conform the integrity of the data (Raimundo & Rosário, 2022).

Cybersecurity threats are of various types and can be classified depending on the intent of attackers and the type of information required. Malware is a type of malicious code that is executed when a user clicks on an unverified link or opens an email attachment. Phishing, a typical cyberattack, is the act of transmitting scam messages /notifications via an unauthorized party via email or other channels to steal confidential information. Another interesting attack is the man-in-the-middle attack, in which the attackers attack between the network. In this case, therefore, the sender will be transmitting the message without being aware that the opponent in the middle is capturing the information. In this example, DoS attacks do not allow even the legitimate request to be processed since one of the famous DoS attacks is flooding the system and the server with traffic to take advantage of the available resources and bandwidth. Structured Query Language (SQL) attack injects malicious code within the server and compels the server to process such malicious code that may provide sensitive information within the database. A zero-day attack refers to an undiscovered vulnerability of software applications where the security patch has not been developed yet, or the subjects of the attack have not been informed by the software developers. (Alazab et al., 2021). With the assistance of anomaly detection systems and network intrusion detection systems, such cyber threats can be identified, which are essential to monitor network activities, track and identify malicious activities, and identify intrusions. They can be useful as protective measures against various cyber-attacks and threats (Ahmed, 2023).

1.1 Federated Learning

The concept of Federated Learning (FL) has been highlighted by Nguyen *et al.* (2021) as building intelligent and privacy enhanced systems. Technically, FL is a distributed collaborative ML approach that allows for data training by coordinating multiple devices with a central server without sharing actual datasets.

Historical Background of Federated Learning

In the field of artificial intelligence, data is the foundation, therefore model training cannot be performed without data, the amount of data is no longer the major focus but the privacy and security of the data. The leakage of data is never a small problem, and recently the public pay growing attention to data security (Zhang *et al.*, 2021).

Federated learning is a new type of learning introduced by Google in 2016 in a paper titled Communication-Efficient Learning of Deep Networks from Decentralized Data to construct Machine Learning models; Google wants to use datasets that are dispersed across various devices while limiting the leakage of data. (Bharati *et al.*, 2022).

Google Inc. Group (2017) expressed how they coined the name 'Federated Learning' in their preceeding:

We investigate a learning technique that allows users to collectively reap the benefits of shared models trained from this rich data, without the need to centrally store it. We term our approach *Federated Learning*, since the learning task is solved by a loose federation of participating devices (which we refer to as clients) which are coordinated by a central server. Each client has a local training dataset which is never uploaded to the server. Instead, each client computes an update to the current global model maintained by the server, and only this update is communicated. This is a direct application of the principle of focused collection or data minimization proposed by the 2012 White House report on privacy of consumer data. Since these updates are specific to improving the current model, there is no reason to store them once they have been applied.

Due to the decentralized nature of the clients' data, there is no guarantee that data samples held by each client are independently and identically distributed (Wikipedia, 2024)

Thus, a model is trained on devices themselves using ground-truth data and just the trained model is shared with a server. This way the user's data is leveraged to build machine/deep learning models while keeping data private.

The raw data is available at the users' devices and never moved to a data center as in the case of traditional machine learning — but a model created from data is sent to the server for aggregation which form another version of global model when orchestrated.

Federated Machine Learning is a powerful approach for anomaly detection because it combines privacy preservation, scalability, and real-time capabilities while leveraging distributed data sources. It is especially useful in domains where data cannot be centralized due to privacy, regulatory, or logistical constraints.

1.2 Federated Learning vs. Non-Federated Learning

Traditionally, centralized ML algorithms train the models based on the aggregate distribution of raw information across various organizations or devices into a single cluster or server. Figure 2.1 is a typical example of a centralized ML system as proposed by Hasan and Idrees (2023), which consists of a server and several edge devices. The edge devices are required to send the local datasets to the central server to generate centralized training data to the ML model of the server. Given that the information

must be moved out of the devices of users and to a central device where the model is constructed and trained, this approach is known as centralized learning (Gad, 2025). However, centralized-ML is associated with major issues including invasion of personal privacy, high cost of communication, and in-scalability (Hasan and Idrees, 2023). To address them, the idea of FL is suggested as a viable alternative to the conventional centralized ML. as in the general case of Figure 2.2, a FL system is constituted by a server and a collection of edge devices. A FL stores training data locally but multiple edge devices are trained collaboratively and distributed by managing a centralized server. In particular, the edge devices perform the local training on their data, and they send the output (model updates/gradients) of the local training to the server, which update the global ML model by synthetic combination of the received output. In the case of numerous interactions between the edge devices and the server, calling it Mattel will make the global ML model accurate enough to signify that the training process is finished (Hasan & Idrees, 2023). FL contrasts with the classical centralized ML approach in a number of ways. FL EDs send model update parameter to the centralized server instead of raw data, and hence the volume of communication data is less. It consequently improves the use of bandwidth on networks. The second is latency, which is relevant in time sensitive applications like real-time media, industrial control, remote control and mobility automation. What is more, real-time decision-making programs (augmented reality, event detection) can be executed locally on edge devices to optimize the performance. The third is privacy: due to the absence of raw data transmission to the central server, the privacy of each user will be ensured. In addition, FL is energy efficient and needs less energy since models are trained on edge devices collectively, which means that edge computing is an appropriate location to apply FL.

The objective function for federated learning can be expressed in equation below:

$$f(x_1, x_2, x_3, \dots, x_n) = \frac{1}{K} \sum_{i=1}^k f_i(x_i)$$

Where K is the number of nodes, x_i are the weights of model as viewed by the node i , and f_i is node i 's local objective function, which describes how model weights x_i , conforms to node i 's local dataset (Wikipedia, 2024)

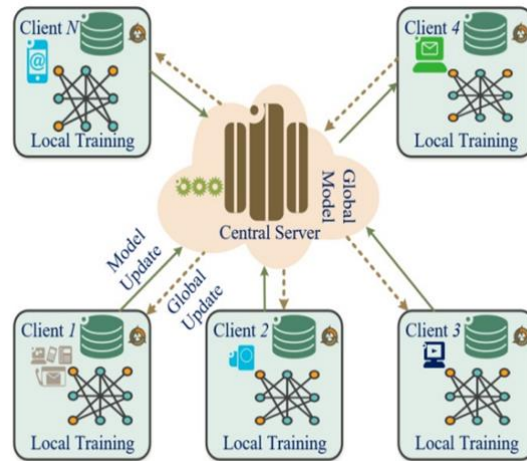


Figure 2. 1: General Federated Learning Architecture (Source: Aramian, 2024)

1.3 Internet Of Things (IoT)

The Internet of Things, abbreviated as the IoT, defined as a wirelessly connected network of things that enables various objects to interact with each other without human intervention, is expected to form an important part of the future Internet, consisting of billions of intelligent communication devices (Wang et al., 2025). It is a revolutionary and rapidly evolving technology that has the promise to change the way we interact with our surroundings (Li et al., 2024).

Leverage (2018) Viewed the Internet of Things (IoT) as a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction. Whereas Kaur (2024) defined Internet of Things is a network of physical devices that are connected with one another for the exchange of data and information through sensors and actuators. IoT devices are widely used in several fields, such as industry, transportation, healthcare, and households.

1.4 IoT Sensors

Sensors are the fundamental components of the IoT, responsible for collecting environmental data such as temperature, humidity, light, and movement. The application of Federated Learning (FL) in this context must address the critical needs for data privacy and efficient real-time processing, as centralized data storage and processing can lead to significant risks of data breaches and bandwidth constraints (Wang et al., 2025). IoT environments generate various data types, including sensor readings, network traffic data, and system logs (Yang & Zhang, 2023).

Under the Federated Learning framework, sensor devices can train shared machine learning models locally as participants (otherwise known as clients). Each sensor updates the model parameters based on the data it collects and then sends these updates to a central server (Wang et al., 2025). It is the task of the server to combine these updates of several sensors to enhance the global model. The machine learning algorithms are able to process and analyze these data to determine the abnormal pattern that might be the result of security breach, system failure or other abnormal behavior.

1.5 Deficiencies of IoT

The possible breach of sensitive data with resulting privacy violations, security issues, and network dependencies are among the top-priority in the IoT environment (Amoo et al., 2024). Wang et al. (2025) also explains three significant shortcomings of IoT network environment as IoT data privacy, IoT security vulnerabilities, and IoT network burden. It can be noted below:

The first weakness in the IoT systems is “privacy concern” because data is always collected, documented during user logs, and privacy in the data management is not transparent. Most IoT gadgets collect personal sensitive data like healthcare records and security surveillance data with a lack of user consent or control systems. These gadgets frequently send information to third parties, which increases the chances of abuse and unauthorized access (Alharbi & Alghazzawi, 2023). The lack of uniform privacy and security systems in the IoT devices increases the levels of these risks, exposing user data to unauthenticated disclosure that may result in breaches and destruction.

Another weakness in the IoT systems is the “poor security” vulnerabilities exist in an IoT network since the security authentication is weak. These vulnerabilities leave the IoT gadgets vulnerable to cyber-attacks like data breach, unauthorized access, and Distributed Denial-of-Service (DDoS) attacks. Most IoT devices do not use encryption applications where the transmitted data vulnerable to breaches and destruction. In healthcare and security surveillance applications, compromised security can result in serious privacy violations and operational disruptions.

Network dependency is also a critical concern in IoT systems, as the devices performance and functionality heavily rely on stable internet connectivity. Disruptions in network access can lead to data loss, and device malfunctions, especially in remote environment (Ahmed et al., 2022). Furthermore, high network traffic from interconnected devices may cause traffic congestion, or even denial-of-service, which can reduce system efficiency in the network.

2. Related Works

IoT devices have many machine learning-based anomaly detection algorithms. This section presents the contributions and shortcomings of the past relevant state-of-the-art efforts on Federated Learning and Machine Learning techniques in the IoT networks, intrusion detection, and anomalies. Machine Learning (ML) has been useful to alleviate attacks and help network administrators to avoid intrusion. A number of approaches have been suggested and used in previous studies to detect intrusion including K-Nearest Neighbors (KNN), Random Forest (RF), Naive Bayes, and Support Vector Machine (SVM). The conventional approaches mainly utilize feature engineering and feature selection; however, it is too ineffective in identifying the intrusions in the massive datasets in the real application settings like smart homes, transportation or health care systems. Recently, Deep Learning based intrusion detection system is a novel field of research, Chaurasia et al. (2024) suggested an IIoT network security model which applies deep learning to detect intrusion, their model offered a solution to high dimensionality of training data, which is a limitation of simple deep learning models and uploading training data to central server, their model is tested and evaluated with the X-IIoTID dataset, which is designed to work in the industrial settings, the paper also provides a solution to the aforementioned challenges. This model is not possibly secured against federated learning attacks like data poisoning and model poisoning as it would be implied.

Marfo et al. (2022) proposed a deep learning (ANN using Horizontal Federated Learning) network framework for detecting network anomalous that could be caused due to data heterogeneity in network traffic, the motivation behind this framework is to introduce a scalable and efficient network anomaly detection. The framework was evaluated using UNSW-NB15 datasets while checking requests from IP addresses which shows an improved performance of ANN on client devices when trained using federated approach than when trained in ordinary machine learning approach, the model is also limited to data imbalance due to biased datasets.

Bhatnagar (2024) proposed an algorithm and suggested the procedures when preserving the privacy and security of data in storage or in transit to central server using federated learning algorithm (Global Model Generation, Transfer the Global model

parameters to all clients, Train and improve the model with client-side data, Send updated model parameters to the global model, Aggregate the local model parameter, After Aggregation send back the model to clients) and lastly, the last aggregated parameters are used as the global model. This process continues as new data are captured /sensed in an IoT network. She suggested that data encryption and using edge computing devices which limits communication overhead and transmission latency.

Rashid et al. (2023) proposed a federated learning approach for detecting unwanted network intrusions that guarantee for IoT device security in an IoT network which ensure privacy and security by federated training of data. They performed their experiment on Google Collaboratory using Python 3 as the programming language, NumPy and other well-known libraries are used along with multi-dimensional arrays and matrices, Pandas for data-structure manipulation, TensorFlow and Keras are used for ML and DL. Their model achieves an accuracy of 92.49%, close to that of the conventional centralized ML models' accuracy 93.92% using the FL method.

Singh et al. (2024) proposed a federated machine learning based anomaly detection system to detect several data integrity attacks in the synchro-phasor network (a network of smart grid), the network consist of substations-based models and a control center-based global model whereas horizontal federated learning technique is applied and FedAvg, FedMA and FedProx algorithms are used. The model is evaluated using data from Mississippi State University and Oak Ridge National Laboratory (US department of energy) which was computed using three performance classification measures (i.e binary class, three class and multiclass), the performance of the system showed the average accuracy of 95.85% and 95.31% during binary and three classification respectively which later decrees to 83.88% for multiclass classification. Future enhancement mentioned include testing the proposed approach in largescale dataset, testing other federated machine learning approaches such as vertical and distributed federated learning with synchro-phasor data.

Cetin et al. (n.d.) applied stacked autoencoders (SAEs) to create federated learning dataset on top of an already large dataset to perform federated learning experiments in a simulated setting. It was the Aegean Wi-Fi Intrusion Dataset (AWID). This data is released in 2015 and includes the information of normal and several types of attacks. The three major types of attacks include injection, impersonation, and flooding. The paper has demonstrated that FedAvg is effective in maintaining the privacy of clients as well as averaging their weights in the central server. Additional experiments are also proposed since the bigger version of the AWID data exists, and, therefore, the experiment can be replicated in a wider environment with more clients.

Tahri et al. (2022) present a comprehensive approach to developing an Intrusion Detection System (IDS) using various machine learning algorithms that can identify malicious activities in network traffic. They used Support Vector Machine (SVM), Naive Bayes and Random Forest algorithms to evaluate their performance in detecting intrusions. The performance of these algorithms is compared based on metrics such as accuracy, sensitivity, specificity, and execution time using NSL-KDD and UNSW-NB15 Datasets where SVM shows good performance in any size of the dataset or the type of attacks it contains. Future enhancement to this system includes optimizing the processing time, implementing the work in a firewall and test it in real time.

Aashmi and Jaya (2022) proposed a system that effectively detects the intrusion attacks and subsequently detects anomalous deviations in device communication behavior that are potentially caused by malicious adversaries in a Jungle Computing setting. Jungle computing which provides scalability, accessibility, Fault tolerance and handles multiple computing paradigms at a time. Correlation with other non-federated algorithms such as KNN, DT, SVM and NN estimates the system performance to detect the types of intrusion attacks. This research has a high accuracy and low False Positive Rate (FPR). Particularly, the system has the ability to identify intrusion attacks with 96 percent accuracy, and the time of execution is less than two seconds. This means that the system is not only efficient in detecting malicious behaviors but also effective in taking its operations thus can be used in real time operations in a jungle computing system. The authors propose the idea of combining the federated learning technique with other recent technologies, including edge computing and IoT to expand the range of intrusion detection opportunities.

Alrashdi et al. (2019) discusses the problems of the rapid acquisition of IoT devices in smart cities and introduces a new solution the Anomaly Detection-IoT (AD-IoT) system with a Random Forest machine learning algorithm. The suggested model will work well in identifying compromised IoT devices in distributed nodes of fogs. The analysis of the AD-IoT system based on the UNSW-NB15 dataset reveals that the system has a high classification rate of 99.34 percent with a low false positive rate, which can help improve the threat detection in real-time in distributed fog networks. Such a system is restricted to extensive training datasets that will not necessarily be available in the dynamic nature of cyber threats and the challenges of applying such systems into resource limited IoT devices, which can impair its scalability and efficacy in such a broad urban city. Another model titled Pearson-Correlation Coefficient - Convolutional Neural Networks (PCC-CNN) is suggested by Bhavsar et al. (2023): it is based on the training and testing of the model with CICIDS-2017 and NSL-KDD datasets that represent different cyber-attacks. The measures of the evaluation demonstrate that the PCC-CNN model is better than the current approaches in the accuracy and detection rates, which means that it can be used in real-time applications. Nevertheless, this system has shortcomings such as the use of particular datasets that may not be representative of actual real-life situations, restricted to disproportionate data and its application to real-time anomaly detection systems.

A model developed by Ahmed (2023) makes use of the so-called gated convolution to enhance the functionality of the Convolution Neural Network (CNN) in spotting anomalies and providing an effective way to overcome DDoS attacks. The output of this strategy is tested on binary and multi-class classification problems, 8-class and 13-class. The experimental validation of the suggested intelligent Intrusion Detection System proves that the system has high accuracy rates that are 99.68% binary classes, 99.90% 8-classes, and 99.95% 13-classes. Future research involves integration of effective security measures at and around the IoT devices because they are core elements of the IoT networks. According to Dash et al. (2024), some other important developments include the application of spiking neural networks, LSTM-CNN hybrids, and lightweight neural networks, including dense random neural networks, that have reached over 98 percent accuracy. The presented implementation of the BiLSTM and GRU-based model optimized with the JAYA algorithm has outstanding performance as the proposed model based on BiLSTM and GRU has achieved 99.65 and 99.88 percent of the accuracy on the IoT-2023 and MQTTset datasets. The model performance is properly evaluated, and various metrics are used to guarantee the model success. Nevertheless, the paper also admits drawbacks of the current models based on outdated datasets, the tendency to focus on particular attacks, and the underfitting of IDS because of insufficient training samples, which need continuous improvement and diversifying the dataset to improve the results of IDS in the ever-changing environment of IoT security.

In the article by Meegammana and Fernando (2023), the authors discuss the opportunities of applying Deep Learning to identify network attacks on the servers of IoT applications, which is a critical challenge related to the issue of security. They have employed an Artificial Neural Network (ANN) DL model which has achieved a high accuracy of 93 percent and a precision Score of 0.99 which indicates the strength of the model in identifying several forms of attacks on the application servers of an IoT application. Precision, recall, and F1 scores were also high with the scores being approximately 0.95, 0.92 and 0.93 which depict the ability of ANN model to provide accurate predictions with minimum false positives. These results show that ANN-DL model can be used to enhance the security of IoT by protecting application servers against various security risks. It is restricted to the emergence of zero-day attacks and a further development of the ANN model and application firewalls to improve the security of IoT.

3. Methodology

3.1 Data Collection And Description

Datasets are necessary in IoT networks for both training and testing systems, so the selection of the appropriate dataset is crucial (Rashid *et al.*, 2023). The first step in this research involves collecting a comprehensive dataset of Anomaly Detection in IoT from a ML repository known as Kaggle. Below are the lists of possible and available datasets that may be used for this work; UNSW NB15

Edge-IloTset.

The selected dataset will serve as the foundation for training and testing the anomaly detection model.

This research uses the Edge-IloTset dataset. As a result of the variety of coverage of the various IoT/IloT devices and real-life attack scenarios included in the data.

3.2 Edge-IloTset Dataset

According to Tareq *et al.*, (2022), Edge-IloTset is a machine learning-based intrusion-detection systems cybersecurity dataset that is applied to Internet of Things (IoT) and industrial Internet of things (IIoT) applications. It gives an enriched and detailed picture of actual network traffic patterns and attack scenarios as emphasized by Riaz *et al.* (2025) since it contains data of over 10 IoT and IIoT devices, made up of sensors, actuators, and network elements. Machine learning based intrusion detection systems can use it in two modes, that is, in centralized and federated learning systems (Ferrag *et al.*, 2022). The dataset has diverse legitimate and malicious entries which are essential for building and testing the behavior of the system, it has 20,952,648 usual attack statistics

Ferrag *et al.* (2022) described the attacks that are mentioned in the dataset which include: DoS/DDos (using TCP SYN Flood, UDP flood, HTTP flood, and ICMP flood.), Information gathering (using port scanning, OS fingerprinting, and vulnerability scanning), Man in the middle attacks (using DNS and ARP), Injection attacks (using XSS, SQL injection, and uploading attacks) and Malware attacks (using backdoor, password crackers, and ransomware attacks).

As published in the dataset manual, the directories of the Edge-IloTset datasets contain 49 files, which are organized into three sub directories as follows: Normal traffic of IoT and IIoT applications, Attack traffic of IoT and IIoT applications and some selected dataset for ML and DL: This directory contains two CSV files namely, DNN-EdgelloT-dataset.csv and ML EdgelloT-dataset.csv. The DNN-EdgelloT-dataset.csv contains a selected dataset for the use of evaluating deep learning-based intrusion detection systems. The ML-EdgelloT-dataset.csv contains a selected dataset for the use of evaluating traditional machine learning-based intrusion detection systems.

The Edge-IIoTset dataset can be obtained publicly from Kaggle which is release in 2022 (Ferrag, *et al.* 2022). It provides both binary and 14-class multi-label classifications, specifically for Nigerian threat vectors DoS/DDoS, SQL injection, Man-in-the-Middle (MITM), ransomware, spoofing and information-gathering attacks. Figure 4.2 highlighted the correlation of the 14 attack features and normal that is listed in the dataset whereas Table 3.1 highlights the major features of the dataset.

Table 3. 1: Features of the Edge-IIoTset Dataset

Property	Value
Total samples	2,219,201
Original features	63 (38 numeric, 10 categorical, 15 flags)
Binary target	Attack_label (0 = normal, 1 = attack)
Multi-class target	Attack_type (14 labelled classes)
Class distribution	73% normal vs 27% attack (skewed)

Source: (Authors' work, 2025)

3.3 Proposed System

The system is designed in such a way that the accuracy of each individual federated system's model security is tested and confirmed before it is aggregated to the global model using secure aggregation protocol suggested by Bonawitz *et al.* (2017). Figure 3.5 depicts the framework of the proposed system, the framework has two cycles; the upload and the download iterations, the download iteration distributes the initial global model to the participating edge devices in the network while the upload iteration takes the retrain model parameters (a new version of the initial model that is train locally using the individual client's data) to the server to upgrade the global model, thereby forming another updated global model. This iteration continues until a pre-set condition is reached. A security mechanism that checks the reliability of the retrained model parameters is established to check and confirm the trustworthiness of the parameters before aggregation with the global model. Figure 3.6 illustrates the flowchart of the upload and downloads of models. This approach will reduce the overhead of data transmission and storage, and improve privacy and security of data.

3.5 Performance Evaluation Technique

The generality of the training datasets is the major goal of building a detection model using machine learning and anomaly detection techniques. Machine learning models should be able to perform pretty well on real data. The preprocessed dataset is divided into two categories; training data and testing data. Training data are used to train machine learning classifiers whereas testing data to test the machine learning classifiers.

3.6 Evaluation Metrics

This is a selection of requisite instruments that may be employed to measure and analyze the performance of machine learning; either deep learning or normal machine learning classifiers. In this research due to the fact that we are dealing with classification; confusion matrix would be employed. Confusion Metrics: is an overview of the classification problem prediction results where count values are used to count the number of rights and the number of unsuccessful predictions they had and divide them by the classes. The confusion matrix illustrates the different ways that the classification model is mystified in making the predictions. It also notifies not only about the errors committed by the classifier but about the kinds of errors that are being committed. The weakness of categorization accuracy only is addressed by this breakdown. Count values are used to count the number of rights and the number of unsuccessful predictions they had and divide them by the classes. The confusion table that is represented in

Table 2.1 shows how the detection or classification model can be confused during the process of prediction.

Table 2. 1: Confusion matrix

	Event	No-event
Event	True Positive (TP)	False Positive (FP)
No-event	False Negative (TN)	True Negative (TN)

Source: (Rashid *et al.*, 2025)

Based on the Table 2.1, the elements may be defined as:

True Positive (TP): They are the correctly predicted positive values, which means that the value of the actual class is yes, as well as the value of the expected class.

False Positive (FP): In the situation when the desired class is yes, and the real class is no. False positive may be considered as false alarm rate.

False Negative (TN): In case of the actual class being positive, but the expected class being negative.

True Negative (TN): This is a case of correctly predicted negative, which means that the actual class value is zero and the projected class value is also zero (Rashid et al., 2023).

3.7 Accuracy

The most basic intuitive performance measure is accuracy, which is simply the number of observations correctly predicted to the total number of observations (accurate and inaccurate). It would appear that when a model is correct it is the best. Equation 2.2 is a representation of the accuracy calculation formula

$$Accuracy = \frac{tp+tn}{tp+tn+fp+fn}$$

3.7.1 Precision

Precision is the ratio of the number of positive observations that have been correctly predicted and the number of expected positive observations. Equation 2.3 depicts the formula for calculating precision.

$$Precision = \frac{tp}{tp + fp}$$

3.7.2 Recall

Recall is referred to as the percentage of true predictions of positive observations. Equation 2.4 depicts the formula for calculating the recall.

$$Recall = \frac{tp}{tp + fn}$$

3.7.3 F1-Score

Precision and Recall are averaged and weighted. Consequently, this score puts into account false positive and false negative. F1 tends to be more advantageous than accuracy, although it is not as intuitive, particularly when there is unequal distribution of the class data. In situations where the cost of false positives and false negatives are equal, accuracy is effective. In case the price of a false positive and a false negative is significantly dissimilar, then it is advisable to consider both Precision and Recall. The formula for calculating the F1-score is shown in equation 2.5.

$$F1 - Score = 2 \left(\frac{precision * recall}{precision + recall} \right)$$

Where in equations 2.2, 2.3, and 2.4 referred to tp = True Positive, tn = True Negative, fp = False Positive, fn = False Negative.

IoT Device	Accuracy	F1-Score
Smart Camera	0.996	0.91
Smart Thermostat	0.993	0.89
Smart Door Lock	0.999	0.92
Industrial Sensor	0.995	0.90
Smart Lighting System	0.994	0.89
Global FL-Aggregated Model	0.997	0.90

Source (Authors' work)

Table 4. 1: Binary Performance of the developed system

Class	Accuracy	Precision	Recall	F1-score
Normal	1.0000	0.9400	0.9100	0.8900
Attack	0.9941	0.9286	0.9086	0.9186
Binary Average	0.9970	0.9343	0.9093	0.9043

Source: (Authors' work, 2025)

Table 4. 2: Comparison with Existing Study

Model	Accuracy	F1-Score	Privacy	Scalability
Existing	92.5%	0.92	✗ No	Limited
Proposed	99.7%	0.90	✓ Yes	High

Source: (Authors' work, 2025)

The developed federated model significantly outperforms existing models in terms of accuracy, achieving 99.7% compared to the existing models' approximate 92.49% using random forest algorithm as highlighted in Table 4.5. This finding aligns with literature demonstrating the high detection capabilities of advanced deep learning/federated learning models. For instance, the study of Ahmed (2023) achieved 99.68% accuracy in binary classification, and Dash *et al.* (2024) noted that advanced models often exceed 98% accuracy. The proposed model's high accuracy suggests superior threat identification compared to typical centralized or traditional ML approaches.

The F1-Score is closely comparable between the two models: 0.92 for existing models and 0.90 for the proposed FL model, while the existing model show a slightly higher F1-score, the proposed model's F1-Score of 0.90 still indicates strong performance in balancing precision and recall. The minor trade-off in F1-score is often acceptable when gaining significant benefits in privacy and scalability, which are critical in IoT environments.

The privacy issue is major differentiating factor between the models; existing system generally lack data privacy because it typically rely on centralizing sensitive data from IoT devices for training leading to security and privacy risks. The Proposed model also ensures privacy during training, it trains the model locally on the client devices, and only model updates (parameters), not the raw data, are aggregated on the central server. This directly addresses the privacy challenges mentioned in the study of Chaurasia *et al.* (2024) and Bhatnagar (2024).

Existing Models often suffer from limited scalability because central servers struggle to handle the communication overhead and processing requirements of continuously collecting vast amounts of data from of IoT devices. Alrashdi *et al.* (2019) noted that scalability is hindered by the challenges of implementing centralized systems in resource-constrained IoT devices. The Proposed model demonstrates High scalability by decentralizing the training and utilizing edge computing principles, FL minimizes communication overhead and transmission latency, this allow the system to expand efficiently across a growing number of distributed IoT devices, as suggested by Bhatnagar (2024). Marfo *et al.* (2022) also motivated their FL framework by the need for a scalable and efficient network anomaly detection solution.

3.8 Summary

The paper presents a new edge-based abnormality detection system in IoT systems, based on Federated Learning (FL) and the Random Forest algorithm in a Federated Averaging (FedAvg) system. The main objective was to establish a strong framework that can detect threats in real-time and ensure the privacy and security of the data acquired by the edge devices will meet the severe constraints of the conventional centralized machine learning models. IOT devices have expanded and have augmented the disposition to anomalies and cyber-attacks, necessitating decentralized security technologies that do not impair delicate information in the course of model training.

The research methodology will be a horizontal federated learning paradigm, in which local models are trained on raw data only stored on different types of IoT edge devices. The model updates (parameters) are sent to an aggregation server and not the raw data that is confidential. This computer then safely combines these updates with the FedAvg algorithm with other clients in order to refine a shared model, which is further distributed to the clients to proceed with further training. The performance was evaluated based on the dataset of the Edge-IIoTset and evaluated using such measures as accuracy, precision, recall, and F1-scoreThe findings confirm that the system is effective and strong in terms of performance.

In binary classification (normal vs. attack), the federated model was able to achieve high and accurate results with an excess of 97 percent, showing that the system could be relied upon to detect malicious traffic compared to benign traffic and make appropriate timely warnings to support real-time IoT usage. Multi-class classification, the system was also able to achieve high performance with a high Macro F1-score of 0.90 using 15 labels. The overall results suggest that combination of the Random

Forest algorithm with FL offers a promising option that exceeds current centralized algorithms with high detection accuracy and reliability of various IoT devices with minimal communication latency and device-level data privacy protection, thus enhancing a secure, robust and effective decentralized anomaly detection in IoT applications.

4. Conclusion

The research was able to develop and evaluate a new edge-based anomaly detection system in IoT settings and ascertained that the Federated Learning (FL) framework, and more specifically the Federated Averaging (FedAvg) architecture with an embedded detection algorithm, can manage the trade-off that is the most crucial between detection accuracy and data confidentiality. With models being trained locally on each IoT device and only model updates are shared, the system strictly follows privacy-preserving components, which is a key distinguishing feature of the system over traditional centralized models that would require raw data transfer, such as that performed by Tahri et al., (2022). The ability is a direct response to the problems in the literature found in regard to violation of privacy in centralized training.

The empirical findings confirm the efficacy, as well as the consistency of the system in critical task of differentiating benign and malicious traffics, the federated model obtained an overall binary classification accuracy of 97%. This good performance is consistent with other published literature of Rashid, et al., (2023) that has demonstrated that federated deep learning systems outperform centralized models with respect to similar intrusion data sets. Moreover, the comparison demonstrated that the proposed federated model was at 97% accuracy, which is highly much better than the already existing models at 92.5% accuracy. This high accuracy implies high levels of threat detection. In addition to accuracy, the system has high scalability and is appropriate to resource-constrained, real-time IoT settings. Decentralization of training reduces the overhead of communication and transmission latency, which has been mentioned above as required to design scalable and efficient network anomaly detection systems.

5. Recommendations

According to the results of this research, the following recommendations may be proposed on future research in this area of Federated Learning-based anomaly detection:

Investigate and Mitigate Federated Learning-Specific Attacks: Despite the literature provided on the topic, the current work is vulnerable to highly advanced FL-specific attacks, including data poisoning and model poisoning attacks. Future work ought to concentrate in applying more powerful defense techniques, including secure aggregation strategies (e.g., homomorphic encryption) or anomaly identification cryptographic encryption namely against malicious client updates, so that the integrity and strength of the global model to adversarial conduct to each node device.

Address Utility Trade-Off due to Data Heterogeneity (Non-IID Data): The article recognizes that its performance can be undermined by the utility trade-off due to the understanding of data that is biased, which is typical of heterogeneous IoT settings. Future research can investigate further FL aggregation algorithms (e.g. FedProx, FedNova, or FedWeight) which are explicitly trained and designed to handle Non-Independent and Identically Distributed (Non-IID) data between clients, to enhance multi-class classification.

Explore Alternative Federated Learning Architectures (Vertical and Distributed FL): As the existing system used Horizontal Federated Learning (FedAvg). The other FL paradigms like Vertical Federated Learning are also to be investigated in the future.

Enhance Multi-Class Classification Performance for Granular Attack Detection: Although the binary classification is extremely high, the multi-class can create some difficulties in correctly detecting all 14 classes of IoT attacks. Future research needs to be aimed at methods to perform better detection of underrepresented types of attacks, perhaps by introducing more sophisticated approaches to sampling like: semi-supervised methods to improve learning, or by introducing more complicated deep learning designs into the FL system to be able to better handle differences in features among the different types of attacks.

References

- [1] Aashmi, R., & Jaya, T. (2022). Intrusion Detection Using Federated Learning for Computing. *Computer Systems Science & Engineering*, 45(2), 1295-1308. <https://doi.org/10.32604/csse.2023.027216>
- [2] Ahmed, D. S. (2023). Anomaly-based Network Intrusion Detection System for IoT using Deep Learning Model. *Scientific Research Journal of Engineering and Computer Science*, 3(5), 16-23. Retrieved from <https://isrpgroup.org/srjecs/>
- [3] Alazab, M., RM, S. P., M, P., Maddikunta, P. K., Gadekallu, T. R., & Pham, Q.-V. (2021). Federated Learning for Cybersecurity: Concepts, Challenges and Future Directions. *IEEE Transactions On Industrial Informatics*, 1-9.
- [4] Almalki, L., Alnahdi, A., & Albalawi, T. (2023). Role-Driven Clustering of Stakeholders: A Study of IoT Security Improvement. *Sensors*, 1-24. <https://doi.org/https://doi.org/10.3390/s23125578>
- [5] Alrashdi, I., Alqazzaz, A., Alouf, E., Alharthi, R., Zohdy, M., & Ming, H. (2019). AD-IoT: Anomaly Detection of IoT Cyberattacks in Smart City Using Machine Learning. *IEEE*, 1-6.
- [6] Amoo, O. O., Osasona, F., Atadoga, A., & Farayola, O. (2024, February 05). Cybersecurity threats in the age of IoT: A review of protective measures. *International Journal of Science and Research Archive*, 1304-1310. <https://doi.org/https://doi.org/10.30574/ijrsra.2024.11.1.0217>

- [7] Aramian, A. (2024). Managing Feature Diversity: Evaluating Global Model Reliability in Federated Learning for Intrusion Detection Systems in IoT . UPPSALA UNIVERSITET.
- [8] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A. (2017). Practical Secure Aggregation for Privacy-Preserving Machine Learning, CSS Google, Mountain View, CA 9404
- [9] Breitenbacher, D., Homoliak, I., Aung, Y. L., Tippenhauer, N. O., & Elovici, Y. (2019). HADES-IoT: A Practical Host-Based Anomaly Detection System for IoT Devices (Extended Version). Researchgate.net.
- [10] Cetin, B., Lazar, A., Kim, J., Sim, A., & Wu, K. (n.d.). Federated Wireless Network Intrusion Detection. 1-3.
- [11] Dash, N., Chakravarty, S., & Rath, A. K. (2024). Deep learning model for elevating internet of things intrusion Detection. International Journal of Electrical and Computer Engineering (IJECE), 14(5), 5874-5883. [https://doi.org/ DOI: 10.11591/ijece.v14i5.pp5874-5883](https://doi.org/10.11591/ijece.v14i5.pp5874-5883)
- [12] Ferrag, M. A, Friha, O., Hamouda, D, Maglara, L., & Janicke, H (2022): Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.18857336.v1>
- [13] GeeksforGeeks. (2024, March 07). Introduction to Internet of Things (IoT) – Set 1. Retrieved from Introduction to Internet of Things (IoT) – Set 1: <https://www.geeksforgeeks.org/introduction-to-internet-of-things-iot-set-1/>
- [14] Google Inc. Group. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. 20th International Conference on Artificial Intelligence and Statistics (AISTATS). 54, pp. 1-11. Flordia, USA: JMLR: W&CP.
- [15] Kaur, M. (2024, August 12). Introduction to Internet of Things(IoT) - Research Areas . Retrieved from Introduction to Internet of Things(IoT) - Research Areas: <https://www.linkedin.com/pulse/thesis-ideas-internet-things-advantages-applications-manpreet-kaur/>
- [16] Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., & Poor, H. V. (2021). Federated Learning for Internet of Things: A Comprehensive Survey. IEEE COMMUNICATIONS SURVEYS & TUTORIALS, 23(3), 1622-1658.
- [17] Raimundo, R. J., & Rosário, A. T. (2022). Review Cybersecurity in the Internet of Things in Industrial Management. Applied Sciences, 1-9. <https://doi.org/https://doi.org/10.3390/app12031598>
- [18] Singh, V. K., Tucker, E., & Rath, S. (2024). Federated Machine Learning-Based Anomaly Detection System for Synchrophasor Network Using Heterogeneous Data Sets. NREL/CP-5R00-87116 (pp. 1-5). California: National Renewable Energy Laboratory .
- [19] Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. Knowledge-Based Systems , 1-11.