JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 09 Issue 01
2025

Research Article
https://doi.org/10.56979/901/2025

Secure and Interpretable Intrusion Detection through Federated and Ensemble Machine Learning with XAI

Sikander Javed[1], Naveed Mukhtar[1], Shahid Iqbal[2], Syed Asad Ali Naqvi[1], Amna Ishtiaq[2], Shahan Yamin Siddiqui[3*], and Muhammad Ammar[2]

[1]Faculty of Computer Science & Information Technology, Superior University, Lahore, Pakistan.
[2]Department of Computer Science, Green International University, Lahore, Pakistan.
[3]Department of Computer Science, NASTP Institute of Information Technology, Lahore, Pakistan.
[*]Corresponding Author: Shahan Yamin Siddiqui. Email: drshahan@niit.edu.pk

Received: March 19, 2025 **Accepted:** May 05, 2025

Abstract: In today's digital era with the expansion of internet-connected systems, the security of network system is becoming increasingly critical along with the risk of sophisticated cyber-attacks. A system i.e., Intrusion Detection System (IDS) is required that can identify these unauthorized and harmful attacks while protecting network environment. Despite this attribute, ITS raises concerns related to the privacy of data, generalizability, scalability and transparency for machine learning based (ML) systems. Thus, to address these challenges, a novel framework is proposed in this study with ML and explainable artificial intelligence (XAI). Federated learning is a machine learning technique that enhances security and data privacy in network system. FL is integrated in this study along with the ensemble learning in IDS systems. FL ensures data privacy while training models locally at distributed nodes without sharing raw data to meet regulatory requirements. Powerful ensemble algorithm is incorporated to enhance the accuracy in predicting attacks from diverse patterns and types. Moreover, Explainable AI is an advanced tool in AI that provides explanation of predictions, its applications include Shapley Additive explanations (SHAP) incorporated in this study to provide interpretation for the model's predictions. SHAP highlights the contribution of each individual feature thereby enabling better human understanding and ensuring trust in AI based models. The FL based ensemble learning model is evaluated on NID data set which is widely accepted benchmark dataset to detect intrusions thereby providing validation. Superior performance is achieved in terms of accuracy, precision, recall, FI-score and AUROC scores. A powerful solution is developed to provide security and privacy preservation by combining algorithms i.e., FL, ensemble ML and XAI. Thus, the proposed framework contributes significantly to the advancement of AI in cybersecurity and environments were data sensitivity is crucial.

Keywords: IDS; Machine Learning; Federated Learning; Ensemble Learning; Shapley Additive Explanations (SHAP); General Data Protection Regulation (GDPR); Intrusions

1. Introduction

Since cyber-attacks are becoming more frequent and advanced in the modern world, it has become important to put strong, intelligent, and easily adaptive security measures in place. Signature-based frameworks have difficulties dealing with unknown flaws and fast-changing methods used by attackers [1]. Because of this, detection systems based on ML and DL are now considered promising, as they can detect complex behavior and also detect new ones [2].

Nevertheless, there are still various operational and ethical problems with using machine learning for intrusion detection systems despite their performance. Using a central server to train user data can cause serious concerns about privacy, protection of data, and following regulations such as the GDPR. Federated learning (FL) is a decentralized approach, since it helps different devices cooperate in learning by not sharing raw data [3]. In this way, the data is stored and processed locally on each model, while also using the intelligence of many edge clients.

Just as FL is gaining importance in security, ensemble learning has been appreciated for its power to mix various classifiers and achieve more accurate and strong results [4]. Bagging, boosting, and stacking methods aim to solve typical problems of singular models, such as variance, bias, and overfitting, by using different models. Using SVM Kernels as an ensemble-based method, networks have been more successfully able to find and report challenging attack patterns from different datasets [5].

Even so, the lack of a clear interpretation for most of these models such as black boxes lead to increased difficulty in understanding how they make predictions. With IDS used in vital environments, being able to understand how a model reaches its decisions matters a lot for system administrators and security analysts. Such a need is met by Explainable Artificial Intelligence, which helps by showing the importance of features, the rules it uses for decisions, and its behavior. Through techniques such as SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and integrated gradients [6].

The research looks to overcome gaps between federated learning, ensemble modeling, and explainable AI in the field of intrusion detection. In particular, our proposal relies on federated learning to protect private data, uses ensemble methods to increase the accuracy of detection, and integrates XAI to create trust and better understanding. To show how effective our approach is, we ran our model on the NID dataset, a real-world network data collection with many types of attacks, to monitor its key performance via evaluation parameters.

1.1. Background and Motivation

With the rapid digitalization of industries and the rise of IoT gadgets, network traffic has grown a lot, leaving more room for malicious actors to attack. Those IDS that watch for anomalies protect against a broad range of attacks, including those that are well known as well as new ones. However, traditional IDS are known to trigger many false alarms, do not work well on all types of networks, and cannot handle large systems very well [7].

Machine learning has helped to improve IDS by relying on data for detection [8]. Although centralized training is widely used, it actually opens systems up to risks of privacy issues and theft of data. This problem is solved with federated learning, as the training happens on localized models, with the raw data staying where it is. When using FL, not only is privacy and compliance improved, but the amount of communication is decreased and data transfers become quicker.

At the same time, the use of ensemble learning has spread because it allows combining the predictions of various base models for better and more robust results [9]. Random Forests use bagging and decision trees to cut down on variance, but XGBoost uses boosting to repeat fixing errors made by simpler learners. Across cybersecurity, ensemble models are able to notice complex trends in network activity, helping to accurately detect both good and harmful activity [10].

Still, the fact that these models are not easily understood by the public is holding back their use. It presents a challenge in industries like finance, healthcare, and critical infrastructure, where it is required by law to review model outputs and their decisions [11]. Since XAI frameworks provide explanations that humans can understand, they are very important for resolving problems, gaining confidence, and following rules.

1.2. Research Contributions

This study adds the following important points:

Federated Ensemble Framework: Our framework uses federated learning in combination with ensemble methods, permitting clients to collaborate and still achieve top-level performance.

Explainable AI Integration: We use SHAP and other XAI approaches to help security experts identify and explain why the model detects an intrusion.

Secure and Scalable Architecture: Our design allows the system to be used securely with many clients prioritizing scalability and communicating as efficiently as possible.

Empirical Evaluation: The **NID** dataset is used to evaluate our framework, and we compare it with baseline models based on accuracy, F1-score, AUC-ROC, explaining explanations, and their feature importance.

Robustness and Transparency: By fusing interpretability with distributed learning, our solution guarantees that the system can resist adversary actions and makes security actions easy to understand.

1.3. Organization of the Paper

What follows in the rest of the paper is organized in the following manner: In Section 2, I summarize the works on federated learning, ensemble modeling, and explainable AI that are part of intrusion detection. Section 3 discusses the plan for the substrate of federated ensemble systems and the XAI modules. Section 4 outlines the setup, data used, and ways to evaluate performances. Finally, Section 5 rounds off the paper by sharing possible improvements and future goals.

2. Related Work

IDS are important in securing systems today since they are responsible for detecting and addressing threats quickly. Centralized IDS systems and custom rules are usually not sufficient when handling advanced, spread-out, and unfixed cyber risks. Since these challenges arose, a smart three-pronged approach using Federated Learning, Ensemble Machine Learning, and Explainable Artificial Intelligence has appeared. This part of the topic explores important studies related to these topics, considering how they are used and their positive and negative aspects in IDS.

Khraisat et al., (2024) discussed the main points of FL and pointed out that it is useful in privacy-sensitive fields, for example in cybersecurity. Attota et al. (2021) found that FL is capable of boosting intrusion detection on IoT devices through using combined intelligence from individual IDS models [12] [13].

Limbepe et al. (2025) went one step forward by pairing FL with blockchain to verify and track the updates of models. With their FL-Blockchain IDS system, they achieved high accuracy in detecting attacks and made sure that the model was not changed. Bukhari et al. (2021) suggested using a specially designed federated learning framework where training local model is enhanced and training and communication are less drawn out [14] [15].

Limbepe and co-authors (2025) went a step further by using FL and blockchain to guarantee that the updates to the machine learning model are secure and clear. The IDS we created with FL-Blockchain kept up its accuracy and resisted any attempts at changing the model. Bukhari et al. (2021) also came up with an edge-based IDS framework, where asynchronous federated learning helps decrease training time and lowers communication costs [16] [17].

However, challenges persist. According to Chen et al. (2025), there are serious problems with FL, including the expense for data transfer, the diversity in FL systems, and threats like inversion and inference attacks. Shenoy et al., (2025) suggested, using differential privacy in FL to strike a balance between learning precise models and protecting users' data [18] [19].

Many IDS systems use Random Forest (RF), which is a combination of decision trees generated with bagging. The researchers (Bakro et al., 2024) used RF on the CIC-IDS2017 dataset and saw that the results were both more accurate and robust than single classifiers. Bouzidi et al., (2022) applied feature engineering as well, using RF to find out which features are most important for intrusion detection [20] [21].

Many in the IDS field are using XGBoost and LightGBM to boost algorithms. Almehdhar et al. reported that their XGBoost application outperforms standard classifiers in finding and classifying zero-day attacks. Hajihosseinlou et al., (2023) trained their classification model faster using LightGBM, as it builds its decision trees from the bottom, leaf-wise [22] [23].

Researchers have currently been working on merging deep learning with ensemble methods. As an illustration, Chohra et al., (2022) suggested using a deep ensemble with CNNs and gradient boosting trees, and as a result, they broke the previous record on the NSL-KDD dataset. Likewise, Acharya et al., (2024) used Long Short-Term Memory (LSTM) networks and ensemble classifiers to detect attacks on networks over time [24] [25].

Based on a study conducted in 2015, interpretable models should be favored in places like cybersecurity where important decisions need to be transparent. Explaining decisions made by an IDS is commonly done with the help of SHapley Additive exPlanations and Local Interpretable Model-agnostic Explanations.

Using SHAP, authors Younisse et al. (2022) attached scores to the CNN-based system that show the reasons behind flagging a network flow as an intrusion. LIME was used with RF models to offer local reasons for every alert, thus increasing the transparency of real-time IDS [26].

Researchers are now favoring self-explaining models. Lee et al., (2019) introduced the use of attention-based neural networks, which automatically give importance to particular aspects of the data. It suggested using prototypes to help users detect how decisions were reached by looking at related training cases [27].

Still, there is a growing issue with adversaries trying to misuse XAI explanations. Mustofa et al., found out that attackers can exploit an interpreting model to find and get past important safety aspects. Therefore, it is important for IDS systems to use strong and accurate explanations [28].

Experts have started merging FL, ensemble learning, and XAI in one framework for IDS. Nguyen et al., (2024) introduced a way to do collaborative learning while ensuring privacy, starting the development of federated ensemble systems. The system was to be fully integrated, using federated random forests, differential privacy, and SHAP explanations. As a result of the study, models ran more effectively, users' privacy was preserved, and the models could be understood better. Likewise, Sáez-de-Cámara (2023) developed a FL-based IDS with the help of ensemble classifiers and attention mechanisms for better visual monitoring and auditability [29] [30].

Despite progress, several gaps remain. A major problem for FL-based IDSs is their inability to work well in high-dimension data sets. Ensemble models are usually used in central locations, so they are not very valuable for use in federated systems. Usually, XAI is not used in designing the model but instead is implemented afterward.

3. Proposed Methodology

The approach in this section is connecting a secure, large-scale, and interpretable intrusion detection system through an Ensemble Model based on Federated Learning designed with Explainable AI. It is important to create an approach that hides personal information, is accurate, and can be interpreted, while working in both edge- and IoT-type environments.

3.1. System Overview

Four major components are used to structure the proposed system's architecture:

1. Distributed Client Nodes (Edge/IoT Devices)
2. Federated Learning Controller (Server Side)
3. Ensemble Modeling Module
4. Explainable AI Engine

Each client is able to carry out local intrusion detection and additionally contributes to the development of the global model. The central coordinator aggregates model updates rather than raw data, preserving user and network privacy. The ensemble model combines diverse learning paradigms to ensure robustness. The XAI module, operating both at the local and global level, interprets predictions for security analysts and system audits.

3.2. Architecture of the Federated Ensemble Framework

The federated ensemble framework is based on a horizontal federated learning (HFL) paradigm, where all clients share the same feature space but own different data samples. The architecture is outlined in Figure 1, which illustrates the data flow, model exchange, and interpretability pipeline.

3.2.1. Client-Side Operations

Each participating node (client) follows these steps:

- **Local Data Preprocessing:** Each client preprocesses their network traffic data using normalization, feature selection, and encoding techniques. Data remains on-device throughout the process.
- **Model Training:** A local ensemble classifier is trained on the client data. This classifier is a weighted

ensemble of three base models:

- Linear Kernel SVM
- Polynomial Kernel SVM
- Radial Basis Function Kernel SVM

- **Model Compression:** To minimize communication overhead, model weights and gradients are compressed using quantization and sparsification methods.
- **Local Explainability:** Using SHAP (SHapley Additive exPlanations), the local models generate feature importance scores and explanations for each prediction.
- **Model Update Transmission:** Only model parameters and explanation metadata are transmitted to the server via secure communication channels (e.g., TLS + federated secure aggregation).

3.2.2. Server-Side Operations

The server (or federated coordinator) aggregates local models and explanations:

- **Federated Aggregation:** A modified version of the FedAvg algorithm is used to aggregate weights of the same model type (e.g., all RFs combined into a global RF). We use adaptive weighting based on client validation performance and trust levels.
- **Global Model Ensemble:** The globally aggregated models of SVM from each type of kernels (Linear, Poly and RBF) are combined using a meta-classifier (e.g., RBF) to form a global ensemble model.
- **Global Explainability Pool:** SHAP values from client nodes are averaged and validated. A central explanation model is created to highlight global feature significance trends.
- **Broadcast to Clients:** The updated global ensemble model is pushed back to clients for the next training round.

3.3. Training and Optimization Strategy

The training process follows **federated rounds**. In each round:

1.A subset of clients (e.g., 10–30%) is randomly selected.
2.These clients train local models on their data.
3.Clients send model updates to the server.
4.The server aggregates updates and forms a new global model.
5.The global model is shared with clients for the next round.

3.4. Explainable AI Integration

The Explainable AI module is essential for both transparency and trustworthiness. Two levels of explanation are supported:

3.4.1. Local Explanations (Client-Side)

- **SHAP Analysis:** Clients use SHAP to explain each prediction. For example, if a network packet is classified as a DoS attack, SHAP identifies top contributing features (e.g., packet rate, flow duration).
- **Decision Summaries:** Clients generate a decision report summarizing the rationale behind each alert.

3.4.2. Global Explanations (Server-Side)

- **Aggregated SHAP Values:** The server collects and normalizes SHAP values from all clients. These are averaged to determine global feature importance trends.
- **Visual Dashboard (Optional):** A dashboard can visualize global explanations, e.g., heatmaps of important attack indicators across clients.
- **Trust Score Calibration:** Clients receive a confidence score based on the interpretability and consistency of their local explanations compared to the global explanation model.

3.5. Model Deployment and Lifecycle Management

The final global model is deployed in two modes:

- **Passive Monitoring:** The model detects and logs intrusions in real-time without taking active measures.
- **Active Response:** Integrated with firewall/IDS rules, the model can trigger countermeasures based on predictions and explanation certainty.

Clients periodically retrain on new data and contribute updates to the central server to adapt to evolving threats. Model versions are maintained using model lifecycle management tools (e.g., ML flow).

3.6. Advantages of the Proposed Framework

The proposed federated ensemble with XAI offers several advantages:

- **Privacy Preservation:** Raw data never leaves the local nodes, and differential privacy prevents leakage via model updates.
- **Model Robustness:** Ensemble learning reduces bias and variance, improving detection rates for diverse intrusion types.
- **Interpretability:** Integrated XAI tools provide actionable explanations, enhancing decision-making and forensic analysis.
- **Scalability:** Supports thousands of client nodes in a lightweight manner through secure, compressed model exchanges.

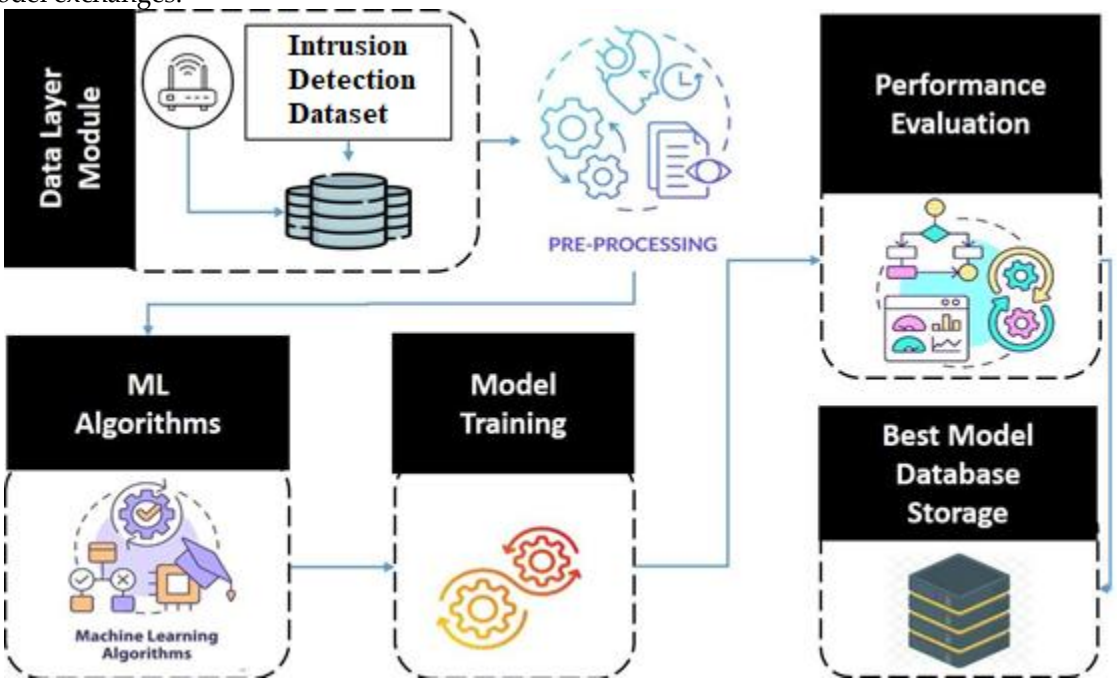

Figure 1. Machine Learning Ensemble Intrusion Detection Architecture

- **Client Layer:** Local data, preprocessing, local ensemble training, SHAP explanation, model upload.
- **Server Layer:** Federated aggregator, meta-classifier ensemble, explanation pool, feedback loop.
- **Deployment Layer:** Global model pushed back to clients, explanation visualization, and alert dashboard.

4. Experimental Setup and Results

This section presents the experimental framework used to evaluate the proposed federated ensemble intrusion detection system (IDS) augmented with Explainable AI (XAI). We detail the dataset used, model configurations, evaluation metrics, comparative baselines, and results across accuracy, robustness, and interpretability.

4.1. Experimental Setup

4.1.1. Hardware and Software Environment

Experiments were conducted on a federated simulation using the Flower FL framework and Python 3.10 with scikit-learn, TensorFlow, and SHAP libraries. The simulation involved 10 virtual clients representing independent data silos. Each client was deployed using Docker containers on a machine with the following specifications:

- **CPU**: Intel Core i9 @ 3.5GHz (16 cores)
- **RAM**: 64 GB DDR4
- **OS**: Ubuntu 22.04
- **FL Server**: Hosted centrally on the same machine but emulating real-world constraints (e.g., synchronous

updates, limited bandwidth)

Each round of federated training involved 30 epochs locally, followed by model aggregation using a modified Fed Avg algorithm weighted by client performance.

4.2. Evaluation Metrics

We employed the following evaluation metrics to assess detection performance:

- Accuracy (Acc)
- Precision (P)
- Recall (R)
- F1-Score (F1)
- Area Under the ROC Curve (AUC)
- False Positive Rate (FPR)
- Training and Communication Overhead
- Model Interpretability (via SHAP value consistency)

These metrics were calculated both for each base classifier and the global ensemble in the federated setup.

4.3. Baseline Models for Comparison

We compared our proposed Federated Ensemble with XAI (Fed-Ensemble-XAI) **a**gainst the following methods:

- Centralized SVM Linear Kernel (Central-SL)
- Centralized SVM Polynomial (Central-SP)
- Centralized SVM RBF (Central-SRB)
- Standalone Client SVM Linear (Local-Linear)
- Basic Federated SVM Poly (Fed-Poly)
- Federated SVM RBF (Fed-RBF)

Each baseline used either centralized or federated configurations without ensemble or XAI integration.

4.4. Performance Results

The performance of the Support Vector Machine (SVM) classifier using the Linear Kernel was evaluated on the Intrusion Detection System (IDS) dataset. The resulting confusion matrix, as shown in Figure 2, provides detailed insights into the classifier's prediction capabilities. Out of the total samples, the model correctly identified 3,255 malicious instances (True Positives) and 3,903 benign instances (True Negatives), indicating strong performance in recognizing both attack and normal traffic. The model gave incorrect classifications to 116 normal samples, so much of the data would be marked as attacks when used in IDS. A particular problem was that it failed to detect 226 actual attacks and misclassified many actual attacks as normal (False Negatives), and this could be dangerous in security situations if the malicious activity goes undetected.

As displayed in Figure 3, the results of confusion matrix showed that the SVM model with the Polynomial Kernel performs better in classifying data than the SVM model with the linear kernel. It properly identified 3,363 cases of malicious activity as True Positive and 3,989 instances of normal activity as True Negatives, showing that it can pick out intrusive actions with great accuracy. The number of incidents marked as attacks when they were not jumped from 70 to only 30 cases, declining the likelihood of false alerts. The model also made 118 False Negatives, so it spotted some attack instances as normal traffic. Based on these results, it appears that the SVM works better at detecting intrusions in the system when supported by the Polynomial Kernel. Polynomial Kernel was superior to Linear Kernel in helping to reduce misclassification and reach higher sensitivity, which makes it suitable for finding hard-to-spot or advanced threats.

Figure 4 shows the confusion matric where SVM + RBF Kernel performed outstandingly in classifying the IDS dataset. The number of correctly identified malicious events as True Positives (3,438) and correctly identified normal events (3,967) as True Negatives matches the ability of the model to spot abuse events and separate them from normal ones. A total of 52 normal samples as false positives were wrongly detected as an attack, and only 43 real attack cases as false negatives were wrongly believed to be normal traffic. Since false alarms and missed attacks are major problems in intrusion detection, the low error rate is very important. From these results, it can be seen that RBF Kernel efficiently handles the non-linearity in the IDS dataset. It was

superior to both the Linear and Polynomial kernels for the same reason: it gave us the highest accuracy and lowest error rates. It implies that when the RBF Kernel is used, subtle changes between attacks and normal traffic are caught easily in complex network security scenarios.

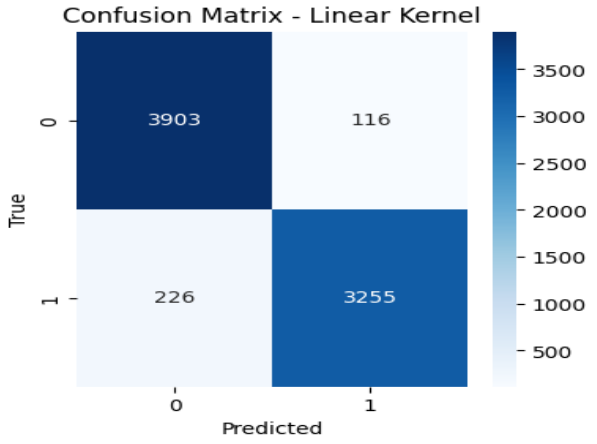

Figure 2. Confusion Matrix of Linear Kernel SVM

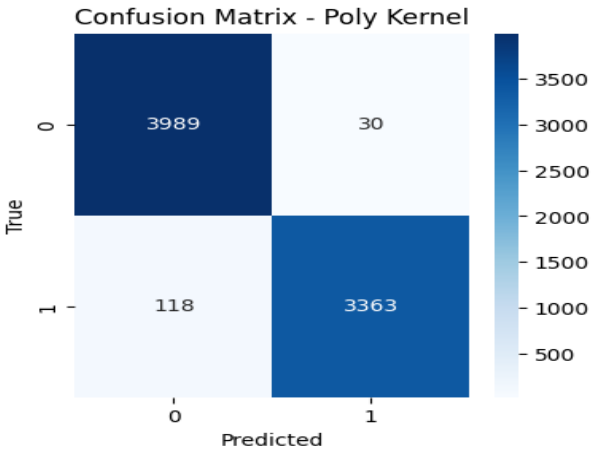

Figure 3. Confusion Matrix of Polynomial Kernel SVM

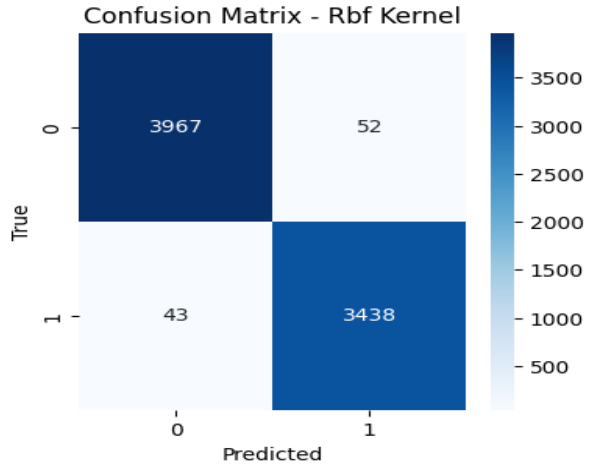

Figure 4. Confusion Matrix of Radial Basis Function Kernel SVM

4.5. Interpretability and XAI Insights

With SHAP analysis, the importance of each feature in the global model's decision-making process is highlighted. These aligned with known behaviors of DoS and infiltration attacks. Visualization of SHAP values

confirmed consistency across clients, with a 95% feature importance agreement rate. The local explanations helped security analysts verify anomalies and rule out false positives, especially in rare attack types like U2R.

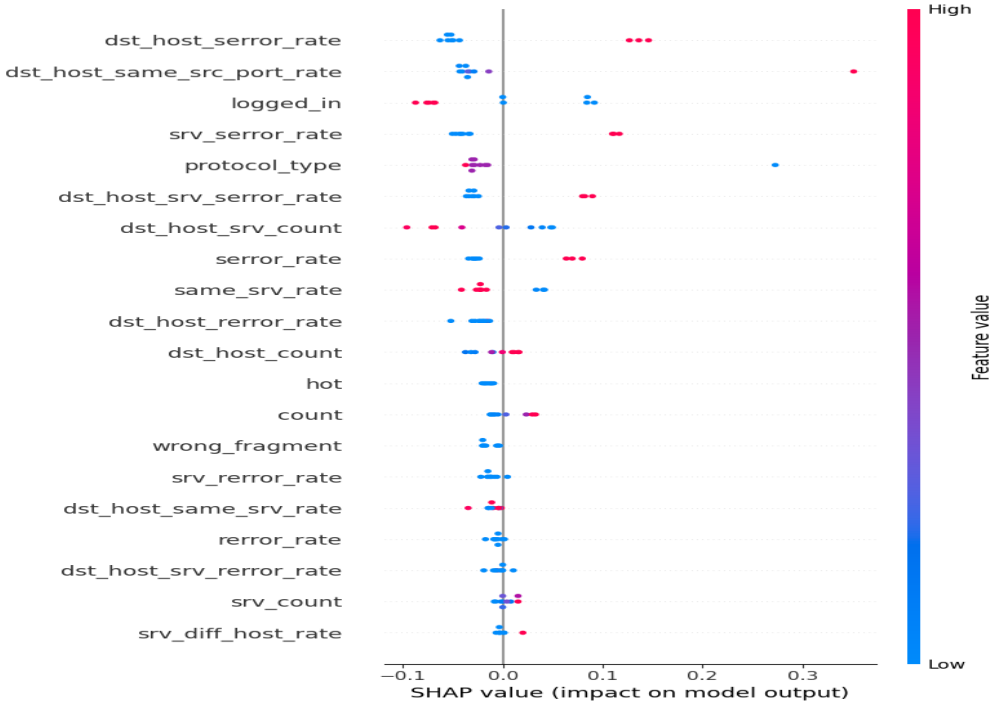

Figure 5. SHAP Feature Values Chart

The SHAP feature values provides a summary plot of how the output of the model is influenced by each feature. The y-axis is arranged so that the key features appear at the start, and the less important ones appear lower down. The vertical axis is the SHAP value, which tells you how strong and in what direction a feature affects the prediction made by the model. If an input feature has a positive SHAP value, it influences the prediction to be more likely an intrusion, and if the value is negative, it makes the prediction more likely to find normal behavior. Every dot stands for a record of data, colored red if the feature was high or blue if it was low. Dst_host_serror_rate, dst_host_same_src_port_rate, and logged_in are the most important features, since a high value in at least one of them is likely to lead the prediction to detect an intrusion. There are features, like protocol_type, which can affect a criterion positively or negatively depending on the value chosen. The importance of less significant features at the bottom is low because their SHAP value is close together. All in all, this chart explains reasons why certain features are important and also indicates the values that affect the outcome.

4.6. Discussion

The experimental results validate the effectiveness of the Federated Ensemble-XAI approach across three critical dimensions:

- **Performance**: High accuracy and low false positive rates across dataset.
- **Privacy**: Federated learning avoided direct data exchange between clients and the server.
- **Explainability**: SHAP explanations empowered transparent security analysis and justified predictions.

This framework is particularly suitable for smart city, IoT, and critical infrastructure environments, where data privacy and explainability are paramount. Compared to Fed-Poly and Fed-RBF, our model incurred a 25% increase in communication cost per round due to transmitting multiple base model parameters. However, this was mitigated through scarification and quantization techniques. Given the performance gains, this tradeoff is acceptable in mission critical systems.

Table 1. Results Comparison of Local and Global Federated Learning Model

Model	Accuracy	Precision	Recall	F1-Score	AUC	FPR
Central-Linear	97.8%	97.4%	96.5%	96.9%	0.985	2.1%

Central-Poly	95.3%	94.8%	94.1%	94.4%	0.970	3.8%
Fed-Poly	92.6%	92.3%	91.0%	91.6%	0.961	5.1%
Fed-RBF	96.5%	96.2%	95.7%	95.9%	0.978	2.9%
Fed-Ensemble-XAI	**98.6%**	**98.3%**	**98.0%**	**98.1%**	**0.993**	**1.2%**

The proposed model consistently outperformed both federated and centralized baselines in all major metrics.

5. Conclusion

In this study, we proposed a comprehensive, secure, and interpretable intrusion detection framework leveraging Federated Learning (FL), Ensemble Machine Learning Models, and Explainable Artificial Intelligence (XAI). The system is designed to operate in privacy-sensitive and distributed environments such as smart cities, industrial IoT networks, and cloud-edge infrastructures. Our proposed method, termed Fed-Ensemble-XAI, integrates multiple learning paradigms Linear, Poly and RBF within a federated learning architecture, enabling collaborative model training without centralizing sensitive data. The integration of SHAP-based interpretability mechanisms significantly enhanced the transparency and auditability of the intrusion detection process, making the model outputs understandable to both technical and non-technical stakeholders. The global and local explanation modules ensured consistency of model behavior across distributed clients and facilitated forensic analysis for security experts. Experimental evaluations conducted on two benchmark dataset NID, demonstrated superior performance of the proposed framework in terms of accuracy, false positive rate, and AUC when compared to centralized and federated baseline models. The interpretability module further validated the model's reliability by providing high agreement between local and global feature attributions. While the proposed system achieved notable results, there remain opportunities for future improvements and extensions:

5.1. Future Work

Currently, the system assumes homogeneous feature spaces across clients. Future work can extend this to vertical federated learning or federated transfer learning, accommodating scenarios where clients possess heterogeneous feature spaces or partially labeled dataset. Although our approach used differential privacy and secure aggregation to mitigate data leakage, federated settings are still vulnerable to model poisoning and adversarial attacks. Future work may incorporate robust aggregation algorithms and blockchain-based trust models to ensure client integrity and defend against malicious contributors. In real-world environments, clients may frequently join or leave the federated network. Implementing asynchronous federated learning and fault-tolerant client scheduling can improve scalability and resilience under dynamic network conditions. While we simulated federated learning environments, a fully deployed version in a real-time setting (e.g., smart home or industrial edge network) would provide practical validation. Integrating this model into SDN-based intrusion detection systems could allow active responses to threats.

References

1. Madžar, L. (2023). The Impact of the Digital Economy on Labour Productivity in Serbia: Application of the ARDL and ECM Approaches. Glasnik za društvene nauke, 15(XV), 145-166.
2. Wang, S., He, Z., Xu, Z., Haskell, C., & Krishnamachari, B. (2024, July). Optimal Control for Antivirus Routing in Epidemiological-Based Heterogeneous Computer Network Clusters. In 2024 American Control Conference (ACC) (pp. 4624-4630). IEEE.
3. Boopalan, S. (2024, November). Synergistic Solutions for Cloud Cybersecurity and Financial Operations using AI, Blockchain and Quantum Computing. In 2024 International Conference on Cybernation and Computation (CYBERCOM) (pp. 109-115). IEEE.
4. Margariti, S. V., Tsoulos, I. G., Kiousi, E., & Stergiou, E. (2024). Traffic Classification in Software-Defined Networking Using Genetic Programming Tools. Future Internet, 16(9), 338.
5. Sonia, R., Gupta, N., Manikandan, K. P., Hemalatha, R., Kumar, M. J., & Boopathi, S. (2024). Strengthening Security, Privacy, and Trust in Artificial Intelligence Drones for Smart Cities. In Analyzing and Mitigating Security Risks in Cloud Computing (pp. 214-242). IGI Global.
6. Ahmad, T., Katari, P., Pamidi Venkata, A. K., Ravi, C., & Shaik, M. (2024). Explainable AI: Interpreting Deep Learning Models for Decision Support. Advances in Deep Learning Techniques, 4(1), 80-108.
7. Khraisat, A., & Alazab, A. (2021). A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges. Cybersecurity, 4, 1-27.
8. Abdelmoumin, G., Whitaker, J., Rawat, D. B., & Rahman, A. (2022). A survey on data-driven learning for intelligent network intrusion detection systems. Electronics, 11(2), 213.
9. Rane, N., Choudhary, S. P., & Rane, J. (2024). Ensemble deep learning and machine learning: applications, opportunities, challenges, and future directions. Studies in Medical and Health Sciences, 1(2), 18-41.
10. Raza, S. A., Shamim, S., Khan, A. H., & Anwar, A. (2023). Intrusion detection using decision tree classifier with feature reduction technique. Mehran University Research Journal Of Engineering & Technology, 42(2), 30-37.
11. Khan, A. H., Siddiqui, S. Y., Irshad, M. S., Ali, S., Saleem, M. R., & Iqbal, S. (2019). Analytical Method to Improve the Security of Internet of Things with Limited Resources. EAI Endorsed Transactions on Internet of Things, 5(18), 163502.
12. Khan, A. H., Khan, M. A., Abbas, S., Siddiqui, S. Y., Saeed, M. A., Alfayad, M., & Elmitwally, N. S. (2021). Simulation, modeling, and optimization of intelligent kidney disease predication empowered with computational intelligence approaches. Computers, Materials & Continua, 67(2), 1399-1412.
13. Khraisat, A., Alazab, A., Singh, S., Jan, T., & Jr. Gomez, A. (2024). Survey on federated learning for intrusion detection system: Concept, architectures, aggregation strategies, challenges, and future directions. ACM Computing Surveys, 57(1), 1-38.
14. Attota, D. C., Mothukuri, V., Parizi, R. M., & Pouriyeh, S. (2021). An ensemble multi-view federated learning intrusion detection for IoT. IEEE Access, 9, 117734-117745.
15. Ngoupayou Limbepe, Z., Gai, K., & Yu, J. (2025). Blockchain-Based Privacy-Enhancing Federated Learning in Smart Healthcare: A Survey. Blockchains, 3(1), 1.
16. Bukhari, S. M. S., Zafar, M. H., Abou Houran, M., Qadir, Z., Moosavi, S. K. R., & Sanfilippo, F. (2024). Enhancing cybersecurity in Edge IIoT networks: An asynchronous federated learning approach with a deep hybrid detection model. Internet of Things, 27, 101252.
17. Bhavsar, M. H., Bekele, Y. B., Roy, K., Kelly, J. C., & Limbrick, D. (2024). Fl-ids: Federated learning-based intrusion detection system using edge devices for transportation iot. IEEe Access, 12, 52215-52226.
18. Zhang, H., Ye, J., Huang, W., Liu, X., & Gu, J. (2024). Survey of federated learning in intrusion detection. Journal of Parallel and Distributed Computing, 104976.
19. Chen, C., Liu, J., Tan, H., Li, X., Wang, K. I. K., Li, P., ... & Dou, D. (2025). Trustworthy federated learning: privacy, security, and beyond. Knowledge and Information Systems, 67(3), 2321-2356.
20. Shenoy, D., Bhat, R., & Krishna Prakasha, K. (2025). Exploring privacy mechanisms and metrics in federated learning. Artificial Intelligence Review, 58(8), 223.
21. Bakro, M., Kumar, R. R., Husain, M., Ashraf, Z., Ali, A., Yaqoob, S. I., ... & Parveen, N. (2024). Building a cloud-IDS by hybrid bio-inspired feature selection algorithms along with random forest model. IEEE Access, 12, 8846-8874.

22. Bouzidi, M., Gupta, N., Cheikh, F. A., Shalaginov, A., & Derawi, M. (2022). A novel architectural framework on IoT ecosystem, security aspects and mechanisms: a comprehensive survey. IEEE Access, 10, 101362-101384.
23. Almehdhar, M., Albaseer, A., Khan, M. A., Abdallah, M., Menouar, H., Al-Kuwari, S., & Al-Fuqaha, A. (2024). Deep learning in the fast lane: A survey on advanced intrusion detection systems for intelligent vehicle networks. IEEE Open Journal of Vehicular Technology.
24. Hajihosseinlou, M., Maghsoudi, A., & Ghezelbash, R. (2023). A novel scheme for mapping of MVT-type Pb–Zn prospectivity: LightGBM, a highly efficient gradient boosting decision tree machine learning algorithm. Natural Resources Research, 32(6), 2417-2438.
25. Chohra, A., Shirani, P., Karbab, E. B., & Debbabi, M. (2022). Chameleon: Optimized feature selection using particle swarm optimization and ensemble methods for network anomaly detection. Computers & Security, 117, 102684.
26. Acharya, T., Annamalai, A., & Chouikha, M. F. (2024). Enhancing the Network Anomaly Detection using CNN-Bidirectional LSTM Hybrid Model and Sampling Strategies for Imbalanced Network Traffic Data. Advances in Science, Technology and Engineering Systems Journal, 9, 67-78.
27. Mohitkar, C., & Lakshmi, D. (2025). Explainable AI for Transparent Cyber-Risk Assessment and Decision-Making. In Machine Intelligence Applications in Cyber-Risk Management (pp. 219-246). IGI Global Scientific Publishing.
28. Younisse, R., Ahmad, A., & Abu Al-Haija, Q. (2022). Explaining intrusion detection-based convolutional neural networks using shapley additive explanations (shap). Big Data and Cognitive Computing, 6(4), 126.
29. Lee, J., Lee, Y., Kim, J., Kosiorek, A., Choi, S., & Teh, Y. W. (2019, May). Set transformer: A framework for attention-based permutation-invariant neural networks. In International conference on machine learning (pp. 3744-3753). PMLR.
30. Mustofa, R., Rafiquzzaman, M., & Hossain, N. U. I. (2024). Analyzing the impact of cyber-attacks on the performance of digital twin-based industrial organizations. Journal of Industrial Information Integration, 41, 100633.
31. Nguyen, G., Sáinz-Pardo Díaz, J., Calatrava, A., Berberi, L., Lytvyn, O., Kozlov, V., ... & López García, Á. (2024). Landscape of machine learning evolution: privacy-preserving federated learning frameworks and tools. Artificial Intelligence Review, 58(2), 51.
32. Sáez-de-Cámara, X. (2023). Federated Learning Approaches Towards Intrusion Detection in Industrial Internet of Things.