Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 5, 2026

A FEDERATED LEARNING APPROACH FOR BIG DATA ANALYTICS IN ENHANCED INTRUSION DETECTION

Mobashirah Nasir[1*], Muhammad Tahir Mehmood [2], Waheed Raza [3], Hifza Rani[4], Uzma Iqbal[5], and Muhammad Yousif[6]

[1]Department of Informatics and systems, University of Management and Technology, Lahore , Pakistan , mobashirah.nasir@umt.edu.pk

[2]Department of Computer Science, National university of computer and emerging sciences Lahore , Pakistan, mtahirmehmood2000@gmail.com

[3]Department of Computer Sceince , Minhaj Univeristy , Lahore , Pakistan, wraza0449@gmail.com

[4]Department of Computer Sceince , Minhaj Univeristy , Lahore , Pakistan, hifza.sarwar1@gmail.com

[5]Department of Artificial Intelligence, The Islamia Univeristy of Bahawalpur, Bahawalpur, Pakistan, uzmaiqbal901@gmail.com

[6]Department of Computer Science, National University of Modern languages Sub-Campus, Lahore, Pakistan, myousif.cs@gmail.com

DOI:-https://doi.org/10.5281/zenodo.20286271

Keywords
Federated Learning, Data Analytics, Intrusion Detection

Article History
Received: 19 April 2026
Accepted: 18 May 2026
Published: 19 May2026

Corresponding Author: *

mobashirah.nasir@umt.edu.pk

Abstract

As the attacks become more complex and more numerous on the interconnected networks, traditional intrusion detection systems (IDS) failed to adapt because of their reliance on centralized architectures, with very large privacy, scalability and adaptability problems. The Multiple Data Analysis Using Federated Learning with ML Algorithms in Intrusion Detection presents a novel approach to developing an intelligent, privacy-preserving and scalable intrusion detection system (IDS), addressing these challenges. The goal of the research is to develop and evaluate a decentralized FL-based IDS with at least 95% detection accuracy to be deployed in different heterogeneous data sets, offering the data privacy and high functionality in the distributed environment of the real world. These are working on an FL-based prototype using CNN models, comparing the performance of the model with the traditional centralized ML models and evaluating the scalability of the model using different datasets such as TII-SSRC-23, NSL-KDD, and UNSW-NB15 by running multiple federated clients. In summary, study reveals that federated learning is a powerful solution to big data analytics for enhanced privacy-preserving intrusion detection systems. The need for intelligent cybersecurity solutions that can be applied to ensure privacy and data utility in

an era where data sharing has been curtailed by legislation such as GDPR drives this research. The proposed FL system trains the local model along the distributed nodes, thereby decreasing the risk of data leakage and greatly enhancing the flexibility of the FL system to the different network conditions. This is in contrast to the traditional centralized system where raw data need to be aggregated and then used for training the model. Centralized and federated CNN models will then be constructed and trained on, and techniques to protect privacy, such as Differential Privacy and Secure Aggregation will be added to ensure that the data is not revealed during the round of communication. The models will be tested by using accuracy, precision, recall, F1-score and ROC-AU evaluation metrics and scalability will be tested by adding more federated clients to the real-life distributed scenario. The research will also try to visualize the results of performance using confusion matrices and ROC curve and make the performance results easy to understand and transient. The proposed system is expected to be more effective than traditional IDS systems in terms of the detection accuracy, the ability to generalize the action on various datasets and privacy protection without undermining its efficiency. The expected outcomes are the development of a fully functional prototype of an IDS based on FL, a comparative performance report between federated and centralized approach and documentation of data and code repository for future research and reproduction. Lastly, the objective of this research is to create a scalable, privacy-preserving framework that can be successfully used to detect and mitigate intrusions in the network at multiple organizations and data source across different data sources to help the cybersecurity community. The results will aid companies, researchers and policymakers in finding new ways to enhance the network defense mechanisms in the highly data-oriented digital world.

1. Introduction

Federated learning is an approach that improves IDs accuracy while maintaining data privacy by having the models trained collaboratively, without sharing the original data. Rather than pooling sensitive network traffic, each node trains the model on its own and only sends its parts of the training updates in encrypted form, which are pooled to create a global model. This way, IDS can be trained to detect a variety of attack scenarios using different data sets, which helps with the generalization and helps the IDS to perform well in other networks. Meanwhile, privacy is maintained, information leakage is reduced and regulations like GDPR are complied with. The performance of federated learning models is comparable to the accuracy, precision and recall of centralized machine learning, and even outperforms it at times because of the training of data from a more diverse source. Centralized models might have some isolated gains in precision but lack in privacy and scalability aspects of data sharing. To overcome these challenges, federated models are developed that keep data private, are more robust, and have better detection performance, which better suit the real-world use of distributed intrusions in the federated IDS [1]. The evolution of cloud computing, IoT and networking has made attacks more sophisticated, bigger and harder to stop by conventional, centralized intrusion detection systems. Issues with scalability, privacy concerns, and compliance with regulations like GDPR are challenges for these systems [2]. To overcome these shortcomings, Federated learning provides a decentralized, privacy-preserving way of collaboratively training models without exchanging raw data. The goal of this research is to design and implement a practical and accurate federated intrusion detection framework for today's distributed systems. The project compares and measures the effectiveness of centralized and federated learning based on common evaluation metrics like accuracy, precision, recall, F1 score, and ROC-AUC [3]. Experiments were performed with multi-clients to evaluate techniques for preserving privacy and scalability, and to visualize performance with confusion matrices and ROC curves. The outcomes are a working prototype of a federated IDS, in-depth comparison of performance, and reproducible data and code for the prototype.

2. Literature Review

The recent developments in cybersecurity have underscored the need to combine Federated Learning (FL) with Big Data Analytics for enhancing Intrusion Detection Systems (IDS). Traditional IDS has centralized machine learning approaches, where the huge-scale network traffic data is collected to the centralized servers for model training. But centralized architectures come with a number of problems, such as privacy leakage, high communication overhead, scalability problems, and single points of failure. The decentralized approach, known as Federated Learning, now is a practical method that enables multiple devices [13] or organizations to train a model in a collaborative way without sharing raw data. Author proposed a Federated Learning Based Hybrid Deep Learning Intrusion Detection Model for Industrial Internet of Things (IIoT). The results of their study indicated that it is

possible to achieve a large improvement in the detection accuracy by incorporating FL to CNN and LSTM models without sacrificing data privacy. The authors stated that they achieved improved false positives and scalability compared to the traditional centralized IDS solutions. However, there were communication latency and complexity issues in the framework with heterogeneous environments [4].

Author suggested a blockchain-based federated learning intrusion detection system (BFLIDS) for Internet of Medical Things (IoMT) networks. To provide the security of decentralized healthcare systems, the study integrated blockchain, adaptive CNN-BiLSTM models, and FL. Their detection accuracy rates were found to be above 97% for the Edge-IIoTSet data and TON-IoT data. The proposed framework was effective, but also used more resources and blockchain overhead, which could hinder its adoption in lightweight IoT devices. However, the suggested framework faced challenges due to its higher resource usage and blockchain overhead, potentially hindering its adoption in resource-constrained devices like IoT [5].

Authors introduced the IDAC framework, which is an automatic detection of anomalies in an IoT environment using IDS mechanisms based on federated learning. The framework improved the efficiency of anomaly detection while preserving the privacy of the users. Their study revealed FL's ability to lower the dependency on the central system and enhance collaborative intelligence among IoT devices. It was found, however, that the framework had a few drawbacks with regards to the convergence rate and processing of highly imbalanced datasets [6].

Authors carried out a comprehensive survey of FL applications in intrusion detection systems. The study classified and classified existing FL-IDS techniques according to the privacy level, communication efficiency, aggregation techniques, and attack detection. The authors pointed out that FL can significantly reduce communication costs and privacy risks in large-scale data cyber-security scenarios. They, however, found that there are some persistent problems like non-IID data distribution, poisoning attacks, limited computational resources, and interoperability problems between heterogeneous devices [7].

In 2024, scholars presented a resource-efficient federated learning approach to network intrusion detection. They investigated ways to reduce computational expenses and communication overhead while maintaining excellent detection accuracy. The framework worked well in scenarios where the intrusion detection system (IDS) operates in a distributed fashion at a large scale. However, it was necessary to enhance the approach for adaptive threat intelligence and real-time intrusion detection [8].

In recent studies, automated and adaptive FL mechanism were also studied. Lightweight CNN models were automatically optimized for intrusion detection tasks by proposing evolutionary neural architecture search-based federated IDS frameworks for Industrial Control Systems (ICS). The automatic design improved the flexibility and detection ability of the system in various cases. However, training

and optimizing processes were still very complex and time consuming.

Similarly, authors proposed a powerful FL system, FedMADE, for IoT intrusion detection, which addressed the data heterogeneity and class imbalance problem. The dynamic aggregation method greatly enhanced the accuracy of minority attack classification and resistance to poisoning attacks. Despite those improvements, the model still had some additional communication delays during the aggregation phase [9].

Furthermore, researchers pointed out transfer learning and knowledge distillation techniques to improve the transferability and adaptability of the IDS in the heterogeneous environment of IoT. The methodologies facilitated the IDS models to perform better on new attacks and generalize on distributed datasets. However, maintaining consistency of the models and simplifying training are still open research questions.

Overall, research shows that federated learning offers a potent answer for big data analytics in improved intrusion detection systems that protect privacy. IDS architectures based on FL offer greater accuracy, scalability, and data privacy, and minimize the reliance on centralized data. However, the current research still has significant limitations in communication overhead, non-IID data distribution, resource limitations, adversarial attacks, and model convergence efficiency.

High level intrusion detection systems (IDSs) are required to deal with the increasing size and sophistication of cyber threats. The IDSs of the traditional machine learning types usually presuppose the centralized data gathering, which provokes the issues of privacy, scalability, and latency. The proposed solution, Federated Learning (FL), offers an effective alternative since it allows decentralized and privacy-preserving training of models in distributed nodes. FL can be combined with big data analytics to leverage the IDS with large, multivariate, and real-time data.

Table 1. Literature Review with dataset

Reference	Contribution	Limitation
[4]	Proposed hybrid CNN-LSTM with FL for IIoT intrusion detection; improved accuracy and privacy preservation.	High computational complexity and communication latency in heterogeneous networks.
[5]	Integrated blockchain and FL for secure IoMT intrusion detection with high detection accuracy.	Blockchain overhead and high resource consumption for lightweight devices.
[6]	Developed autonomous anomaly extraction framework using FL in IoT IDS.	Slow convergence and difficulty handling imbalanced datasets.
[7]	Comprehensive survey of FL-based IDS architectures, challenges, and future directions.	Identified unresolved issues such as non-IID data and poisoning attacks.
[8]	Proposed communication-efficient FL framework for distributed IDS systems.	Limited support for real-time adaptive intrusion detection.
[9]	Introduced evolutionary neural architecture search for lightweight FL-based IDS models.	Increased optimization and training complexity.

[10]	Enhanced transferability and detection of unseen attacks using FL techniques.	Challenges in maintaining model consistency across clients.
[11]	Improved IDS performance in heterogeneous IoT environments using knowledge distillation.	High training complexity and aggregation overhead.

3. Research Methodology

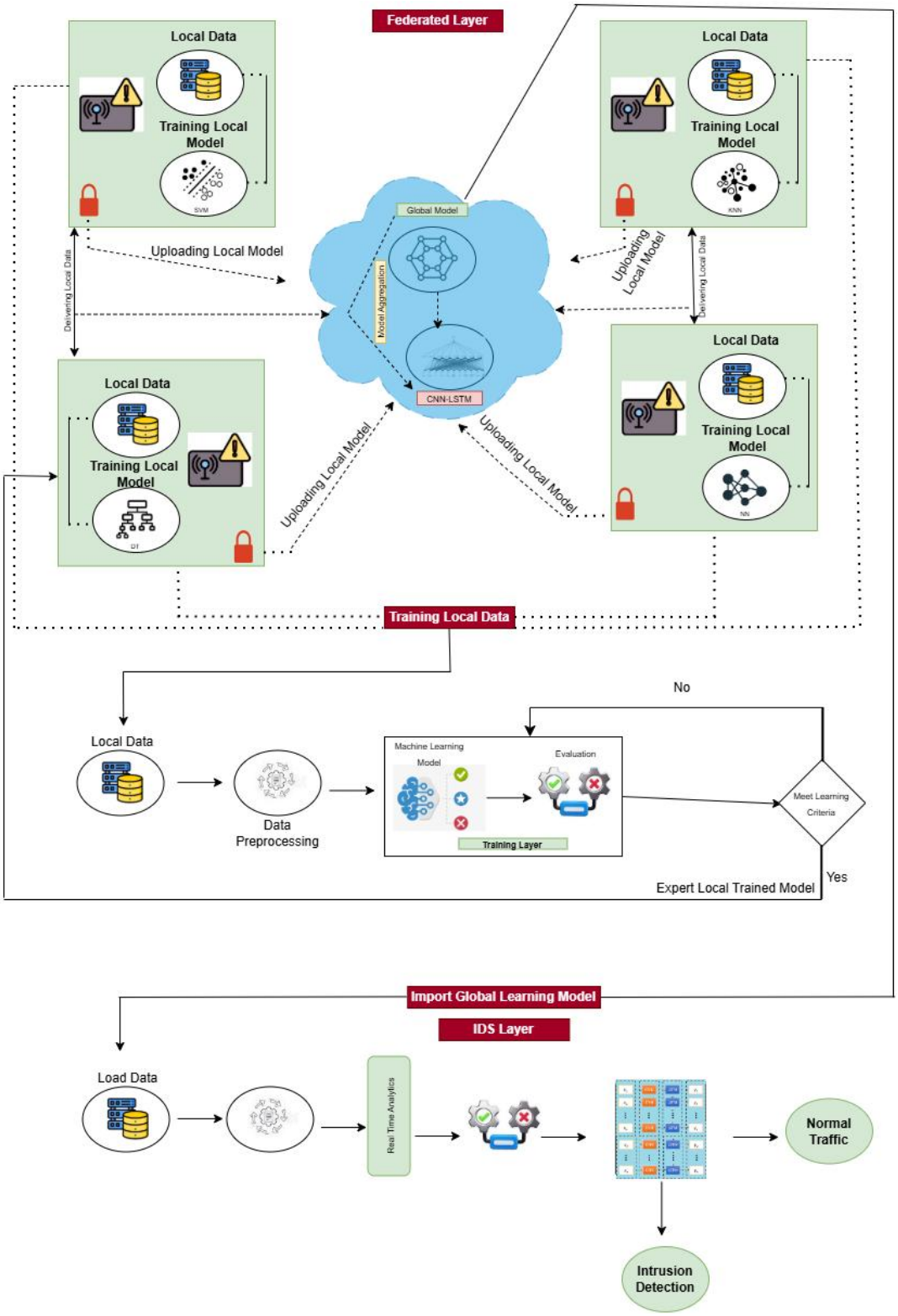
This study uses experimental research method, which is a kind of study that deals with the effect or causal relationship between the independent variable and the dependent variable in which the changes in the independent variable directly impact the dependent variable. The proposed Federated Learning–based Intrusion Detection (FIDL) framework tests the performance of Federated Learning with both client and server techniques, by using accuracy, precision, recall, and ROC metrics. Considering the increasing amount and transformation of malicious network traffic, this study is an extension of the previous studies that investigated various machine learning algorithms to improve the performance of intrusion detection [12].

It is a quantitative approach focusing on the design and evaluation of an intrusion detection system (IDS) using a federated learning framework. This technique consists of multiple network nodes that collaborate to train machine learning models without sharing raw network data. Each node has its own traffic and builds its own IDS model, such as Support Vector Machines (SVM), Decision Trees, Neural Networks, and K-Nearest Neighbors (KNN).

The learning parameters are transmitted securely to a centralized federated server, without sharing sensitive data. At this level, the received parameters are fused together to create an optimized model over the whole globe, like a CNN-LSTM model. The new global model is then sent to the nodes that are participating, and is refined by the nodes on an iterative basis until the desired learning goals and performance levels are achieved.

When the model is converged, the final global IDS is deployed to check the live traffic from the network and quantitatively categorize the activities as normal or intrusive. This federated learning approach has shown to achieve measurable improvements in detection accuracy, preserve data privacy and facilitate effective inter-organizational or inter-distributed network cooperation.

Figure 1. Proposed Methodology Federated Learning approach for big data Analytics using Intrusion detection



The federated learning-based intrusion detection system (IDS) is illustrated in Figure 1, where the local models are trained at distributed nodes using local data. Only model parameters are sent from the modelers to a central server to build a global model, and the global model is sent back to the modelers for better detection performance.

3.1 Research Methods

3.1.1 Local Data Layer

In order to protect privacy and comply with regulations, this layer displays distributed data sources where traffic and records remain local.

3. 1.2 Data Preprocessing Layer

Local data is cleaned and transformed at each site to ensure efficient and accurate model training.

3. 1.3 Local Training Layer

A local intrusion detection model is trained and evaluated by each client on preprocessed data, with algorithms tailored to the environment of each client until the target performance is achieved.

3. 1.4 Secure Model Update Layer

Only encrypted model changes are transferred to the federated server, while raw data remains local to provide privacy and safe communication.

3. 1.5 Federated Aggregation Layer

The federated server collects the informs from the clients and combines them together using techniques such as FedAvg without accessing the limited data.

3. 1.6 Global Model Layer

This layer produces an enhanced total intrusion detection model, with repeatedly updating and distributing the model to the clients to learn various attack patterns.

3. 1.7 Model Validation & Learning Control Layer

Global and local models are evaluated against performance metrics and retrained if needed until a validated expert model is achieved.

3. 1.8 IDS Deployment Layer

The configuration is implemented in the IDS layer and a developed global model is configured to process the live or batch network traffic.

3. 1.9 Decision & Output Layer

The stream of traffic in a distributed system can be either malicious traffic or normal traffic, which facilitates safe and scalable intrusion detection.

4. Simulation and Results

Table 2 presents the client-side performance (in terms of F1-scores) in a federated setting with 10 clients for KNN, SVM, Decision Tree and LightNN. All models are working well with Decision Tree and LightNN models attaining maximum and reliable score followed by that of KNN; SVM is slightly lower. The results indicate effective and robust local learning and LightNN is an attractive option for model aggregation in the global setting.

Table 2. Client-Side Federated Learning Local Data Performances

Client	KNN	SVM	DT	LightNN
1	0.9876	0.9646	0.9950	0.9906
2	0.9869	0.9618	0.9949	0.9909
3	0.9872	0.9634	0.9950	0.9894
4	0.9882	0.9651	0.9935	0.9921
5	0.9881	0.9624	0.9962	0.9916
6	0.9873	0.9619	0.9958	0.9893
7	0.9873	0.9611	0.9944	0.9928
8	0.9870	0.9599	0.9949	0.9892
9	0.9881	0.9615	0.9950	0.9917
10	0.9881	0.9626	0.9954	0.9926

In this table 3, the accuracy, sensitivity and specificity for server-side federated CNN-LSTM and a single client model are compared. The federated way has the best overall accuracy (0.98) that is suggestive of the qualities of combining data of various distributed clients. The sensitivity is high (0.98), which resources that it is effective in detecting intrusion trials correctly and the specificity is high (0.96), which means that it is efficient in discriminating normal traffic with minimum number of false alarms compared to most of the clients. Compared to the individual client models the accuracy is somewhat less, and the specificity is much less, especially for Client 1, Client 2, and Client 3, where the models trained on local data (which could be biased) have a higher false-positive rate.

Table 3. Server-Side Federated Learning CNN_LSTM

	Accuracy	Sensitivity	Specificity
Proposed Approach based on FL (Server Side)	0.98	0.98	0.96
Cleint1	0.97	0.97	0.72
Client2	0.96	0.96	0.76
Cleint 3	0.95	0.95	0.72
Cleint4	0.96	0.97	0.95

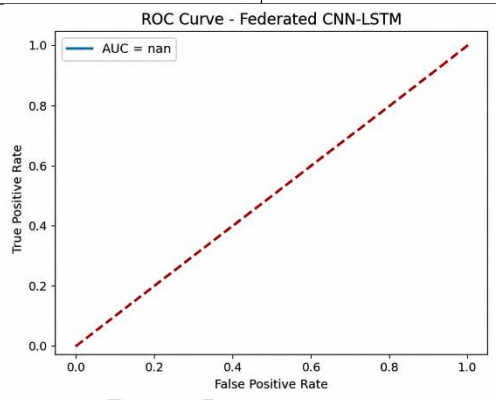


Figure 2. ROC Curve - Federated Learning (CNN-LSTM)

The ROC-plot of the Federated CNN-LSTM network is plotted in this figure, with dashed line at the diagonal corresponding to a random classifier (True Positive Rate = False Positive Rate). The ROC/AUC calculation was not well-posed in this evaluation, since it seems there is no model ROC curve available and the reported AUC is NaN. Typically, this is caused by invalid or degenerate prediction probabilities, e.g. all predictions in one class or no positive or negative samples in the test set. This plot, however, doesn't always reflect the ability to differentiate normal traffic from incursion traffic, but rather an issue with the data or evaluation itself, not necessarily the performance of the model.

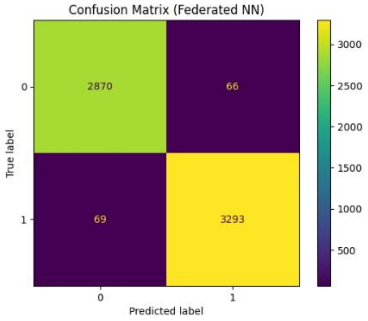


Figure 3. Confusion Matrix NN

Evolution Metrices:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FN+FP} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (3)$$

5. Conclusion

To overcome these drawbacks, a federated learning-based intrusion detection system (IDS) framework is proposed in this study, which is capable of solving the issues of privacy concerns, scalability, and adaptability in heterogeneous network environments. The framework facilitates privacy-preserving learning by training local models on distributed clients instead of sharing raw data, and is still usable in real-world scenarios. The suggested method is based on CNN models and the federated training and centralized training is compared on the datasets [14]. It tests accuracy, precision, recall, F1-score, ROC-AUC and confusion matrices as evaluation metrics, as well as scalability by adding more federated clients. To prevent data leakage during the communication rounds, privacy enhancement techniques such as Differential Privacy and Secure Aggregation are incorporated. The proposed model has achieved 98 % accuracy and 98 % sensitivity in conclusion, with 96 % specificity. It has a great success rate with correctly detecting normal traffic, but it is still weak at detecting attacks. While the federated framework also exhibits good privacy preservation and acceptable specificity, there is still potential for optimizing the model to increase the performance of attack detection and reliability. In summary, the study reveals federated learning as a viable approach for privacy-preserving IDS deployment. It offers a scalable secure basis for future cyber security systems; however, the existing detection capability needs to be increased before it can be used in critical environments.

References:

. Abubakar, M., Sattar, A., Manzoor, H., Farooq, K., & Yousif, M. (2025). Iiot: An infusion of embedded systems, tinyml, and federated learning in industrial iot. Journal of Computing & Biomedical Informatics, 8(02).

. Spalević, Ž., & Vićentijević, K. (2022). GDPR and challenges of personal data protection. *The European Journal of Applied Economics*, *19*(1), 55-65.

. Saeed, S., Haron, H., Riaz, I., Shah, H. K., Riaz, M., & Qadeer, M. (2026). A Federated Learning-based project for predicting disease diagnoses using Artificial Intelligence. *Int J Drug Deliv Technol*, *16*(10s), 975-984.

. Huang, J., Chen, Z., Liu, S. Z., Zhang, H., & Long, H. X. (2024). Improved Intrusion Detection Based on Hybrid Deep Learning Models and Federated Learning. Sensors, 24(12), 4002.

. Begum, K., Mozumder, M. A. I., Joo, M. I., & Kim, H. C. (2024). BFLIDS: Blockchain-Driven Federated Learning for Intrusion Detection in IoMT Networks. Sensors, 24(14), 4591.

. Ohtani, T., Yamamoto, R., & Ohzahata, S. (2024). *IDAC: Federated Learning-Based Intrusion Detection Using Autonomously Extracted Anomalies in IoT*. Sensors, 24(10), 3218.

7. Belenguer, A., Pascual, J. A., & Navaridas, J. (2025). *A Review of Federated Learning Applications in Intrusion Detection Systems*. Computer Networks, 258, 111023.
8. Corin, R. D., Cretti, S., & Siracusa, D. (2024). *Resource-Efficient Federated Learning for Network Intrusion Detection*. IEEE NetSoft Conference.
9. Sun, S., Sharma, P., Nwodo, K., Stavrou, A., & Wang, H. (2024). *FedMADE: Robust Federated Learning for Intrusion Detection in IoT Networks Using a Dynamic Aggregation Method*. arXiv.
10. Ghosh, S., Jameel, A. S. M. M., & El Gamal, A. (2024). *Improving Transferability of Network Intrusion Detection in a Federated Learning Setup*.

1. Shen, J., Yang, W., Chu, Z., Fan, J., Niyato, D., & Lam, K. Y. (2024). *Effective Intrusion Detection in Heterogeneous IoT Networks via Ensemble Knowledge Distillation-based Federated Learning.*
2. Akram, M. I., Yousaf, S., Zubair, M., Riaz, M., & Yousif, M. (2025). CYBERSECURITY CHALLENGES AND THE ROLE OF MACHINE LEARNING IN MODERN MALWARE DETECTION. *Spectrum of Engineering Sciences*, 653-668.
3. Fahad, H. M., & Asif, A. (2021). A simple FPP device for pulsed measurement of sheet resistance. *Instruments and Experimental Techniques*, *64*(6), 898-904.
4. https://www.kaggle.com/datasets/daniaherzalla/tii-ssrc-23