



Privacy-Preserving Intrusion Detection Using Federated Learning: Performance Evaluation and Cyber Law Implications for Data Protection and AI Accountability

Thierry Ezra Adrinaro^{1,*}, Bryan Daniel Angelino²

^{1,2}Department of Information Systems, Faculty of AI and Data Science, Universitas Pelita Harapan, Indonesia

ABSTRACT

This study investigates the use of federated learning for privacy-preserving intrusion detection systems and examines its implications from a cyber law perspective, particularly in relation to data protection and AI accountability. A comparative analysis was conducted between centralized machine learning models, including Random Forest and Extra Trees, and a federated Multilayer Perceptron (MLP) enhanced with the FedProx algorithm under both IID and non-IID data distributions. The results show that the centralized Random Forest model achieved the highest performance, while the federated model under IID conditions demonstrated comparable results, indicating that federated learning can maintain high detection capability without requiring centralized data collection. However, performance degradation was observed under non-IID settings, highlighting the impact of data heterogeneity on federated systems. From a legal perspective, federated learning supports key data protection principles such as data minimization and privacy by design by keeping sensitive data on local devices. Nevertheless, challenges related to model security, limited explainability, and potential information leakage remain, raising concerns regarding AI accountability and regulatory compliance. This study concludes that federated learning offers a promising approach for secure and privacy-aware intrusion detection, but requires additional safeguards to ensure both technical robustness and alignment with evolving cyber law frameworks.

Keywords Federated Learning; Intrusion Detection; Data Privacy; Cyber Law; Explainable AI

INTRODUCTION

The rapid growth of digital systems, cloud computing, and Internet of Things (IoT) technologies has significantly increased the volume and complexity of cyber threats. Traditional intrusion detection systems (IDS), which rely on signature-based or rule-based approaches, are no longer sufficient to detect sophisticated and evolving attacks. As a result, artificial intelligence (AI) and machine learning (ML) have emerged as effective solutions for enhancing intrusion detection by enabling adaptive, data-driven threat analysis [1], [2]. These approaches can identify complex attack patterns and improve detection accuracy compared to conventional methods. However, most AI-based IDS rely on centralized data collection, which raises serious concerns regarding data privacy, security, and regulatory compliance, especially when sensitive network or user data are involved [3].

To address these challenges, federated learning (FL) has been introduced as a

Submitted 4 May 2026
Accepted 26 June 2026
Published 1 September 2026

Corresponding author
Thierry Ezra Adrinaro,
1081230011@student.uph.edu

Additional Information and
Declarations can be found on
[page 203](#)

© Copyright
2026 Adrinaro and Angelino

Distributed under
Creative Commons CC-BY 4.0

privacy-preserving machine learning paradigm that enables collaborative model training without sharing raw data. In FL, data remains on local devices, and only model updates are exchanged, reducing the risk of data leakage and aligning with modern data protection principles such as data minimization and privacy by design [4]. Recent studies have demonstrated that FL can achieve competitive performance in intrusion detection and cybersecurity applications while maintaining data privacy. Furthermore, advanced techniques such as differential privacy, secure aggregation, and federated optimization algorithms have been proposed to enhance both the security and robustness of federated systems [5]. These developments represent the current state of the art in privacy-preserving AI for cybersecurity.

Despite these advancements, several critical challenges remain. First, federated learning systems often suffer from performance degradation under non-IID data distributions, which are common in real-world environments where data varies across organizations or devices. Second, FL introduces new security risks, including model poisoning and information leakage through gradient updates, which may compromise system integrity. Third, most existing research focuses primarily on technical performance, with limited attention to legal and regulatory implications, particularly regarding accountability, transparency, and compliance with data protection laws. The lack of explainability in AI-based IDS further complicates accountability, as decision-making processes are often difficult to interpret and audit.

This study aims to address these gaps by developing and evaluating a privacy-preserving intrusion detection system using federated learning and analyzing its implications from a cyber law perspective. Specifically, this research compares the performance of centralized and federated models under both IID and non-IID data distributions and examines how federated learning aligns with key principles of data protection and AI accountability. By integrating technical evaluation with legal analysis, this study contributes to bridging the gap between AI-driven cybersecurity solutions and evolving regulatory requirements, providing insights into the development of secure, transparent, and privacy-compliant intrusion detection systems.

Literature Review and Related Works

The application of artificial intelligence (AI) in cybersecurity has significantly advanced intrusion detection systems by enabling adaptive and data-driven threat identification. Machine learning and deep learning techniques have been widely adopted to improve detection accuracy, reduce response time, and identify complex attack patterns that are difficult to detect using traditional rule-based systems [6], [7], [8]. These approaches have proven effective in various domains, including IoT, cloud computing, and financial systems, where large-scale data analysis is required to detect anomalies and malicious behavior.

Despite these advantages, conventional AI-based intrusion detection systems typically rely on centralized data collection, which introduces significant privacy and security concerns. Sensitive data such as network traffic logs and user behavior patterns must often be aggregated in a central server, increasing the risk of data breaches and unauthorized access [9], [10]. In response to these challenges, privacy-preserving techniques have been proposed, with federated learning (FL) emerging as one of the most promising approaches. FL enables collaborative model training across multiple clients without sharing raw data,

thereby reducing privacy risks and supporting compliance with data protection regulations [11], [12], [13].

Recent studies have demonstrated that federated learning can achieve competitive performance compared to centralized models in intrusion detection tasks. Federated approaches have been successfully applied in IoT environments, distributed networks, and financial systems, where data privacy is a critical concern [14], [15], [16]. In addition, advanced federated architectures incorporating deep learning models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid models have shown improved detection accuracy and robustness against evolving cyber threats [17], [18]. Techniques such as differential privacy, secure aggregation, and blockchain integration have also been explored to enhance the security and trustworthiness of federated systems [19].

However, several challenges remain in the deployment of federated learning for intrusion detection. One major issue is the performance degradation under non-IID data distributions, which are common in real-world scenarios where data varies across clients. Furthermore, federated systems are vulnerable to adversarial attacks such as model poisoning and gradient leakage, which can compromise both model integrity and data privacy [20]. Another limitation is the lack of explainability in AI-based models, which raises concerns regarding transparency, auditability, and trust in decision-making processes.

From a legal perspective, the integration of AI in cybersecurity introduces new challenges related to data protection, accountability, and regulatory compliance. Modern cyber law frameworks emphasize principles such as data minimization, privacy by design, and organizational accountability, which align with the decentralized nature of federated learning. However, ensuring compliance requires not only protecting data but also providing transparency and explainability in AI-driven decisions. As cyber threats continue to evolve, there is a growing need to bridge the gap between technical advancements in federated learning and the legal requirements governing data protection and AI accountability.

Methodology

Dataset and Preprocessing

This study utilizes a cybersecurity intrusion dataset containing network activity features and a binary target variable indicating whether an attack is detected. The overall research workflow, including data preprocessing, model development, federated training, and evaluation, is illustrated in figure 1.

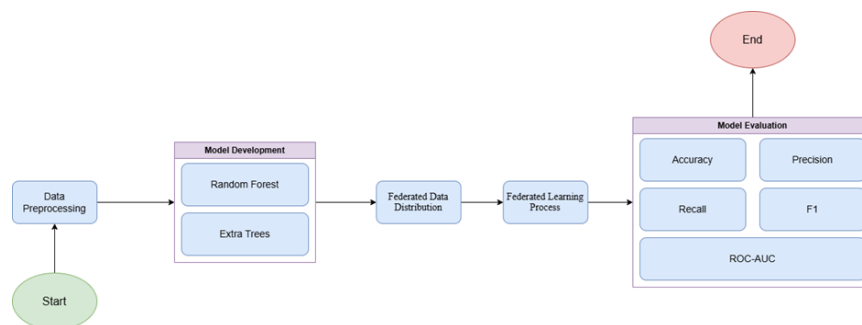


Figure 1 Research workflow of the proposed intrusion detection system

The dataset is first cleaned by removing irrelevant attributes and handling missing values. Categorical features are encoded using one-hot encoding, while numerical features are normalized using standard scaling. Missing values in numerical features are imputed using the median strategy, whereas categorical features are imputed using the most frequent value.

Numerical features are standardized using the following equation:

$$x' = \frac{x - \mu}{\sigma} \quad (1)$$

x is the original feature value, μ is the mean, and σ is the standard deviation.

Centralized Machine Learning Models

Two ensemble models are used: Random Forest and Extra Trees. These models estimate class probabilities and classify samples based on a decision threshold.

The prediction is defined as:

$$\hat{y} = \begin{cases} 1, & \text{if } p(x) \geq t \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

$p(x)$ is the predicted probability and t is the threshold selected to maximize the F1-score.

The F1-score is defined as:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (3)$$

Federated Learning Framework

In federated learning, the global model is trained by aggregating local updates from multiple clients without sharing raw data. The global model parameters at round t are defined as:

$$w^{(t+1)} = \sum_{k=1}^K \frac{n_k}{n} w_k^{(t)} \quad (4)$$

K is the number of clients

n_k is the number of samples at client k

$$n = \sum_{k=1}^K n_k \quad (5)$$

$w_k^{(t)}$ are local model parameters

This weighted aggregation ensures that clients with more data have a greater influence on the global model.

Federated Model Architecture

The model uses a multilayer perceptron (MLP). The output layer produces logits, which are converted into probabilities using the sigmoid function:

$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}} \quad (6)$$

Binary Cross-Entropy (BCE) loss is used:

$$L_{BCE} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (7)$$

Federated Optimization using FedProx

To handle heterogeneity, the FedProx algorithm adds a proximal term to the local loss function:

$$L_{FedProx} = L_{BCE} + \frac{\mu}{2} \|w - w^{(t)}\|^2 \quad (8)$$

μ is the proximal coefficient

w are local parameters

$w^{(t)}$ are global parameters

This term penalizes large deviations from the global model, stabilizing training.

Training Procedure

The federated learning process is conducted iteratively across multiple communication rounds to collaboratively train a global model without sharing raw data between clients. Initially, a global model is initialized, either randomly or based on prior knowledge, and distributed to all participating clients. In each round, individual clients perform local training using their own private datasets, updating the model parameters based on local data characteristics. After local training is completed, the updated parameters from each client are sent back to a central server, where a parameter aggregation step is performed, typically using methods such as weighted averaging to combine the contributions from all clients. The aggregated parameters are then used to update the global model, which is redistributed to clients for the next round of training. This iterative process continues until the model converges or reaches a predefined number of communication rounds, enabling the system to learn from decentralized data while preserving data privacy.

Performance is evaluated using ROC-AUC:

$$AUC = \int_0^1 TPR(FPR^{-1}(x)) dx \quad (9)$$

Evaluation Metrics

The main evaluation metrics include:

Precision:

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

Recall:

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

(12)

These metrics provide a comprehensive evaluation of intrusion detection performance.

Results

Model Performance Comparison

This study evaluates the performance of four models for intrusion detection, namely a centralized Random Forest (RF), a centralized Extra Trees (ET), and two federated learning models based on Multilayer Perceptron (MLP) optimized with the FedProx algorithm under both IID and non IID data distribution settings. The evaluation is conducted using five widely accepted classification metrics, including Accuracy to measure overall correctness, Precision to assess the proportion of correctly predicted positive instances, recall to evaluate the model’s ability to identify all actual positive cases, F1 score as the harmonic mean of Precision and Recall, and ROC AUC to measure the model’s capability in distinguishing between classes. The performance results of all models are comprehensively presented in [table 1](#), which provides a detailed quantitative comparison, and are further illustrated in [figure 2](#) to offer a clearer visual understanding of how centralized and federated approaches perform under different data distribution scenarios.

Table 1 Performance metrics of all models						
Model	Threshold	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Centralized Random Forest (Tuned)	0.50	0.8847	0.9922	0.7479	0.8529	0.8755
Centralized Extra Trees (Tuned)	0.55	0.8679	0.9560	0.7386	0.8333	0.8753
Federated MLP + FedProx (IID)	0.66	0.8816	0.9922	0.7409	0.8483	0.8733
Federated MLP + FedProx (Non-IID)	0.46	0.8291	0.8915	0.7034	0.7864	0.8575

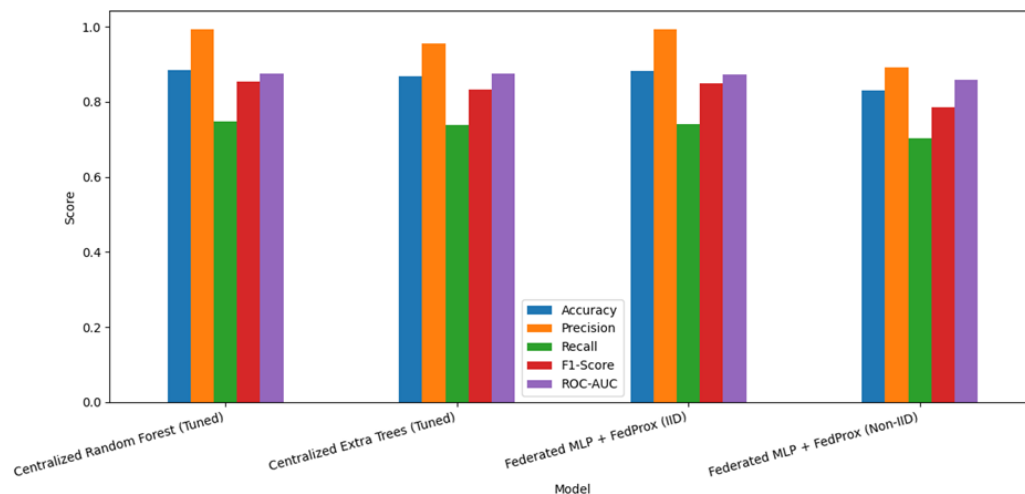


Figure 2 Performance comparison of centralized and federated models

The results indicate that the centralized Random Forest model with hyperparameter tuning achieved the highest overall performance, obtaining an F1 score of 0.8529 and a ROC AUC of 0.8755, which suggests strong capability in both classification balance and class separation. The Federated MLP under IID data distribution performed comparably, with an F1 score of 0.8483, demonstrating that federated learning can closely match centralized approaches when the data across clients is homogeneous. In contrast, the Federated MLP under non IID conditions experienced a clear decline in performance, highlighting the difficulty of training models effectively when data distributions vary significantly across clients. Additionally, across all evaluated models, precision values remain consistently high while recall values are relatively lower, indicating that the models tend to prioritize minimizing false positives over capturing all actual intrusion instances, which may lead to some attacks being missed.

ROC Curve Analysis

The ROC curves of all models are illustrated in figure 3, providing a visual comparison of their classification performance across different decision thresholds. The analysis shows that the centralized models and the Federated MLP under IID conditions exhibit very similar curve patterns, with AUC values ranging from approximately 0.87 to 0.88, indicating strong and consistent discriminative ability in distinguishing between normal and attack traffic. In contrast, the Federated MLP under non IID data distribution presents a noticeably lower ROC curve, which confirms a decline in classification performance when the data across clients is heterogeneous. This difference further emphasizes the impact of data distribution on federated learning effectiveness, where non IID conditions can significantly hinder the model's ability to generalize well across diverse data sources.

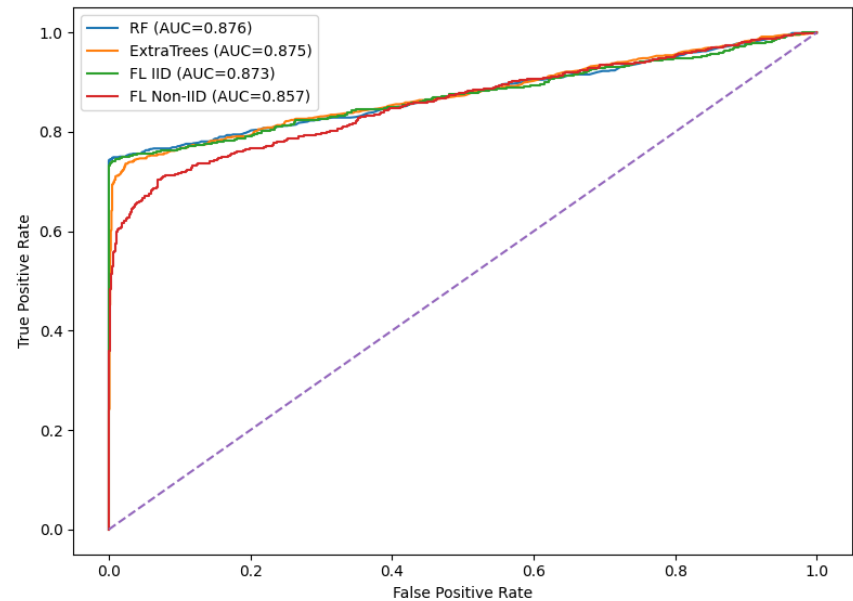


Figure 3 ROC curves of centralized and federated models

Federated Learning Convergence

The convergence behavior of the federated models across training rounds is illustrated in figure 4, showing clear differences between IID and non IID data settings. The Federated MLP under IID conditions demonstrates faster and more stable convergence, reaching an F1 score of approximately 0.8785 and a ROC AUC of 0.9378 after 30 communication rounds, which indicates efficient and consistent learning across clients with similar data distributions. In contrast, the Federated MLP under non IID conditions converges more slowly and stabilizes at a lower performance level, reflecting the challenges introduced by heterogeneous data across clients. This result highlights that data heterogeneity significantly affects the effectiveness of federated training, even when using optimization techniques such as FedProx, as the model struggles to generalize across diverse local data patterns.

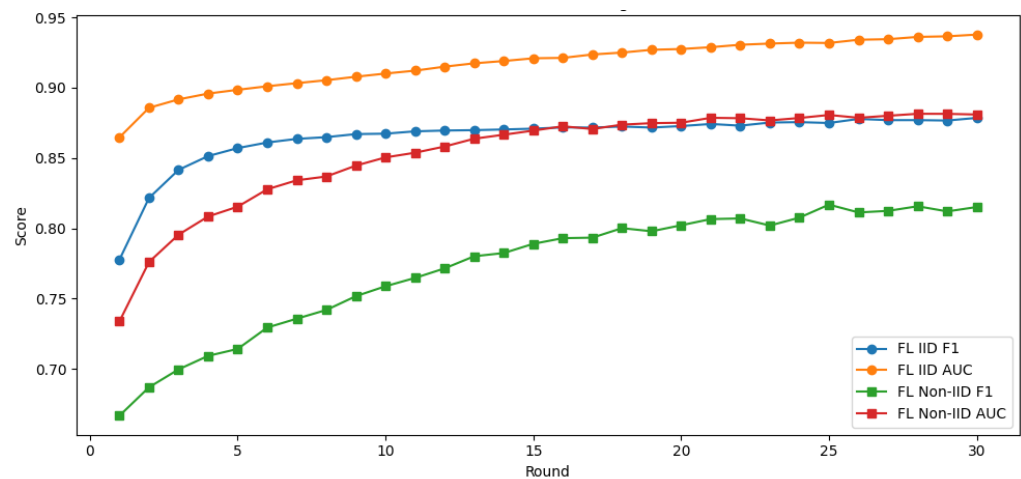


Figure 4 Convergence of federated MLP models (IID vs non-IID)

Feature Importance Analysis

The most important features identified by the Extra Trees model are presented in figure 5, highlighting the key variables that contribute most significantly to the intrusion detection task. Among these, features such as num_failed_logins, num_login_attempts, and num_ip_reputation_score emerge as the top contributors, as they are closely associated with abnormal user behavior and potential network anomalies. The prominence of these features indicates that the model effectively captures critical cybersecurity indicators, particularly those related to authentication patterns and reputation-based risk signals, thereby reinforcing its ability to distinguish between normal and malicious activities in the network.

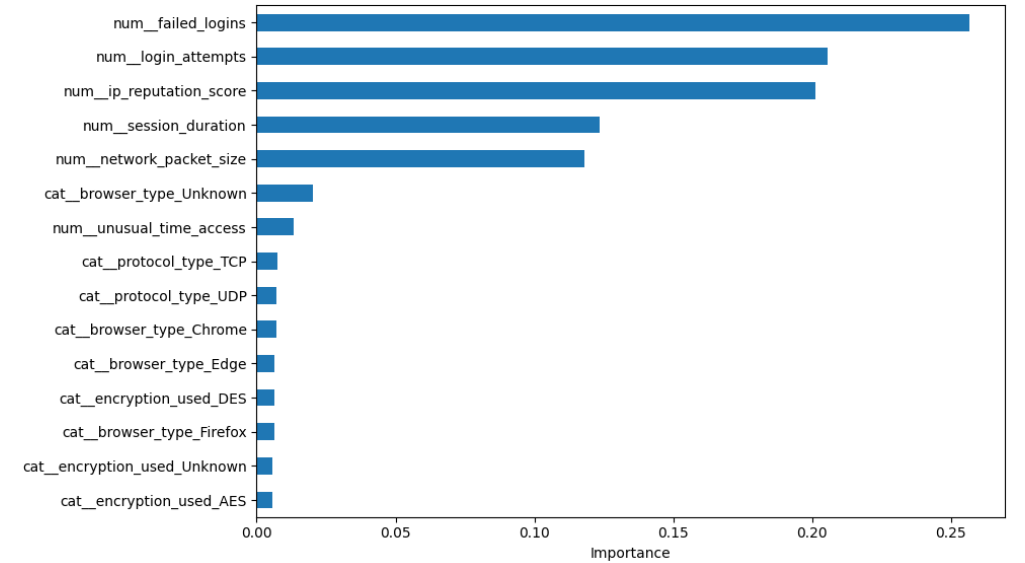


Figure 5 Top 15 feature importances (Extra Trees model)

Confusion Matrix Analysis

The confusion matrix of the best performing model, namely the Random Forest, is presented in figure 6, providing a detailed view of its classification outcomes. The results show a very low false positive rate, indicating that the model is highly effective at correctly identifying normal traffic without incorrectly flagging it as an attack. However, there is still a moderate number of false negatives, meaning that some actual attack instances are not detected by the model. This pattern suggests that the model is more optimized for precision than recall, prioritizing the accuracy of positive predictions over the complete detection of all intrusion events, which could result in certain attacks going unnoticed in practical deployment scenarios.

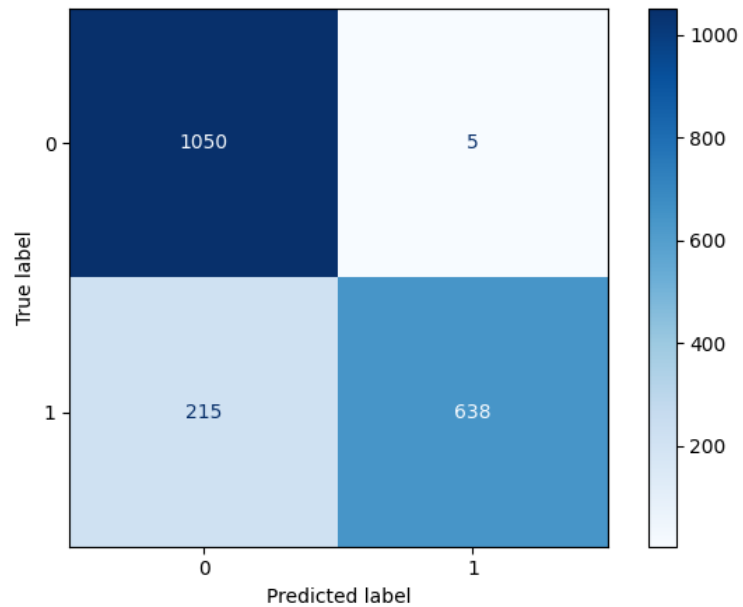


Figure 6 Confusion matrix of the centralized Random Forest model

Summary of Findings

Overall, the results indicate that centralized models still achieve the highest performance in intrusion detection tasks, benefiting from direct access to complete and uniformly distributed data. However, federated learning under IID conditions is able to deliver comparable performance, demonstrating its potential as a privacy preserving alternative without significant loss in effectiveness. On the other hand, the noticeable performance degradation observed in non IID settings underscores the limitations of current federated approaches when dealing with heterogeneous data, highlighting the need for more robust optimization strategies and adaptive techniques to improve model generalization across diverse client distributions.

Discussion

The results demonstrate that centralized models still provide the highest overall performance, with the tuned Random Forest achieving the best F1-score and ROC-AUC. However, the Federated MLP with FedProx under IID conditions shows highly competitive performance, indicating that federated learning can approximate centralized detection capability without requiring data centralization. This suggests that federated learning offers a viable trade-off between performance and privacy, particularly in environments where data sharing is restricted. Nevertheless, the significant performance degradation observed in the non-IID setting confirms that data heterogeneity remains a major challenge in federated systems. Although FedProx improves training stability, it does not fully resolve the imbalance in data distribution across clients. In addition, the models consistently exhibit high precision but moderate recall, indicating that while predictions are highly reliable, a portion of actual attacks remains undetected. This trade-off highlights an important limitation in intrusion detection systems, where missing attacks may pose greater risks than false alarms depending on the operational context.

From a data protection and cyber law perspective, federated learning aligns with

key principles such as data minimization and privacy by design, since raw data remains on local devices and is not transmitted to a central server. This significantly reduces the risk of data breaches and unauthorized access, making federated approaches more suitable for sensitive domains. However, privacy preservation alone is not sufficient to ensure compliance and accountability. Federated systems still face risks such as information leakage from model updates and vulnerabilities to adversarial manipulation, which may undermine both security and trust. Furthermore, the limited explainability of deep learning models raises concerns regarding transparency and legal accountability, particularly in scenarios where automated decisions must be justified. Therefore, while federated learning provides a strong foundation for privacy-preserving intrusion detection, it must be complemented with additional mechanisms such as explainable AI, secure aggregation, and auditability frameworks to meet both technical and legal requirements.

Conclusion

This study demonstrates that federated learning is a promising approach for privacy-preserving intrusion detection, achieving performance comparable to centralized models under IID data conditions while significantly reducing the need for raw data sharing. However, the results also reveal important limitations, particularly the performance degradation under non-IID data distributions and the trade-off between high precision and moderate recall. From a cyber law perspective, federated learning aligns well with key data protection principles such as data minimization and privacy by design, but it does not fully address challenges related to security risks, model robustness, and AI accountability. Therefore, while federated learning provides a strong foundation for secure and privacy-aware intrusion detection systems, its practical adoption requires the integration of additional mechanisms such as explainability, secure aggregation, and governance frameworks to ensure both technical effectiveness and legal compliance.

Declarations

Author Contributions

Conceptualization: T.E.A. and B.D.A.; Methodology: B.D.A.; Software: T.E.A.; Validation: T.E.A. and B.D.A.; Formal Analysis: T.E.A. and B.D.A.; Investigation: T.E.A.; Resources: B.D.A.; Data Curation: B.D.A.; Writing Original Draft Preparation: T.E.A. and B.D.A.; Writing Review and Editing: B.D.A. and T.E.A.; Visualization: T.E.A.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] J. T. Heaton, "Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, 2016, 800 pp, ISBN: 0262035618," *Genetic Programming and Evolvable Machines*, vol. 19, no. 1–2, pp. 305–307, Oct. 2017, doi: 10.1007/s10710-017-9314-z.
- [2] Y. Xin et al., "Machine learning and deep learning methods for cybersecurity," *IEEE Access*, vol. 6, no. May, pp. 35365–35381, 2018, doi: 10.1109/ACCESS.2018.2836950.
- [3] A. Y. Javaid, Q. Niyaz, and W. Sun, "A deep learning approach for network intrusion detection system," *EAI Endorsed Transactions on Security and Safety*, vol. 3, no. 9, pp. 1-8, Dec. 2015, doi: 10.4108/eai.3-12-2015.2262516.
- [4] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, vol. 54, 2017.
- [5] K. Bonawitz, V. Ivanov, B. Kreuter, and A. Marcedone, "Practical secure aggregation for privacy-preserving machine learning," in *Proc. 2017 ACM SIGSAC Conf. Computer and Communications Security (CCS)*, vol. 2017, no. October, pp. 1175 – 1191, Oct. 2017, doi: 10.1145/3133956.3133982.
- [6] M. Ring, S. Wunderlich, D. Scheuring, and D. Landes, "A survey of network-based intrusion detection data sets," *Computers & Security*, vol. 86, no. September, pp. 147–167, Jun. 2019, doi: 10.1016/j.cose.2019.06.005.
- [7] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in *Proc. Military Communications and Information Systems Conf. (MilCIS)*, Canberra, Australia, vol. 2015, no. December, pp. 1-6, Nov. 2015, doi: 10.1109/MilCIS.2015.7348942.
- [8] Z. He, D. Davila, S. Bi, and T. Wang, "Machine learning for cybersecurity: A survey of applications, adversarial challenges, and future research directions," *Electronics*, vol. 14, no. 23, p. 4563, pp. 1-28, Nov. 2025, doi: 10.3390/electronics14234563.
- [9] S. Zawoad and R. Hasan, "Cloud forensics: A meta-study of challenges, approaches, and open problems," *arXiv preprint arXiv:1302.6312*, vol. 2013, no. February, pp. 1-15, Feb. 2013.
- [10] R. Roman, J. Lopez, and M. Mambo, "Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges," *Future Generation Computer Systems*, vol. 78, no. January, pp. 680–698, Feb. 2018, doi: 10.1016/j.future.2016.11.009.
- [11] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *arXiv preprint arXiv:1902.04885*, vol. 2019, no. February, pp. 1-19, Feb. 2019, doi: 10.48550/arXiv.1902.04885.

- [12] J. Konečný, H. B. McMahan, F. X. Yu, and P. Richtárik, "Federated learning: Strategies for improving communication efficiency," *arXiv preprint arXiv:1610.05492*, vol. 2016, no. October, pp. 1-10, Oct. 2016, doi: 10.48550/arXiv.1610.05492.
- [13] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, et al., "Advances and open problems in federated learning," *arXiv preprint arXiv:1912.04977*, vol. 2019, no. December, pp. 1-121, Dec. 2019, doi: 10.48550/arXiv.1912.04977.
- [14] S. J. Hardy, W. Henecka, H. Ivey-Law, R. Nock, et al., "Private federated learning on vertically partitioned data via entity resolution and additively homomorphic encryption," *arXiv preprint arXiv:1711.10677*, vol. 2017, no. November, pp. 1-60, Nov. 2017, doi: 10.48550/arXiv.1711.10677.
- [15] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, et al., "Advances and open problems in federated learning," *arXiv preprint arXiv:1912.04977*, vol. 2019, no. December, pp. 1-121, Dec. 2019, doi: 10.48550/arXiv.1912.04977.
- [16] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *arXiv preprint arXiv:1812.06127*, vol. 2020, no. April, pp. 1-22, Apr. 2020, doi: 10.48550/arXiv.1812.06127.
- [17] W. Wang, Y. Sheng, J. Wang, and X. Zeng, "HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection," *IEEE Access*, vol. 6, no. December, pp. 1792–1806, Dec. 2017, doi: 10.1109/ACCESS.2017.2780250.
- [18] J. Kim, J. Kim, T.-T.-H. Le, and H. Kim, "Long short term memory recurrent neural network classifier for intrusion detection," in *Proc. 2016 Int. Conf. Platform Technology and Service (PlatCon)*, vol. 2016, no. April, pp. 1-5, Feb. 2016, doi: 10.1109/PlatCon.2016.7456805.
- [19] Y. Tang, Y. Zhang, T. Niu, and Z. Li, "A survey on blockchain-based federated learning: Categorization, application and analysis," *Computer Modeling in Engineering & Sciences*, vol. 139, no. 3, pp. 2451–2477, Mar. 2024, doi: 10.32604/cmescs.2024.030084.
- [20] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, "How to backdoor federated learning," *arXiv preprint arXiv:1807.00459*, vol. 2019, no. August, pp. 1-15, Aug. 2019, doi: 10.48550/arXiv.1807.00459.