Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 07 Issue 01
2024

Research Article
https://doi.org/10.56979/701/2024

Securing IoT: Balancing Privacy and Attack Prediction

Zainab[1], Muhammad Azam[1], Gohar Mumtaz[1*], and Zeeshan Mubeen[2]

[1]Faculty of Computer Science and Information Technology, Superior University, Lahore, 54000, Pakistan.
[2]Riphah International University, Lahore, 54000, Pakistan.
[*]Corresponding Author: Gohar Mumtaz. Email: gohar.m@superior.edu.pk

Received: January 13, 2024 **Accepted:** May 11, 2024 **Published:** June 01, 2024

Abstract: The huge network of Connected devices that exchange and gather data is known as the Internet of Things (IoT). But this connectedness also creates vulnerabilities, opening up IoT networks to hackers that might steal information, interfere with operations, or even be physically harmful. Our study suggests a novel method that preserve user privacy in IoT network threat prediction by utilizing federated machine learning. Federated learning mitigates privacy problems by enabling models to be trained on dispersed devices without directly sharing sensitive data. The suggested PPIOTN model makes use of the CIC-IOT-2023 dataset, that was created especially for studies on IoT security. Through the use of federated machine learning with differential privacy to train a model on this dataset, the study seeks to secure user privacy while achieving precise cyberattack prediction. Furthermore, proposed PPIOTN architecture's results are compared with other approaches. Finally, the research is concluded based on tuning the differential privacy parameters and obtaining the satisfied results.

Keywords: Privacy Preservation; Cyber-Attacks; IoT Networks; Federated Machine Learning; CIC-IoT-2023 dataset.

1. Introduction

The landscape of the Internet of Things (IoT) is expanding at an exponential rate [1]. The fig 1 shows a steady increase in the number of active connections. This pattern is a reflection of a larger phenomenon the growing quantity of IoT devices [2]. The number of connected devices is predicted to rise in the upcoming years due to factors like more affordable sensors, easily available cloud storage, and the increasing need for data [3].

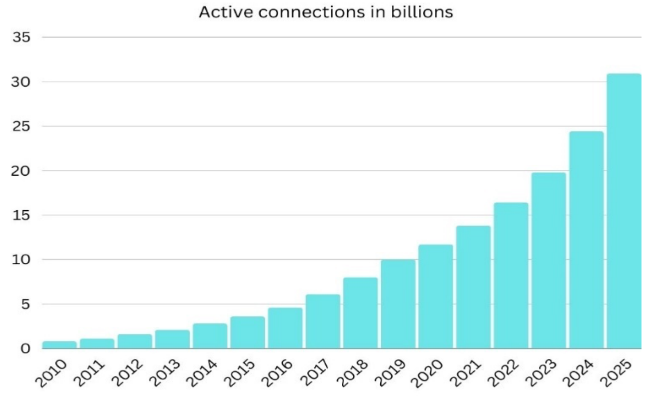


Figure 1. No of IOT Devices

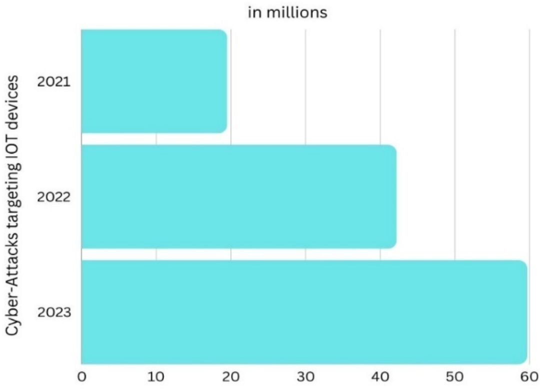


Figure 2. No of Cyber-Attacks

Regretfully, there is an increased susceptibility to cyberattacks associated with this expansion [4]. as shown in the fig 2 These cyberattacks have the potential to steal confidential information, interfere with operations, or even result in physical harm [5]. Strong security measures are required to protect this extensive network [6]. In addition to the wonderful potential that come with the exponential increase in connected devices [7].

Finding a balance between effectively anticipating cyberattacks and protecting user privacy in IoT networks is the primary research challenge [8]. The study suggests that by decentralizing the model training procedure, a unique strategy based on federated machine learning can accomplish this delicate balance. In order to address the combined difficulties of cybersecurity and privacy preservation in the Internet of Things sector, federated learning is introduced as a cutting-edge technique in a brief explanation of the theoretical and conceptual basis of the project [9] [10].

Researchers and policymakers active in influencing the direction of IoT security, along with users, companies, and industries utilizing IoT technologies, comprise the target demographic of interest [11]. One shortcoming of earlier research is the scant investigation of privacy-preserving techniques designed especially for IoT network predictive models [12]. Current methods frequently give priority to accuracy over user privacy; therefore, a thorough and well-rounded solution is required [13].

The study's research is valuable because it offers a novel approach to dealing with the urgent issue of cyberattacks in the context of the Internet of Things [14]. Beyond theoretical ramifications, the value offers real-world applications that complement the dynamic nature of IoT technologies. The suggested method not only closes a significant research gap but also advances the creation of safe and private-preserving products that can be easily included into practical IoT applications as the integration of IoT grows more widespread [15]. The main contributions are;

1. Integration of Differential Privacy with Federated Learning
2. Tuning Differential Privacy parameters to get satisfied results

This research is organized and divided into sections. Literature review section covers the related works, Methodology section further describes the methodological approach. Then the Results and discussion section presents comparative analysis on the results. Finally, last Section summarizes the article and provide the conclusion of this study.

2. Literature Review

In this section we will discuss about the background and related work of our proposed research integration of this related work is shown in Table 1.

SIDRA ABBAS Introduces a federated learning approach for detecting cyberattacks in IoT devices. This method ensures security and privacy by training models locally without sharing raw data. The approach achieved high accuracy of 89.00% in attack detection. Additionally, federated learning enhances scalability and diversity by allowing multiple clients to contribute to model training [16]. MOUSA'B MOHAMMAD SHTAYAT et al utilized an ensemble DL-based IDS incorporating SHAP and LIME methods for intrusion detection in IIoT networks. This approach enhances security and privacy by providing transparent explanations for decision-making. The models achieved high accuracy rates, showcasing robust performance. The methodology demonstrates scalability and diversity in addressing various types of attacks in IIoT environments [17]. Mounia Hamidouche et al discusses about a Deep Learning (DL) model, specifically an learning approach, for device-type and attack detection identification in IoT networks. This approach enhances security by autonomously extracting features, improving accuracy and scalability. It addresses privacy concerns by reducing the need for manual feature engineering. The model's adaptability ensures high accuracy, while its scalability allows for diverse IoT configurations to be effectively analyzed [18]. Hoor Fatima et al utilized an Extra Tree classifier for feature selection and two classifiers, LSTM and 1D-CNN, for intrusion detection. The proposed model focuses on security and privacy in IoT networks, achieving high accuracy of 90.87%. It demonstrates scalability by handling a dataset with 34 attack classes and emphasizes diversity through the use of ensemble techniques [19]. Adrian Pirtama et al presents a Random Forest algorithm for attack detection on the Internet of Things. This algorithm is known for its security features, ensuring robust detection of suspicious activities. Additionally, it prioritizes privacy by effectively classifying data without compromising sensitive information. With high accuracy rates reaching up to 92.2%, it demonstrates exceptional performance in identifying attacks. Moreover, its scalability and diversity make it suitable for handling various types of attacks in complex IoT networks [20]. Mohammed Alosaimi et al addresses a various anomaly detection algorithm such as Markov Chain-based models, LSTM Auto encoders, and Random Forest. These algorithms aim to enhance security and privacy by detecting abnormal behaviors in IoT networks. They also focus on improving accuracy and scalability while ensuring diversity in detecting different types of anomalies effectively [21]. Hadeel Q. Gheni et al utilized the Gaining-Sharing Knowledge (GSK) optimization algorithm for intrusion detection. This algorithm enhances security by identifying intrusion information effectively. It prioritizes privacy by reducing data dimensionality without losing important information. The model demonstrates high accuracy in attack detection, offering improved network security trust. Additionally, the algorithm's scalability and diversity contribute to its efficiency in handling various intrusion scenarios [22]. Denis Parfenov et al presents a Random Forest, Catboost, XGBoost, and MLP-Prod. These algorithms were evaluated for security, privacy, accuracy, and scalability in detecting attacks in IoT networks. The study focused on developing robust machine learning models to improve cybersecurity. The results demonstrated high accuracy and effectiveness in detecting adversarial attacks. Further research is needed to enhance the diversity and reliability of detection methods [23]. Andy Reed et al addresses a systematic packet sampling algorithm for efficient Slow DoS detection in IoT networks. This approach enhances security by reducing computational overheads. It maintains privacy by selectively filtering packets with zero-byte payloads. The accuracy is upheld through a balanced dataset post-sampling. The scalability and diversity of the algorithm support resource-constrained IoT environments effectively [24]. FARIS ALASMARY et al utilized ShieldRNN, a novel training and prediction approach for RNN/LSTM models. ShieldRNN offers enhanced security and privacy for IoT devices, ensuring accurate DDoS attack detection. It demonstrates high accuracy in detecting anomalies and is scalable to handle diverse IoT network traffic effectively [25]. jawad ahmed et al presents Feed Forward Neural Networks (FFNN), Long Short-Term Memory (LSTM), and Random Neural Networks (RandNN) algorithms. These algorithms enhance security, privacy, accuracy, and scalability in IoT networks. They offer diverse capabilities for detecting intrusions effectively [26]. Diego Abreu et al introduces a combination of Ensemble Learning, Deep Learning (LSTMs), and Stream Machine Learning (HAT) algorithms. These algorithms enhance security by detecting and classifying IoT network attacks with high accuracy. They also prioritize privacy by processing data in real-time streams. The system demonstrates high accuracy and precision above 90%, showcasing scalability and diversity in handling various attack types [27]. Nevetha Govindaraju et al utilized Decision Trees, K-Nearest Neighbors, Support Vector Machines, Local Outlier Factor, Isolation Forest, and One-class SVM algorithms. These algorithms offer a balance of security, privacy, accuracy, and scalability in detecting IoT attacks. Their diverse capabilities contribute to robust threat detection and prediction [28]. Nazia Butt et al proposed a hybrid model combining KNN, DT, and LSTM algorithms for intrusion detection in IoT-based smart homes. This approach addresses security vulnerabilities, enhances privacy, improves accuracy, and ensures scalability. The diverse combination of machine learning and deep learning techniques results in robust security measures for smart home networks [29]. Kahraman Kostas et al introduces a SVM algorithm that addresses a Generalizable Models for Behavior-Based IoT Attack Detection and achieving a high accuracy building a generalizable ML model for attack detection. The accuracy achieved varies from 87.97 across different machine learning algorithms lack of focus on demonstrating the generalizability of the proposed approach for IoT network attacks the diverse combination of machine learning [30].

Table 1. Related Work

References	**[16]**	**[17]**	**[18]**	**[19]**	**[20]**	**[21]**	**[22]**	**[23]**	**[24]**	**[25]**	**[26]**	**[27]**	**[28]**	**[29]**	**[30]**	**Proposed PPIoTN**
Security	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Scalability	✕	✕	✓	✓	✓	✓	✓	✓	✓	✕	✓	✓	✕	✓	✓	✓
Privacy	✓	✓	✕	✓	✓	✕	✓	✓	✓	✓	✓	✕	✓	✓	✓	✓
Accuracy	✓	✓	✓	✓	✓	✓	✕	✓	✕	✓	✓	✓	✓	✓	✕	✓
Diversity	✓	✕	✓	✕	✓	✓	✓	✕	✓	✓	✓	✓	✕	✕	✓	✓

3. Methodology

This section, we discuss our methodology and present our suggested PP-IOTN, as shown in Fig. 3, which is a privacy-preserving IOT network. Firstly, we describe how data was interpreted, study as follows. Following that, we look at the techniques of classification and analysis, which include pre-processing, Hyperparameter, and simple DP and advanced DP respectively.

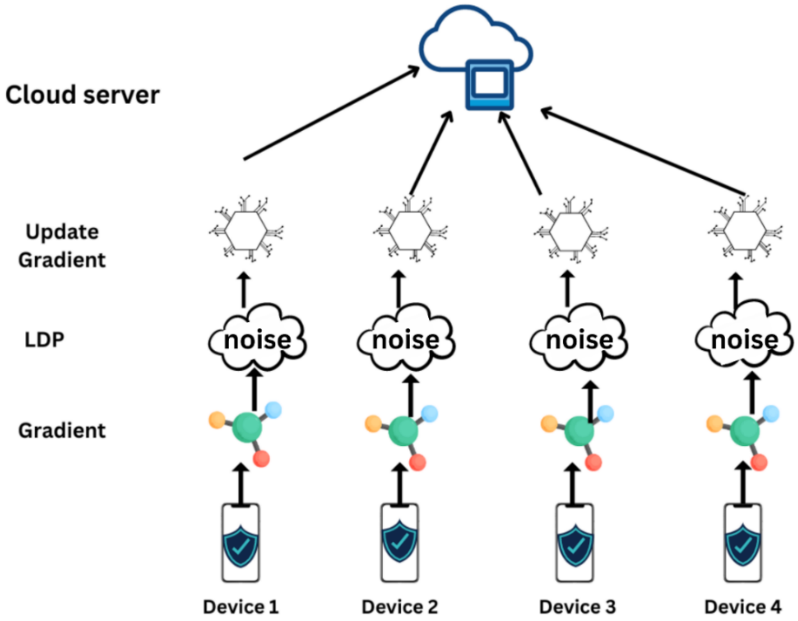


Figure 3. Illustration proposed PP-IOTN architecture.

3.1. Data Interpretation:

Our approach of this methodology includes analysis. the CIC-IOT-2023 dataset are commonly used dataset Table 2 shows the seven categories into which these attacks are divided: DDoS, DoS, Recon, Web-based, Brute Force, Spoofing, and Mirai. Lastly, all attacks are carried out by malicious IoT devices that are directed towards other IoT devices. They gather information from 33 attacks on IoT devices, broken down into 7 types, and show how they may be replicated to identify and categorize IoT network traffic as either benign or harmful.

Table 2. Data Interpretation

Main class	Sub class	Actual	Used
	ACK Fragmentation	285,104	14499
	ICMP Flood	7,200,504	364558
	HTTP Flood	28,790	1468
	ICMP Fragmentation	452,489	22891
	RSTFIN Flood	4,045,285	0
DDOS	Synonymous Flood	3,598,138	182095
	TCP Flood	4,497,667	47777
	UDP Fragmentation	286,925	0
	SYN Flood	4,059,190	206147
	PSHACK Flood	4,094,755	207972
	UDP Flood	5,412,287	0
	SlowLoris	23,426	1177
	SYN Flood	2,028,834	102050
DOS	HTTP Flood	71,864	3702
	TCP Flood	2,671,445	135181

	UDP Flood	3,318,595	168754
Recon	Ping Sweep	2262	2262
	Vulnerability Scan	37,382	37,382
	Host Discovery	134,378	134,378
	Port Scan	82,284	82,284
	OS Scan	98,259	98,259
Web-Based	Sql Injection	5245	5245
	Uploading Attack	1252	1252
	Command Injection	5409	5409
	Backdoor Malware	3218	3218
	Browser Hijacking	5859	5859
	XSS	3846	3846
Brute Force	Dictionary Brute Force	13,064	13,064
Spoofing	DNS Spoofing	178,911	178,911
	Arp Spoofing	307,593	307,593
Mirai	UDPPlain	890,576	890,576
	Greeth Flood	991,866	991,866
	GREIP Flood	751,682	751,682

3.2. Research Motivation

As the number of IoT devices increases, so does the frequency of cyberattacks [31]. This growth can be linked to the growing attack surface created by the proliferation of IOT devices. Hackers take advantage of flaws in these devices, which are generally the result of insufficient security measures or weak passwords. The increasing interconnection of IoT devices is a serious cybersecurity challenge, requiring more awareness and proactive actions to prevent possible threats.

Our main motivation aims study is to address the important requirement for privacy preservation while predicting cyber-attacks on IoT networks. With the fast expansion of IoT devices, cyber-attacks are becoming more frequent and complicated, posing substantial challenges to cybersecurity. The interconnectedness of IoT networks increases the attack surface, making them vulnerable to exploitation by hackers. Given these problems, it is critical to create proactive ways to protect user privacy while maintaining the accuracy of cyber-attack prediction models. Using privacy-preserving approaches such as differential privacy, this study aims to provide a safe framework for detecting cyber-attacks in IoT networks while protecting user privacy rights. This research aims how privacy-preserving approaches affect cyber-attack prediction accuracy and efficacy, with the goal of developing privacy-aware cybersecurity solutions for IoT settings.

3.3. Method Of Analysis And Classification

3.3.1. Preprocessing

In the preprocessing part of our research, we start with the CIC-IOT-2023 dataset, which consists of 169 files that are systematically transformed into 17 separate groups [32]. Each of these categories contains statistics on seven types of cyber-attacks: DDoS, DoS, Reconnaissance, Web-based attacks, Brute Force, Spoofing, and Mirai. Our goal is to detect and isolate the occurrences of these common attack types throughout the dataset. Following that, we begin the process of combining all relevant information into a single file, synthesizing the aggregate occurrences of the aforementioned attacks This aggregation method speeds the subsequent analysis, allowing for a more concentrated assessment of the dataset's frequent attack vectors [33]. By collecting the data into a single file, we assure clarity and consistency in the following stages of our research, allowing for a thorough examination of the detected attack patterns. This careful preprocessing methodology strengthens our inquiry into privacy-preserving strategies for anticipating cyber-attacks in IoT networks, ensuring the reliability and integrity of our findings.

3.3.2. Hyperparameters

This study series of experiments investigated the effect of Hyperparameters on privacy-preserving federated learning's (FL) ability to anticipate cyberattacks in Internet of Things networks. Using the CIC-IOT-2023 dataset. The research concentrated on varying the noise and delta parameters using (non-IID) data. Our studies had 10 number of clients, having an output size 10 with a batch size 64. Five local iterations (E), a fixed epsilon value of 0.5, and a sample rate (Q) of 0.1 were employed in the model. The trials evaluated algorithm convergence over 50 rounds. The model's output was subjected to different delta values of 1e-5 and 1e-3 in addition to the Laplace noise values of 1.2, 2.0, 4.0, and 8.0. Through analysis of these Hyperparameters, the purpose of the study was to comprehend how they fit into the FL framework for predicting cyberattacks in Internet of Things networks while maintaining privacy and accuracy as shown in Table 3.

Table 3. Hyperparameters

Hyperparameters	**Traditional Federated Learning CIC-IOT-2023(IID And NON-IID)**	**FL with DP CIC-IOT-2023 (NON-IID)**
Batch size	32	64
Number Of Clients	100	10
Epoch (E)	5	5
Output size	-	10
Learning rate	0.1	0.01
Epsilon (Ɛ)	-	0.5
Noise	-	1.2, 2, 4, and 8
Q (sample rate)	-	0.1
Delta (¥)	-	0.00001, 0.001
Aggregation Method	FedAvg	FedAvg
Number of rounds	50	150
Model	MLP	MLP
Optimizer	SDG	SDG

3.3.3. Simple Differential Privacy & Advanced Differential Privacy

Differential Privacy is a strategy, maintaining privacy, characterized by Ɛ and ¥. In simple Differential Privacy, only Ɛ is used that prohibits any leaking. In Advanced differential privacy, the value of ¥ is chosen to be smaller than the dataset's inverse size [35]. The CIC-IOT-2023 dataset, which has 6627142 total instances—4639000 for training and 1988142 for testing—is used for this suggested approach. The dataset was split up into 10 distinct clients. We added Laplace noise to the dataset's average value, Xi, in order to preserve privacy. Laplace noise was taken into using the following equation.

$$Z = A + \mathrm{LAP}(s) \tag{1}$$

where A represents dataset's average height and s = 1/ Ɛ. Laplace distribution's random variable with scale s is denoted by LAP(s). The value of ¥ should be smaller than the dataset's inverse size in order to avoid privacy concerns [36]. The probability that any given data point Xi will be disclosed is provided by

$$P\left[\,|Z - Xi| \leq t\right] = 2 * \exp(-\mathcal{E} * t) \leq ¥ \tag{2}$$

The equation 2 show this is lesser than or equal to ¥. demonstrate this, we may compute the natural logarithm (ln) of both sides, and finally isolating t:

$$t \leq (\ln(¥) - \ln(2)) / \mathcal{E} \tag{3}$$

This demonstrates that according to Eq. 2, if a single data point Xi satisfies the condition in equation 3, the risk of it being leaked is limited by the desired privacy parameter ¥, where ¥ is the desired level of privacy and t is the sensitivity of the data. In many tests (1.2, 2, 4, and 8), the Laplace noise applied to the average height can be varied, but Ɛ remains constant at 0.5. Furthermore, le-5 and le-3 may have different

¥ values. When Laplace noise is introduced to the average height within the ±s range, specific data points are less likely to be revealed [42].

3.3.4. Classification

In this study, we used the Privacy-Preserving IoT Networks (PP-IOTN) approach to anticipate cyber-attack in the CIC-IOT-2023 dataset. We used Multilayer Perceptron (MLP) models to train our prediction models on this dataset. To enhance privacy and security during model training, we added Laplace noise to model updates, as stated in [37-38]. We used the Stochastic Gradient Descent (SGD) optimizer and the FedAvg aggregation technique, with particular Hyperparameter listed in Table 3. Our findings demonstrate the effectiveness of the PP-IOTN technique in improving security and privacy while retaining accuracy of model. Furthermore, using the FedAvg aggregation strategy resulted in enhanced model convergence. These findings demonstrate the usefulness of our proposed PP-IOTN architecture for training distributed models that priorities privacy preservation. This study contributes to the development of privacy-preserving strategies for predicting cyber-attacks in IoT networks. It highlights the feasibility and advantages of our approach in this vital sector.

3.3.5. Privacy Analysis

This evaluation exhibits Preserving privacy in Federated Learning. We used DP by adjusting the delta and noise levels while maintaining the epsilon constant. Differential Privacy is a PP strategy that adds random noise to the data to safeguard secure information, guaranteeing statistic of analysis remains accurate without compromising the user data privacy points [39]. The delta and noise levels are adjusted to establish a compromise between privacy and accuracy. To guarantee optimal PP, we employed statistics metrics to assess the degree of delta values and loss of a privacy and ensured that adjusted correctly to retain the intended privacy level. Overall, the study centered on establishing effective PP approaches for Federated Learning, allowing the sensitive analysis material while opposing the specific user privacy.

4. Results and Discussion

In this part, we describe the experimental result from our research on privacy preservation while forecasting cyber-attack in IoT networks using the MLP model. Our evaluation includes both quantitative and qualitative analyses to determine the efficacy of our strategy [40] [43]. We use quantitative measurements like accuracy and loss to analyses the model's performance over a range of characteristics such as client count, dataset size, and training round count. This investigation provides insights into the best circumstances for using privacy-preserving federated learning approaches in IoT network security. We assess the usefulness of the MLP model in answering our study objectives qualitatively, with an emphasis on its capacity to preserve privacy while detecting cyber-attack. In addition, we compare our results to LDP-Fed, a comparable strategy that employs Local Differential Privacy (LDP) approaches. By contrasting our results, we learn about the relative efficiency of various privacy-preserving strategies in IoT network security. Our experimental results show that the MLP model effectively preserves privacy and predicts cyber-attack. This provides useful insights for deploying privacy-preserving federated learning in real-world IoT network contexts.

4.1. Simple Federated Learning

In the first part of our study on Simple Federated Learning with the CIC-IOT-2023 dataset, we used a (MLP) model and tested it on both non-IID and IID datasets [41] [44]. The findings, shown in Figs. 4 and 5, demonstrate the model's performance in both cases, with the IID option outperforming the non-IID configuration. Specifically, Fig. 4 shows that the IID dataset achieves approximately 98% accuracy, whereas the non-IID dataset achieves just 90.5%. Our method emphasizes the iterative process of model training, in which clients do local calculations on their particular datasets and subsequently exchange updated central server with model parameters. This server collects the parameters and generates a new set for the next training round. Furthermore, Fig. 5 displays the analysis of loss rate, which shows that the IID dataset has a lower loss rate due to better preprocessing. These results demonstrate the effectiveness of privacy preservation approaches in our study's setting, providing the framework for better cyber-attack prediction while preserving user privacy in IoT networks.

4.2. Federated Learning with Differential Privacy

In the second part of our study into federated learning with differential privacy for privacy preservation using the CIC-IOT-2023 dataset. Using a non-IID dataset with defined characteristics, our evaluation sought to determine the effective way of privacy preservation by adjusting delta and a noise value while maintaining a 0.5 as the constant epsilon. Our goal was to find the best combination of noise-delta that balances privacy and model accuracy.

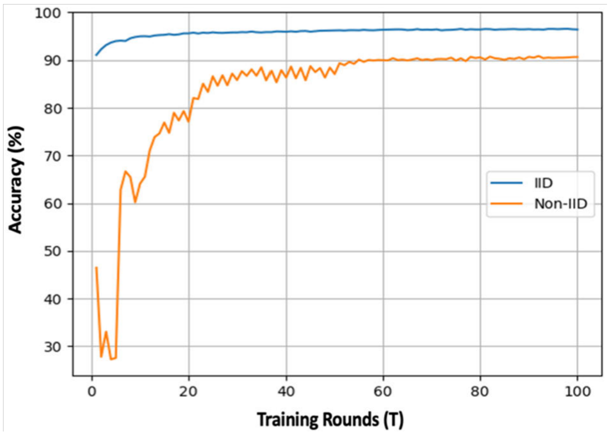


Figure 4. Test set accuracy (CIC-IOT-2023 non-iid and iid) setting with Federated Learning.

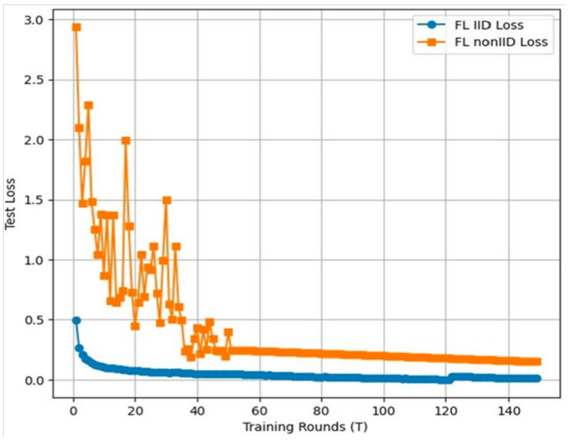


Figure 5. Loss rate Test set (CIC-IOT-2023 non-iid and iid) setting with Federated Learning

We assessed for example privacy outside of the federated learning framework, concentrating on non-IID examples as shown in the Table 4 The multi-layer perceptron (MLP) model was trained over a numerous epoch until it reached the current maximum bound of the (Ɛ, ¥)-differential privacy guarantees. Training used 64 batch sizes, 5 clipping thresholds, and 0.01 learning rates, with runs ending at a minimum acceptable ¥ value of 1e-5. Our results maintained Ɛ (epsilon) values of 0.5 throughout 150 rounds, regardless of noise level. For example, with a level of noise of 1.2, accuracy was 92.3% with a 3.12 epsilon value, but the 8.0 greatest level of noise, accuracy dropped to 78.6%. Despite noise changes, epsilon values for differential privacy parameters remained constant at 0.5, showing the delicate balance between PP and accuracy in our technique for predicting cyber-attacks in IoT networks.

Table 4. FL with differential privacy (Delta = 0.00001).

Variables				
Privacy	3.12	1.92	1.02	0.75
Noise	1.2	2.0	4.0	8.0
Accuracy	92.3	89.0	86.0	78.6
Rounds	150	150	150	150
Epsilon	0.5	0.5	0.5	0.5

4.3. Federated Learning with Tuned Differential Privacy (TDP)

In the third part of our study, federated learning with tuned differential privacy for investigated privacy preserving cyber-attacks in IOT network (PP-IOTN) as a way to handle privacy issues while predicting cyber-attacks in IoT networks, using the CIC-IOT-2023 dataset. This phase maintained the noise and other parameters from the previous phase while adjusting the value of delta.

4.3.1. Quantifying Privacy and Individual Data Disclosure

We evaluated PP-IOTN's privacy guarantees based on two fundamental metrics: chance of individual data leakage and epsilon (Ɛ). Epsilon is a key parameter in PP, measuring the extent of privacy leakage. Lower Ɛ values indicate better privacy preservation. Furthermore, we evaluated the likelihood of individual data point disclosure during the PP process, which is critical for ensuring that no one data point has a substantial influence on the model predictions or learning process.

4.3.2. Probability of Data Disclosure

We attempted to minimize the danger of disclosing specific data points by limiting the noise levels during gradient aggregation, which is consistent with the key concepts of differential privacy. our experiments included a careful examination of PP-IOTN across 150 rounds, each with 5 local iterations. We got an impressive 93% accuracy rate, outperforming prior findings reported in the literature, including those obtained with CLDP (cyclic local differential privacy). The improved accuracy results demonstrated the effectiveness of our parameter setting in achieving the best balance between PP and accuracy within the Federated Learning. This significant difference in accuracy results demonstrated the uniqueness of our strategy over LDP-Fed.

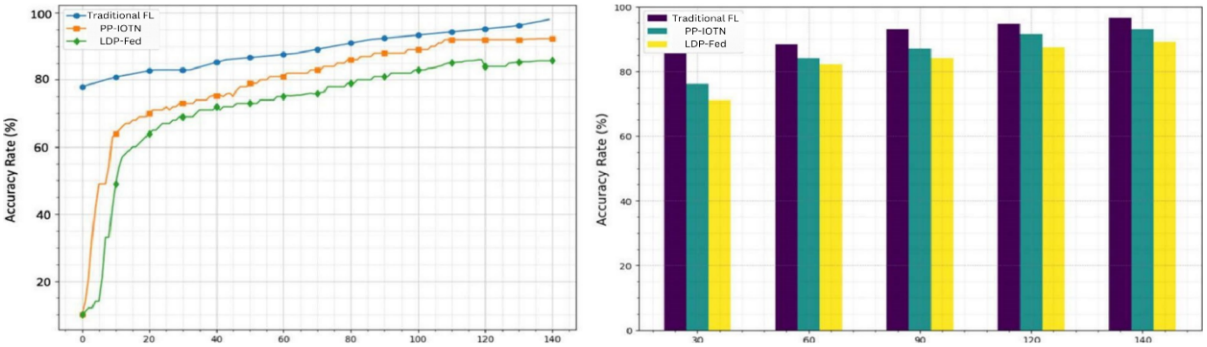


Figure 6. CIC-IOT-2023 NON-iid dataset's accuracy for different rounds (delta = 0.00001), compared to LDP-Fed

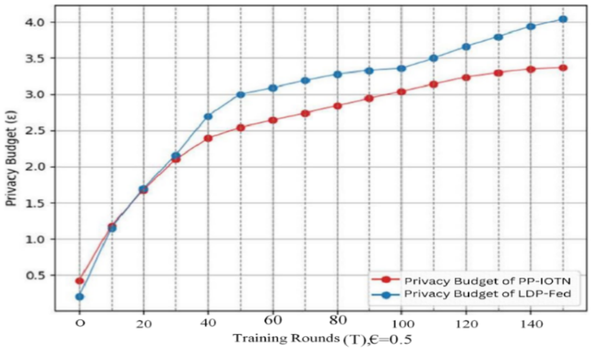
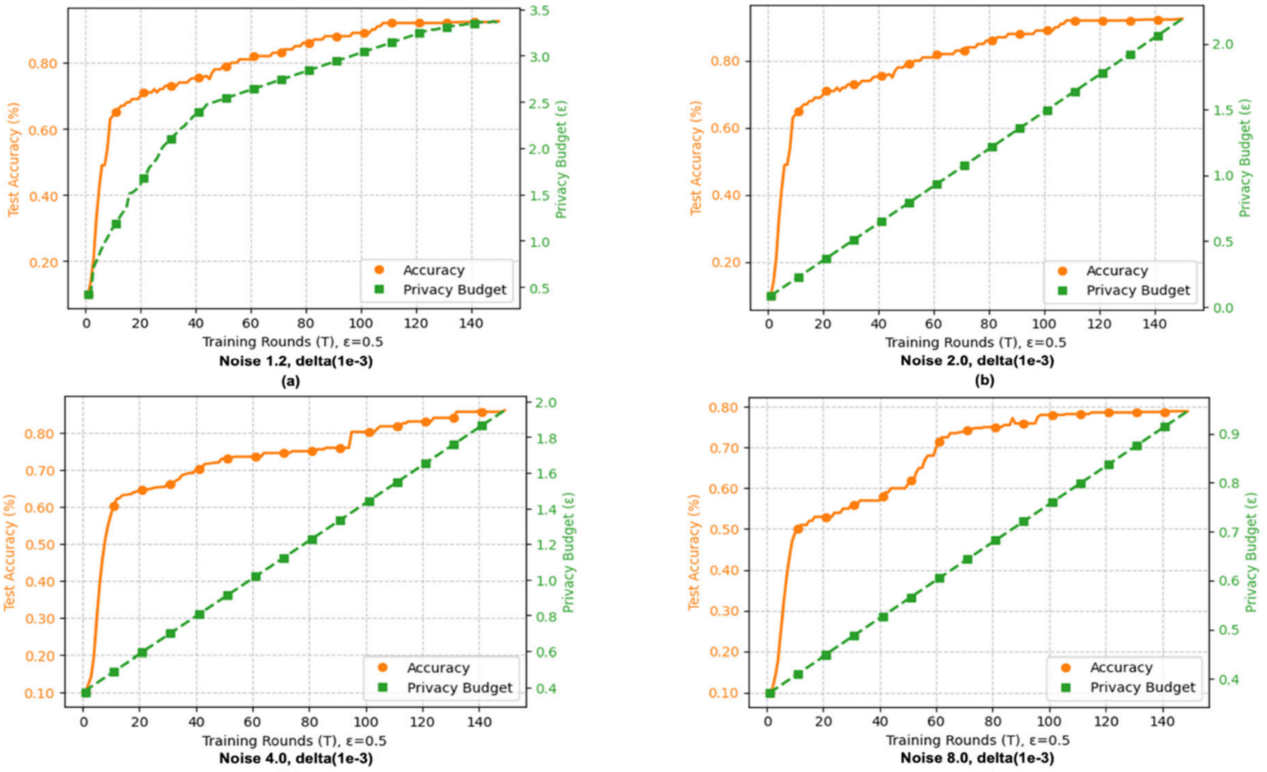


Figure 7. Privacy Budget of PP-IOTN and LDP-Fed

Furthermore, we performed a thorough comparison across a variety of factors, including epsilon, number of customers, number of iterations, and noise levels and report our findings in Figs. 6(a) and 6(b). These data illustrated PP-IOTN's performance in terms of accuracy, privacy budget, and loss rate, demonstrating its ability to reduce privacy concerns while retaining accuracy. Notably, Fig. 7 compares the privacy budgets of PP-IOTN and LDP-Fed, demonstrating the major benefit of our suggested technique in terms of privacy preserving. The findings of our experiment are shown in Table 5 and Figs. 8(a) to 8(d), which represent the test's budget for privacy and accuracy in the proposed PP-IOTN. To evaluate the suggested FL with PP-IOTN, we continuously fixed the delta (¥) value at 1e-3. We maintained the epsilon

(Ɛ) constant 0.5, for noise levels of 1.2, 2.0, 4.0, and 8.0, respectively. Every experiment was run through 150 rounds. At noise level = 1.2, the accuracy was notable (92.3%), 89.2% at noise = 2.0, 86.0% at noise = 4.0, and a notable decline to 78.6% at noise = 8.0. The calculated losses in privacy are 2.19, 1.95, 0.95, and 3.36.

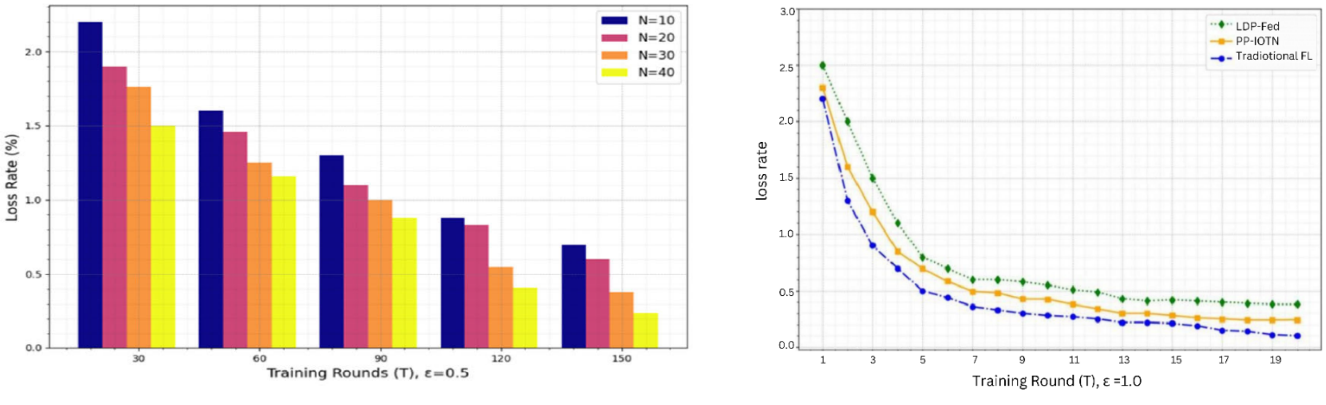


(a) (b) (c) (d)

Figure 8. The accuracy and privacy budget of PP-IOTN (delta 0.001) and varying noises.

As noise levels grew, the model's accuracy decreased but provided better privacy assurances, indicating a trade-off. This study concludes that FL and PP-IOTN approaches effectively solve privacy concerns. Tests conducted using the Opacus module in Python demonstrate that adding noise and varying the delta value can improve user privacy without sacrificing model accuracy. The study found that adding noise influences model accuracy, with lower levels leading to a bigger gain in accuracy than higher levels. Furthermore, the trials demonstrate that adjusting the delta value can improve privacy while maintaining model accuracy.

Furthermore, our experiments revealed noteworthy patterns in loss rates across multiple numbers of customers and epochs, as shown in Fig. 9(a) and 9(b). These statistics showed that PP-IOTN gradually improved through iterations, finally achieving a 92.5% final score. n Fig. 10(a), Fig. 10(b) LDP-Fed showed sporadic variations, showing the stability of PP-IOTN.



(a) (b)

Figure 9. Comparative loss rates for (a) PP-IOTN with respect to client count and (b) PP-IOTN with other approaches

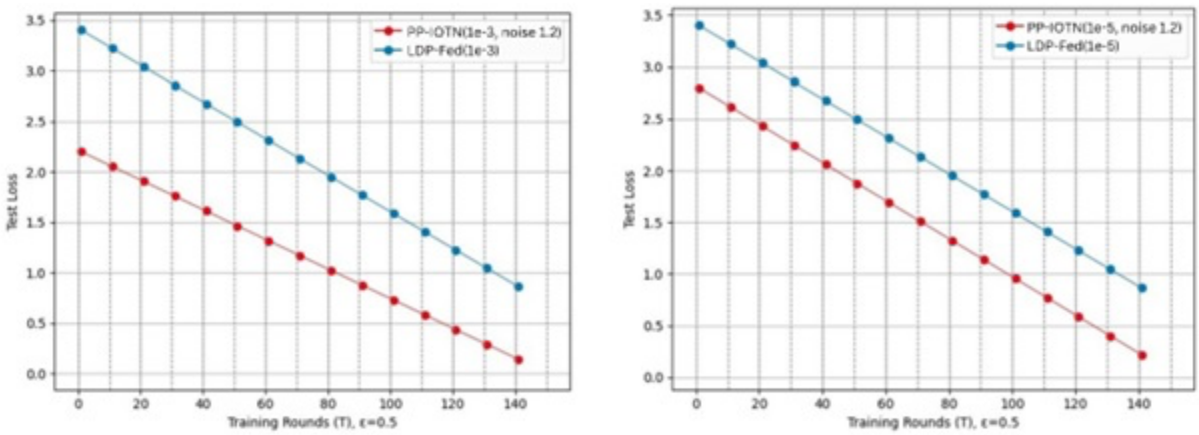


Figure 10. Loss rate comparison of (PP-IOTN and LDP-Fed with le-3 and le-5)

Table 5. FL with differential privacy (Delta = 0.001)

Variables				
Privacy	3.36	2.19	1.95	0.95
Noise	1.2	2.0	4.0	8.0
Accuracy	92.3	89.0	86.0	78.6
Rounds	150	150	150	150
Epsilon	0.5	0.5	0.5	0.5

5. Conclusion

Finally, this study emphasizes the usefulness of privacy-preserving approaches, particularly in the context of forecasting cyber threats in IoT networks. Our studies, which use the PP-IOTN framework, show how to effectively integrate tuned Differential Privacy (PP-IOTN) into Federated Learning. We demonstrated via thorough research with the CIC-IOT-2023 dataset that introducing noise and altering delta values dramatically improves user privacy while maintaining model accuracy. Our results show that the amount of noise supplied affects model accuracy, with lower levels resulting in bigger accuracy increases. Additionally, altering delta values improves privacy without affecting model accuracy. This research advances privacy-preserving strategies in IoT security, highlighting the significance of differential privacy in protecting sensitive data and detecting cyber threats in IoT networks.

Reference

1. Malik Bader Alazzam et al "Federated Deep Learning Approaches for the Privacy and Security of IoT Systems", Wireless Communications and Mobile Computing, vol. 2022, Article ID 1522179, 7 pages, 2022. https://doi.org/10.1155/2022/1522179
2. Briggs, C et al (2021). A Review of Privacy-Preserving Federated Learning for the Internet-of-Things.) Federated Learning Systems. Studies in Computational Intelligence, vol 965. Springer, Cham. https://doi.org/10.1007/978-3-030-70604-3_2
3. Viraaji Mothukuri, Prachi Khare, et al "Federated Learning-based Anomaly Detection for IoT Security Attacks": INTERNET OF THINGS JOURNAL, 2021
4. M. A. Ferrag, O. Friha, L. Maglaras, H. Janicke and L. Shu, "Federated Deep Learning for Cyber Security in the Internet of Things: Concepts, Applications, and Experimental Analysis," in *IEEE Access*, vol. 9, pp. 138509-138542, 2021, doi: 10.1109/ACCESS.2021.3118642
5. Uprety, D. B. Rawat and J. Li, "Privacy Preserving Misbehavior Detection in IoV Using Federated Machine Learning," *2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)*, Las Vegas, NV, USA, 2021, pp. 1-6, doi: 10.1109/CCNC49032.2021.9369513.
6. Randhir Kumar, "A deep blockchain-based trustworthy privacy-preserving secured framework in industrial internet of things systems" Department of Information Technology, National Institute of Technology, G.E Road, Raipur,Chhattisgarh 492010, India.
7. Arzoo Miglani a, Neeraj Kumar "Blockchain management and machine learning adaptation for IoT environment in 5G and beyond networks" 5 February 2021
8. Makkar, Aaisha, et al. "Fedlearnsp: Preserving privacy and security using federated learning and edge computing." IEEE Consumer Electronics Magazine 11.2 (2021): 21-27.
9. Makkar, U. Ghosh, D. B. Rawat and J. H. Abawajy, "FedLearnSP: Preserving Privacy and Security Using Federated Learning and Edge Computing," in *IEEE Consumer Electronics Magazine*, vol. 11, no. 2, pp. 21-27, 1 March 2022, doi: 10.1109/MCE.2020.3048926.
10. P. Zhang, Y. Wang, N. Kumar, C. Jiang and G. Shi, "A Security- and Privacy-Preserving Approach Based on Data Disturbance for Collaborative Edge Computing in Social IoT Systems," in *IEEE Transactions on Computational Social Systems*, vol. 9, no. 1, pp. 97-108, Feb. 2022, doi: 10.1109/TCSS.2021.3092746.
11. D. C. Attota, V. Mothukuri, R. M. Parizi and S. Pouriyeh, "An Ensemble Multi-View Federated Learning Intrusion Detection for IoT," in *IEEE Access*, vol. 9, pp. 117734-117745, 2021, doi: 10.1109/ACCESS.2021.3107337.
12. Kan, Xiu, et al. "A novel IoT network intrusion detection approach based on adaptive particle swarm optimization convolutional neural network." Information Sciences 568 (2021): 147-162.
13. Liu, Gaoyang, et al. "Keep your data locally: Federated-learning-based data privacy preservation in edge computing." IEEE Network 35.2 (2021): 60-66.
14. Puri, Vikram, et al. "Blockchain meets IIoT: An architecture for privacy preservation and security in IIoT." 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA). IEEE, 2020.
15. Magaia, Naercio, et al. "Industrial internet-of-things security enhanced with deep learning approaches for smart cities." IEEE Internet of Things Journal 8.8 (2020): 6393-6405.
16. Pirtama, Adrian, et al. "Improvement Attack Detection on Internet of Thinks Using Principal Component Analysis and Random Forest." Media Journal of General Computer Science 1.1 (2024): 14-19.
17. [17] Khanday, Shahbaz Ahmad, Hoor Fatima, and Nitin Rakesh. "A Novel Data Preprocessing Model for Lightweight Sensory IoT Intrusion Detection."
18. Hamidouche, Mounia, Eugeny Popko, and Bassem Ouni. "Enhancing iot security via automatic network traffic analysis: The transition from machine learning to deep learning." Proceedings of the 13th International Conference on the Internet of Things. 2023.
19. Abbas, Sidra, et al. "A novel federated edge learning approach for detecting cyberattacks in IoT infrastructures." IEEE Access (2023).

20. Hasan, Mohammad Kamrul, et al. "An Explainable Ensemble Deep Learning Approach for Intrusion Detection in Industrial Internet of Things." IEEE Access (2023).
21. Alasmary, Faris, et al. "Shieldrnn: A distributed flow-based ddos detection solution for iot using sequence majority voting." IEEE Access 10 (2022): 88263-88275.
22. Bakhsh, Shahid Allah, et al. "Enhancing IoT network security through deep learning-powered Intrusion Detection System." Internet of Things 24 (2023): 100936.
23. Abreu, Diego, and Antonio Abelém. "Ominacs: Online ml-based iot network attack detection and classification system." 2022 IEEE Latin-American Conference on Communications (LATINCOM). IEEE, 2022..br
24. Govindaraju, Nevetha. "Towards Examining Supervised and Unsupervised Learning for IoT Attack Detection." (2023).
25. Butt, Nazia, et al. "Intelligent deep learning for anomaly-based intrusion detection in IoT smart home networks." Mathematics 10.23 (2022): 4598.
26. Kostas, Kahraman, Mike Just, and Michael A. Lones. "IoTGeM: Generalizable Models for Behaviour-Based IoT Attack Detection." arXiv preprint arXiv:2401.01343 (2023).
27. Reed, Andy, Laurence S. Dooley, and Soraya Kouadri Mostefaoui. "Packet Filtering and Sampling for Efficient Slow Denial of Service Detection in Resource Scarce IoT Networks." 2023 International Symposium on Networks, Computers and Communications (ISNCC). IEEE, 2023.
28. Denis Parfenov* , Lubov Grishina, Artur Zhigalov, and Anton Parfenov Orenburg State University, "Investigation of the impact effectiveness of adversarial data leakage attacks on the machine learning models" Orenburg, 460018, Russia
29. Gheni, Hadeel Qasem, and Wathiq L. Al-Yaseen. "Two-Step Data Clustering for Improved Intrusion Detection System Using Ciciot2023 Dataset." Available at SSRN 4762201.
30. Alosaimi, Mohammed, Omer Rana, and Charith Perera. "Testbeds and evaluation frameworks for anomaly detection within built environments: A systematic review." (2023).
31. Fu, Anmin, et al. "VFL: A verifiable federated learning with privacy-preserving for big data in industrial IoT." IEEE Transactions on Industrial Informatics 18.5 (2020): 3316-3326.
32. Zhao, Yang, et al. "Privacy-preserving blockchain-based federated learning for IoT devices." IEEE Internet of Things Journal 8.3 (2020): 1817-1829.
33. Thilakarathne, Navod Neranjan, et al. "Federated learning for privacy-preserved medical internet of things." Intell. Autom. Soft Comput 33.1 (2022): 157-172.
34. Abdel-Basset, Mohamed, et al. "Privacy-preserved learning from non-iid data in fog-assisted IoT: A federated learning approach." Digital Communications and Networks (2022).
35. Li, Jiachun, et al. "A federated learning based privacy-preserving smart healthcare system." IEEE Transactions on Industrial Informatics 18.3 (2021).
36. Wang, Ruijin, et al. "RPIFL: Reliable and Privacy-Preserving Federated Learning for the Internet of Things." Journal of Network and Computer Applications 221 (2024): 103768.
37. Al-Marri, Noor Ali Al-Athba, Bekir S. Ciftler, and Mohamed M. Abdallah. "Federated mimic learning for privacy preserving intrusion detection." 2020 IEEE international black sea conference on communications and networking (BlackSeaCom). IEEE, 2020.
38. Li, Yong, et al. "Privacy-preserving federated learning framework based on chained secure multiparty computing." IEEE Internet of Things Journal 8.8 (2020): 6178-6186.
39. Liu, Tian, et al. "High-accuracy low-cost privacy-preserving federated learning in IoT systems via adaptive perturbation." Journal of Information Security and Applications 70 (2022): 103309.
40. Sahinbas, Kevser, and Ferhat Ozgur Catak. "Secure multi-party computation-based privacy-preserving data analysis in healthcare IoT systems." Interpretable Cognitive Internet of Things for Healthcare. Cham: Springer International Publishing, 2023. 57-72.

41. Jalali, Nasir Ahmad, and Hongsong Chen. "Federated learning security and privacy-preserving algorithm and experiments research under internet of things critical infrastructure." Tsinghua Science and Technology 29.2 (2023): 400-414.

42. Aqsa Ijaz, Ammar Ahmad Khan, Muhammad Arslan, Ashir Tanzil, Alina Javed, Muhammad Asad Ullah Khalid, & Shouzab Khan. (2024). Innovative Machine Learning Techniques for Malware Detection. Journal of Computing & Biomedical Informatics, 7(01), 403–424.

43. Ammar Ahmad Khan , Muhammad Arslan , Ashir Tanzil , Rizwan Abid Bhatty , Muhammad Asad Ullah Khalid , Ali Haider Khan. (2024). Classification Of Colon Cancer Using Deep Learning Techniques On Histopathological Images. Migration Letters, 21(S11), 449–463

44. Abbas, F., Iftikhar, A., Riaz, A., Humayon, M., & Khan, M. F. (2024). Use of Big Data in IoT-Enabled Robotics Manufacturing for Process Optimization. Journal of Computing & Biomedical Informatics, 7(01), 239-248.