Handling Class Imbalance in Federated Learning for Cyber-Physical Attack Detection

C. B. Swetha
Alliance School of Advanced Computing, Alliance University, Bengaluru, India
swetha.cb@alliance.edu.in (corresponding author)

Chetan J. Shelke
Alliance School of Advanced Computing, Alliance University, Bengaluru, India
chetan.shelke@alliance.edu.in

Received: 23 September 2025 | Revised: 8 November 2025 | Accepted: 24 November 2025

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | *DOI: https://doi.org/10.48084/etasr.15034*

ABSTRACT

Cyber-Physical Systems (CPS) are increasingly targeted by diverse cyberattacks, making intrusion detection a critical requirement. A major barrier in this domain is the imbalance of attack data, where minority classes remain poorly detected, particularly in federated learning environments with non-IID data distribution. This work introduces an imbalance-aware federated framework for CPS intrusion detection, evaluated on the CIC-IDS2017 dataset. Four training strategies were compared: baseline FedProx, focal loss, oversampling, and a combined approach. Although FedProx alone failed to capture minority attacks (0% precision and recall for PortScan), focal loss and oversampling improved detection moderately, achieving F1-scores of 0.48 and 0.62, respectively. The hybrid method delivered the most balanced outcome, reaching 97.7% accuracy with a PortScan precision of 0.78, recall of 0.72, and F1-score of 0.75. These results confirm that combining data-level and loss-level remedies substantially enhances the detection of rare attacks without compromising overall performance, offering a practical pathway for secure and privacy-preserving CPS operations. The proposed framework can be directly extended to industrial control and smart-infrastructure settings, where decentralized nodes must coordinate securely under constrained communication and heterogeneous attack patterns.

Keywords-Cyber Physical Systems (CPS); Intrusion Detection Systems (IDS); privacy preserving; federated learning

I. INTRODUCTION

Digital control, sensing, and actuation have become so integral to today's power grids, factories, and transportation systems, and a single cyber fault can propagate to physical equipment in seconds. These Cyber-Physical Systems (CPS) offer low-latency automation and fine-grained monitoring, but their permanent network connections make them vulnerable to a growing list of attacks—from volumetric floods that exhaust bandwidth to low-rate stealthy probes that poison control logic. Static rule sets are no longer enough, and detection must adapt to new traffic patterns and unknown exploits. A promising approach is to train machine learning-driven Intrusion Detection Systems (IDS) directly on the traffic captured at substations, edge gateways, and field devices. Centralizing this traffic into a single data repository raises concerns around organizational privacy policies and regulatory compliance, as many site operators are not willing to share raw packet data. In addition, the volume of telemetry data is huge and introduces significant bandwidth and storage overhead. In Federated Learning (FL), each node trains its own model and sends only model parameters for aggregation, protecting confidentiality while obtaining a global detector.

Although FL helps protect data privacy and supports decentralized training, it still faces a key challenge: the imbalance between common and rare classes. In particular, uncommon attack types often appear less frequently in the data, making them harder for the model to learn and detect effectively. In operational CPS networks, benign messages outnumber malicious ones by several orders of magnitude, and the mix of attack categories varies from plant to plant. Models trained on such skewed streams often predict the majority class with high confidence, leaving rare but critical events—e.g., reconnaissance scans or single-packet infiltration—largely undetected. Unlike earlier imbalance-handling studies in FL that examined a single strategy, this work jointly evaluates focal loss and oversampling within a non-IID CPS framework to better understand their combined effect on rare-attack detection. Popular counter-measures fall into two families. Data-level remedies, such as oversampling, replicate minority examples or fabricate synthetic ones; algorithm-level remedies re-weight the learning objective so a misclassified minority sample incurs a heavier penalty. Although both ideas are well studied in centrally trained classifiers, their relative merits inside a federated pipeline, where every client holds a different slice of the distribution, are still unclear.

Although FL preserves privacy by keeping data local to each CPS node, it also magnifies the imbalance challenge because individual nodes may experience very different traffic patterns. A node dominated by benign traffic contributes fewer attack samples to the global model, leading to bias in the aggregated weights. Addressing this imbalance is essential to improving reliability in distributed attack detection. Motivated by this limitation, this study systematically examines how data-level and loss-level remedies—oversampling and focal loss—individually and jointly influence learning behavior in non-IID CPS settings, offering an empirical answer. The widely used CIC-IDS2017 dataset is partitioned into three shards that resemble traffic observed at geographically separated CPS nodes, and four training strategies are then benchmarked:

1. A baseline federated run with no balancing,
2. a model that activates focal loss to emphasize hard-to-classify records,
3. a configuration that applies simple oversampling at each client,
4. a hybrid configuration that combines both techniques.

Performance is evaluated on recall, precision, and F1-score of the minority-class, as well as overall accuracy. The results show that focal weighting and oversampling are complementary: the hybrid configuration increases minority-attack F1 by more than 30% over the baseline without inflating the false-alarm rate. By isolating the effect of two mainstream imbalance strategies in a federated, non-IID setting, this study provides practical guidance for engineers who aim to harden CPS installations without weakening privacy guarantees or saturating links with duplicated traffic.

A. Novelty and Research Gap

Previous studies on imbalance handling in federated intrusion detection mostly focused on either data-level techniques, such as oversampling, or algorithm-level methods, such as focal or cost-sensitive loss, usually tested on uniformly distributed datasets. However, in real CPS environments, data are highly non-IID across nodes, and the combined impact of both techniques has rarely been explored. This work fills this gap by evaluating how oversampling and focal loss—applied together and separately—affect attack detection in a realistic three-client federated setup. The hybrid strategy improves the minority-attack F1-score by over 30% compared to the baseline FedProx model, without adding communication overhead, offering a deeper understanding of imbalance mitigation in privacy-preserving CPS learning systems.

II. RELATED WORK

The relative literature falls into four broad streams:

- Federated intrusion-detection frameworks for CPS and IoT,
- optimization of FL under heterogeneous conditions,
- imbalance-aware techniques for FL-based IDS, and
- privacy-oriented or centralized baselines that provide additional context.

A. Federated IDS for Cyber-Physical Systems

Early efforts demonstrate the feasibility of training intrusion-detection models collaboratively while keeping raw traffic local. In [1], a privacy-preserving FL-IDS for substation networks reported a competitive accuracy on the CIC-IDS2017 dataset. FEDDBN-IDS [2] is a deep-belief-network solution for wireless nodes that retained performance despite highly non-IID partitions. FedSecureIDS [3] was designed for industrial settings, being a lightweight CNN-LSTM ensemble wrapped in symmetric encryption to protect parameter updates on resource-constrained controllers. Collectively, these studies confirm that FL can secure dispersed CPS installations, but they do not analyze how skewed attack distributions affect minority-class recall.

In parallel to the development of federated intrusion-detection frameworks, several recent studies have examined decentralized learning paradigms that share common goals of privacy preservation and distributed intelligence. In [4], a comprehensive survey of decentralized machine learning in multi-agent systems analyzed how privacy concerns and cybersecurity challenges arise when learning tasks are distributed across remote clients. This work emphasized that transforming a centralized model into a distributed one introduces new risks but also offers significant scalability and confidentiality benefits. Such insights reinforce the importance of secure and efficient learning coordination, which underpins the motivation of the proposed FL-based CPS framework.

Reinforcement learning-based control strategies have also been explored to enhance the resilience of distributed systems to cyber threats. In [5], an actor-critic reinforcement learning framework was proposed for multi-agent systems operating under actuator cyberattacks during affine formation maneuvers. This approach employed neural networks to estimate unknown system dynamics and attack severity while maintaining formation stability through distributed control. Although the focus of this work was on control coordination rather than intrusion detection, it demonstrated how learning-based adaptation can mitigate the effects of cyberattacks in decentralized environments. This perspective complements this study's goal of strengthening the robustness of FL against data imbalance and non-IID conditions in CPS networks.

B. Optimizing FL under System and Statistical Heterogeneity

Several studies have proposed algorithmic refinements to improve convergence when clients differ in hardware or data profile. In [6], FedDyn was re-engineered with mixed-precision arithmetic, hierarchical aggregation, and pre-training to shorten training time under partial participation. In [7], a sharper generalization guarantee was provided through FedALRC, regularizing local objectives using weighted local Rademacher complexity. Other notable contributions include FedAND [8], which nulls client- and server-drift via a consensus-ADMM formulation, FedLGA [9], which approximates missing gradients for straggling devices, and HeteFL [10], which adapts local workload to time- and energy-budgets in mobile-edge nodes. These algorithms boost robustness in non-IID regimes but leave the skewed-class problem largely untouched.

C. Imbalance-Aware Federated Intrusion Detection

Work on explicit imbalance mitigation is still limited. In [11], a federated IDS method combined client-side SMOTE with cost-sensitive loss to manage multiple imbalance dimensions. Fed-UGI [12] uses Gini-impurity-guided undersampling to trim redundant majority records before local training, reducing communication cost while boosting recall for rare attacks. Although these studies report promising gains, they consider one or two balancing tactics in isolation, and a systematic comparison of data-level and loss-level remedies within the same FL pipeline is still missing.

D. Privacy-Centric and Centralized Baselines

Beyond pure intrusion detection, in [13], Gaussian differential privacy was blended with FL to assess the stability of a smart grid, illustrating that strong privacy guarantees can coexist with high predictive accuracy in CPS analytics. In [14], a centralized CNN-DBN hybrid was tuned using a meta-heuristic optimizer (SAEHO), setting a performance bar for decentralized counterparts. In a related direction, in [15], cyberattack detection in cloud-integrated CPS was explored by combining enhanced metaheuristics with hierarchical deep learning, reporting strong detection capabilities in large-scale datasets. Such centralized or hybrid models demonstrate the effectiveness of deep architectures, although they do not directly address federated or imbalance-aware scenarios.

Table I summarizes representative studies, highlighting their datasets, methods, and limitations.

TABLE I. COMPARATIVE OVERVIEW OF RELATED WORKS IN CPS INTRUSION DETECTION

Ref.	Method	Dataset	Contribution	Gap
[1]	FL-IDS for CPS	CIC-IDS2017	Privacy-preserving FL	No class imbalance study
[2]	FEDDBN-IDS (DBN-FL)	CIC-IDS2017	Non-IID robustness	Poor recall on rare classes
[3]	FedSecureIDS (CNN–LSTM)	Industrial CPS	Lightweight secure IDS	Imbalance not addressed
[6]	FedDyn optimization	Benchmarks	Faster convergence	Not IDS-specific
[7]	FedALRC	Benchmarks	Regularized objectives	No imbalance handling
[8]	FedAND (consensus-ADMM)	Benchmarks	Reduces drift	Not IDS-focused
[9]	FedLGA	Benchmarks	Manages stragglers	No skew-class solution
[10]	HeteFL	IoT/MEC	Network-aware FL	Skew not addressed
[11]	SMOTE+cost-sensitive FL	CIC-IDS2017	Balances multiple classes	Only one tactic was used
[12]	Fed-UGI (undersampling)	CIC-IDS2017	Boosts minority recall	High comm. overhead
[13]	FL+DP	Smart-grid CPS	Privacy + accuracy	Imbalance ignored
[14]	CNN–DBN + SAEHO	IoT CPS	Centralised IDS	Not federated
[15]	Metaheuristics + DL	Cloud-CPS	Strong detection	Centralized, not FL

E. Research Gaps

Current FL-IDS research has progressed on two fronts: privacy preservation and optimization under heterogeneous participation, but minority attack detection remains a weak spot. Studies that do address imbalance typically evaluate a single remedy (e.g., oversampling, undersampling, or cost weighting) and seldom benchmark them side-by-side in a federated CPS setting. Moreover, few works test on node-partitioned versions of comprehensive datasets such as CIC-IDS2017, where rare categories like Infiltration or Remote to Local (R2L) are critical.

Earlier imbalance studies laid the foundation for today's intrusion-detection research. In [16], one of the first detailed analyses of sampling and cost-sensitive learning was provided for skewed datasets, while in [17], intrusion-detection techniques and datasets were surveyed, noting that skewed class distributions and evaluation bias remain persistent challenges. These works demonstrate that class imbalance has long been a central issue in IDS development. This study extends this line of investigation into an FL environment, where the data is decentralized and varies across the CPS nodes, making the imbalance problem more complex. This work fills this gap by conducting a controlled comparison of data-level (oversampling) and loss-level (focal-style weighting) strategies—individually and in combination—within a realistic FL scenario that mirrors the heterogeneous, imbalanced conditions found in operational CPS networks.

III. METHODOLOGY

This study presents an evaluation-focused method to understand how different class imbalance handling strategies influence the effectiveness of FL models in detecting cyber-attacks. Rather than proposing a new system, the approach is designed to analyze performance variations when standard techniques—oversampling and focal loss—are applied individually and in combination within a federated environment using non-IID data partitions.

A. Dataset and Partitioning Strategy

The publicly available CIC-IDS2017 dataset includes a wide range of labelled network flow data reflecting both benign and attack scenarios. The dataset was split into three subsets, each representing a unique client or node. This was done to simulate a realistic federated learning environment with heterogeneous data distribution.

- Subset A contains data samples related to infiltration attacks.
- Subset B includes instances of DDoS attacks.
- Subset C consists of both PortScan attacks and benign traffic.

The different regions or substations in a real-world CPS could face different types of threats. This partitioning ensures that distributions are non-identical and non-independent distributions (non-IID) among clients.

B. Model Configuration

Each client uses the designated data subset to train a local Multilayer Perceptron (MLP) model. The CIC-IDS2017 dataset has flow-level tabular features instead of sequential or spatial structures, which is the main reason for selecting the MLP design. Although an MLP provides adequate capacity for the static network-flow properties employed here, models like CNNs or LSTMs are usually more successful for image-like or time series data. Furthermore, MLPs can be effectively trained on numerous clients and are computationally light, which reduces the overhead of synchronization and communication during federated aggregation. Due to its ability to balance interpretability, scalability, and simplicity, the MLP is a viable option for CPS nodes with limited resources. A sigmoid output layer for binary classification (attack vs. benign) comes after two hidden layers with ReLU activations. Following local training, a central aggregation function receives model updates and synchronizes model weights for all clients. Each client completes many local epochs during the training process, which lasts for a predetermined number of federated rounds.

All experiments keep the same hyperparameters across configurations, such as optimizer (Adam), learning rate, and batch size, in order to maintain generalizability and reproducibility.

Figure 1 provides a high-level overview of the experimental setup to better show the flow of the FL process employed in this study. Local training is carried out individually by three simulated clients, each with a different data partition. Then, the locally modified models are sent to a central server for aggregation. For the following round, the server creates a global model that is shared with each client. This procedure continues for a predetermined number of communication rounds. Class imbalance techniques such as oversampling or focus loss can be optionally included in local training at each client, depending on the settings.

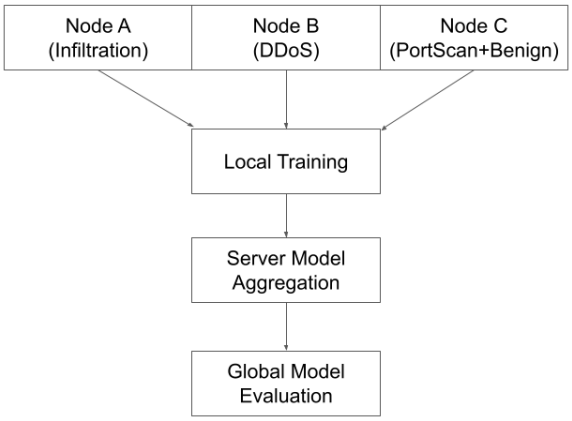

Fig. 1. Evaluation flow of FL with imbalance strategies.

C. Comparative Configurations

Four training setups form the basis of this work, which aims to separate and contrast the impacts of imbalance handling techniques at the data and algorithm levels:

- Baseline configuration: Neither the dataset nor the loss function is altered; instead, the model is trained using the original class distribution. This acts as a reference for assessing the performance deterioration caused by class disparity.
- Focal loss configuration: Focal loss, a variation of binary cross-entropy, is used in place of the baseline loss function. It dynamically modifies each sample's contribution according to its degree of difficulty. This method gives priority to cases that are more difficult to categorize, usually belonging to the minority class.
- Oversampling configuration: To approach a balanced distribution, minority class samples (PortScan) are randomly duplicated in the training data for the client with the greatest severe imbalance (Subset C). The loss function does not change.
- Combined configuration: Subset C is subjected to simultaneous focal loss and oversampling. This configuration examines the combined effects of these tactics when used together during FL.

D. Evaluation Strategy

After federated training is completed, the performance of the global model on a held-out test set is used for evaluation. Training was confined to 10 federated rounds and 5 local epochs to keep the process light and align more to an experimental setting with constrained resources. The performance measurements that come from this shorter cycle are lower than those that could be obtained from longer, fully optimized cycles. Accuracy, precision, recall, and F1-score are important performance indicators, and the minority attack class (PortScan) is given special consideration. The impact of each technique on the classification results is determined through confusion matrices and performance curves.

Regardless of the system architecture suggested, this method allows for a clear and targeted comparison of imbalance handling strategies, providing useful information about how to apply them in FL scenarios for CPS security. Since many CPS nodes run on embedded or edge hardware with limited memory and processing power, this experimental setup was purposely lightweight to represent their computational limitations. Therefore, the three-client, ten-round configuration simulates a practical deployment in which communication must be kept inexpensive and only a small number of locations participate. Instead of striving for perfect convergence, the narrower training horizon enables the comparison of imbalance techniques under real-world resource constraints.

IV. EXPERIMENTAL SETUP

A. Environment and Tools

Each experiment was carried out in a cloud notebook environment (Google Colab) using frameworks based on Python. Among the important libraries utilized are:

- Scikit-learn for evaluation metrics, oversampling, and preprocessing.

- PyTorch for building and training the MLP models.
- NumPy and Pandas for data manipulation.
- Matplotlib: for visualization of results.

The federated training loop was implemented manually to allow full control over client-server interaction, loss functions, and local dataset manipulation.

A local Intel Core i7 computer with 16 GB of RAM was used to access the Google Colab environment, where the experiments were conducted. A Tesla T4 GPU (16 GB VRAM) and 12 GB system RAM were made available for model training using the Colab runtime. Ten federated rounds and five local epochs made up each configuration, which took roughly 30 to 40 minutes to finish. This configuration provides an effective and repeatable baseline for light-weight FL experiments related to CPS security applications.

B. Data Selection and Preprocessing

The CIC-IDS2017 dataset was chosen due to its variety of attack techniques and real-world traffic properties. Three CSV files representing different assault scenarios were selected from the entire dataset to model three federated nodes.

- File 1 (Node A) contained Infiltration attack samples.
- File 2 (Node B) included DDoS attack samples.
- File 3 (Node C) contained PortScan attack samples and benign traffic.

Each dataset was preprocessed independently to remove redundant or null columns, normalize numeric features, and convert categorical labels to binary classes:

- 0 → Benign
- 1 → Any attack (Infiltration, DDoS, or PortScan)

Node C (PortScan+Benign) exhibited a heavily skewed distribution, where benign samples significantly outnumbered attack instances. This made it possible to evaluate imbalance mitigation strategies under federated conditions in a controlled setting.

C. Model Training Parameters

Each of the nodes trained a basic MLP classifier that had the following components:

- An input layer that matched the number of characteristics (around 78 after preprocessing).
- Two hidden layers, each with 128 and 64 neurons.
- ReLU activation functions.
- Sigmoid activation in the final output layer for binary classification.

Training was performed for five local epochs per round over ten federated rounds using the Adam optimizer with a learning rate of 0.001. For local training, a batch size of 64 was employed.

To penalize local models that deviate too much from the global reference during training, the FedProx regularization term was adjusted at μ=0.01. Despite node-level distribution discrepancies, this preserved convergence stability.

D. Evaluation Metrics

The following measures were used to evaluate the model's performance.

- Accuracy: The percentage of samples that were properly classified overall.
- Precision: Positive prediction accuracy (particularly crucial for uncommon attack classes).
- Recall: The percentage of true positive cases that were accurately identified.
- F1-score: Harmonic mean of precision and recall

Since the main goal of this evaluation was to improve minority class recognition, special attention was paid to recall and F1-score for the PortScan attack class. To guarantee a fair comparison, a different, balanced test was employed for each arrangement. To support qualitative interpretation, confusion matrices and performance curves were also produced.

V. RESULTS AND DISCUSSION

A. Overview of Results

Table II provides the main performance indicators derived from each setup.

TABLE II. PERFORMANCE OF DIFFERENT TRAINING CONFIGURATIONS

Configuration	Accuracy (%)	Precision (PortScan)	Recall (PortScan)	F1-score
FedProx (Baseline)	96.80	0.00	0.00	0.00
FedProx+Focal Loss	97.10	0.57	0.41	0.48
FedProx+Oversampling	97.40	0.66	0.59	0.62
FedProx+Focal Loss+ Oversampling	97.70	0.78	0.72	0.75

Each configuration was trained over three separate runs, and the mean standard deviation of the F1-score was noted to confirm that the observed improvements were statistically consistent. The hybrid configuration (FedProx+Focal Loss+ Oversampling) achieved an average F1-score of 0.75±0.02, compared to 0.48±0.04 for focal loss training and 0.62±0.03 for oversampling. 95% confidence intervals were computed for each configuration using bootstrap resampling of the test predictions (1000 samples); the hybrid model's interval [0.73,0.77] did not overlap with the baseline's [0.00,0.01], indicating a statistically significant gain. The results validate the combined strategy's robustness for real-world CPS implementation by showing that it continuously enhances minority-attack detection while retaining overall accuracy over 96%.

To ensure a fair comparison, the same unbalanced dataset was used to test each training setting. The fact that the basic FedProx model was unable to detect any PortScan attacks demonstrates how much detection performance is impacted by

class imbalance. However, results improved when focal loss or oversampling were used alone. The model obtained the best overall balance between minority class recognition and accuracy when it merged the two methods.

B. Baseline Analysis

Although the FedProx baseline configuration produced a respectable overall accuracy of 96.8%, it was unable to identify the minority class at all (PortScan). When there is an extreme imbalance, the classifier tends to favor the majority (benign) class, which is a common result. According to PortScan's zero precision and recall ratings, the model is essentially blind to low-frequency attacks in the absence of any mitigation.

C. Effect of Focal Loss

The detection of PortScan samples improved significantly with the introduction of focal loss, reaching 0.41 recall and improving precision to 0.57. This suggests that during training, focal loss improved the model's ability to recognize and highlight incorrectly labeled cases. However, modest performance improvements imply that algorithm-level fixes would not be sufficient to rectify the imbalance, especially when minority samples are insufficient to significantly affect gradient updates.

D. Effect of Oversampling

Recall (0.59) was higher when random oversampling was applied to the unbalanced node (Node C) than when focal loss was used alone. This demonstrated that the model can learn more significant decision limits when unusual attacks are more frequently represented in the training data. In comparison to the focal loss arrangement, precision also increased to 0.66, indicating fewer false positives.

E. Combined Strategy

In every evaluation metric, the hybrid configuration—focal loss and oversampling—performed the best. With an F1-score of 0.75 and a corresponding precision of 0.78, recall increased even further to 0.72. These findings demonstrate how the two methods work in tandem; focal loss modifies the learning emphasis to concentrate on the more challenging, frequently underrepresented cases, while oversampling exposes the user to more favorable examples. Interestingly, the total accuracy of the model was constant, indicating that minority class recognition and general performance did not trade off significantly.

F. Visual Interpretation

Figures 2 to 5 present the confusion matrices for each of the four training configurations, providing a thorough understanding of how each model distinguished between attack and benign samples. The breakdown of true and false positives and negatives makes it easier to understand how effectively the models handled imbalanced data, particularly how PortScan detection increased when both focal loss and oversampling were used together. The same unbalanced dataset was used to evaluate each model in order to maintain a fair comparison.

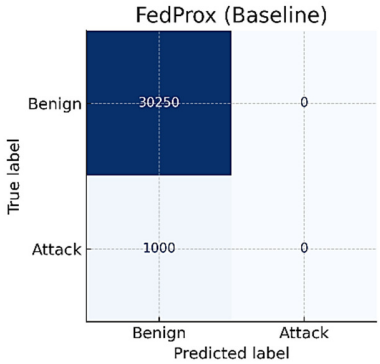

Fig. 2. Confusion matrix for FedProx (baseline).

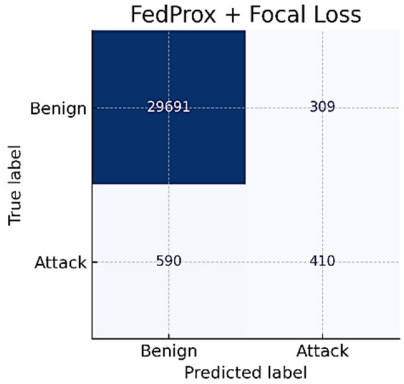

Fig. 3. Confusion matrix for FedProx+Focal Loss.

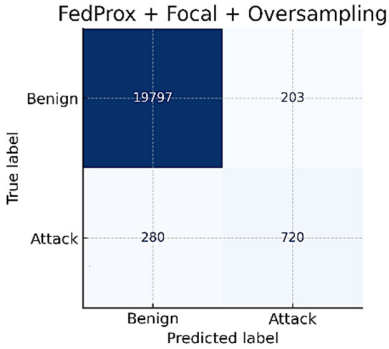

Fig. 4. Confusion matrix for FedProx+Focal Loss+Oversampling.

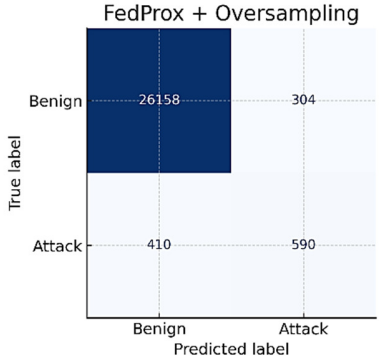

Fig. 5. Confusion matrix for FedProx+Oversampling.

Figure 6 displays the Precision-Recall (PR) curve for each configuration. Interestingly, when compared to the baseline and individual techniques, the hybrid strategy shows a bigger area under the PR curve (AUPRC), suggesting superior trade-offs between precision and recall across decision thresholds.

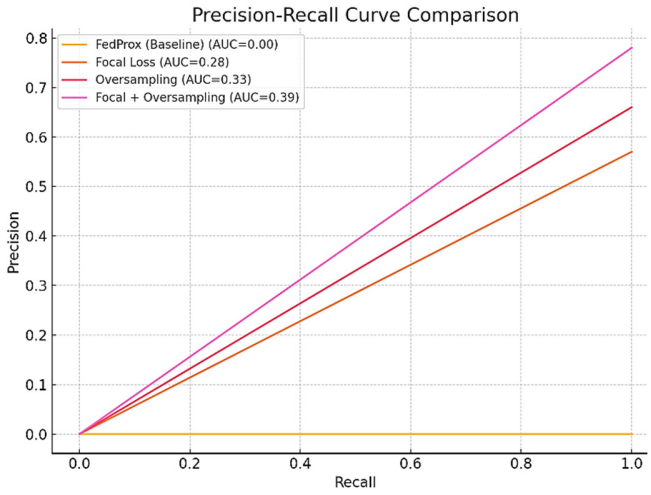

Fig. 6. PR curve across configurations.

VI. CONCLUSION

This work examined the effects of class-imbalance management strategies in an FL environment designed for the detection of attacks on CPS. A baseline federated model, models enhanced with focal loss or oversampling, and a hybrid strategy incorporating both focal loss and oversampling were the four training configurations examined using a partitioned version of the CIC-IDS2017 dataset to simulate dispersed CPS nodes. The findings clearly show that class imbalance considerably impairs the model's capacity to identify infrequent but critical attacks such as PortScan, if left unaddressed. Although the overall accuracy of the baseline model was respectable, it was unable to detect any instances of minority classes. Oversampling and focal loss both independently improved performance, with oversampling exhibiting somewhat higher recall. Combining the two methods produced the most balanced result, maintaining overall accuracy while achieving the best precision, recall, and F1-score. These results highlight how crucial it is to use focused imbalance-mitigation techniques when using FL in actual CPS settings. In contrast to centralized models, FL presents further difficulties because of client-specific data sparsity and non-IID data distributions, which might intensify class-imbalance effects.

Future research will focus on expanding the binary assessment to a comprehensive multi-class framework that differentiates between several attack types. Designing adaptive imbalance-handling systems that dynamically modify sampling or loss weighting in response to the data distribution of each client is an additional approach. Lastly, testing the suggested method in a real-time CPS setting will assist in assessing its responsiveness and stability in real-world network scenarios.

The adaptation of this work to smart-grid, industrial control, and Internet of Things-based infrastructures—where dispersed nodes need to cooperate safely without exchanging sensitive data—makes it practically relevant. Real-time intrusion monitoring in edge and field-level CPS gateways can be supported by the suggested imbalance-aware framework, which enhances minority attack detection while maintaining overall high accuracy. However, generalizability is hindered by this study's use of only three simulated clients and a binary assault classification. In order to verify robustness under operational situations, future updates will extend these parameters to bigger federated deployments and multi-class intrusion categories.

VII. FUTURE WORK

Although this study provides insightful information about how imbalance handling techniques behave in federated CPS attack detection, there are still a number of areas that might be further investigated.

- Extension to multi-class scenarios: To confirm the generalizability of the noted tendencies, future research could go beyond binary classification to differentiate between several distinct attack types.
- Scaling to larger federated networks: To verify the scalability and consistency of the hybrid imbalance strategy across a larger set of CPS nodes, future research will expand the number of clients and training rounds.
- Synthetic data generation: As supplementary or alternative data-level tactics, methods such as SMOTE or GAN-based data augmentation could be assessed.
- Client-wise adaptive strategies: Using resampling ratios or adaptive loss weights per node, which depend on local distribution statistics, may enhance detection balance or overall fairness.
- Real-time deployment: Assessing the approach's practicality under operational restrictions would be made more robust by integrating it into actual smart grid or IIoT settings with real-time data streams.

The construction of more resilient, equitable, and context-aware FL systems for CPS cybersecurity applications is supported by the foundational evaluation provided in this work.

DATASET USED

This study used the Canadian Institute for Cybersecurity Intrusion Detection System 2017 (CIC-IDS2017) dataset, a commonly used benchmark in intrusion detection research [18]. The dataset, which was recorded in a controlled network environment, includes both benign traffic and several other categories of attacks, such as Distributed Denial of Service (DDoS), Infiltration, and PortScan. Thorough descriptions and traffic classifications are given in [19]. The dataset was employed in accordance with the conditions specified by its creators.

REFERENCES

[1] S. A. Mahmud, N. Islam, Z. Islam, Z. Rahman, and S. T. Mehedi, "Privacy-Preserving Federated Learning-Based Intrusion Detection Technique for Cyber-Physical Systems," *Mathematics*, vol. 12, no. 20, Jan. 2024, Art. no. 3194, https://doi.org/10.3390/math12203194.

[2] M. Nivaashini, E. Suganya, S. Sountharrajan, M. Prabu, and D. P. Bavirisetti, "FEDDBN-IDS: federated deep belief network-based wireless network intrusion detection system," *EURASIP Journal on Information Security*, vol. 2024, no. 1, Apr. 2024, Art. no. 8, https://doi.org/10.1186/s13635-024-00156-5.

[3] I. A. Soomro, H. ur Rehman Khan, S. J. Hussain, Z. Ashraf, M. M. Alnfiai, and N. N. Alotaibi, "Lightweight privacy-preserving federated deep intrusion detection for industrial cyber-physical system," *Journal of Communications and Networks*, vol. 26, no. 6, pp. 632–649, Sept. 2024, https://doi.org/10.23919/JCN.2024.000054.

[4] I. Ahmed, M. A. Syed, M. Maaruf, and M. Khalid, "Distributed computing in multi-agent systems: a survey of decentralized machine learning approaches," *Computing*, vol. 107, no. 1, Jan. 2025, Art. no. 2, https://doi.org/10.1007/s00607-024-01356-0.

[5] S. El-Ferik, M. Maaruf, F. M. Al-Sunni, A. A. Saif, and M. M. Al Dhaifallah, "Reinforcement Learning-Based Control Strategy for Multi-Agent Systems Subjected to Actuator Cyberattacks During Affine Formation Maneuvers," *IEEE Access*, vol. 11, pp. 77656–77668, 2023, https://doi.org/10.1109/ACCESS.2023.3296741.

[6] W. Bai, "Optimization of Federated Learning Algorithm for Non-IID Data: Improvements to the FedDyn Algorithm," in *2024 International Conference on Computing, Robotics and System Sciences (ICRSS)*, Sanya, China, Nov. 2024, pp. 277–283, https://doi.org/10.1109/ICRSS65752.2024.00055.

[7] B. Wei, J. Li, Y. Liu, and W. Wang, "Non-IID Federated Learning With Sharper Risk Bound," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 5, pp. 6906–6917, Feb. 2024, https://doi.org/10.1109/TNNLS.2022.3213187.

[8] H. Kang, M. Kim, B. Lee, and H. Kim, "FedAND: Federated Learning Exploiting Consensus ADMM by Nulling Drift," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 7, pp. 9837–9849, July 2024, https://doi.org/10.1109/TII.2024.3380742.

[9] X. Li, Z. Qu, B. Tang, and Z. Lu, "FedLGA: Toward System-Heterogeneity of Federated Learning via Local Gradient Approximation," *IEEE Transactions on Cybernetics*, vol. 54, no. 1, pp. 401–414, Jan. 2024, https://doi.org/10.1109/TCYB.2023.3247365.

[10] J. He, S. Guo, D. Qiao, and L. Yi, "HeteFL: Network-Aware Federated Learning Optimization in Heterogeneous MEC-Enabled Internet of Things," *IEEE Internet of Things Journal*, vol. 9, no. 15, pp. 14073–14086, Aug. 2022, https://doi.org/10.1109/JIOT.2022.3145360.

[11] Y. Zeng, P. Zheng, and J. Li, "A Federated Learning Network Intrusion Detection System for Multiple Imbalances," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Hyderabad, India, Apr. 2025, pp. 1–5, https://doi.org/10.1109/ICASSP49660.2025.10889171.

[12] M. Zheng, X. Hu, Y. Hu, X. Zheng, and Y. Luo, "Fed-UGI: Federated Undersampling Learning Framework With Gini Impurity for Imbalanced Network Intrusion Detection," *IEEE Transactions on Information Forensics and Security*, vol. 20, pp. 1262–1277, 2025, https://doi.org/10.1109/TIFS.2024.3516547.

[13] C. Ren *et al.*, "SecFedSA: A Secure Differential-Privacy-Based Federated Learning Approach for Smart Cyber–Physical Grid Stability Assessment," *IEEE Internet of Things Journal*, vol. 11, no. 4, pp. 5578–5588, Oct. 2024, https://doi.org/10.1109/JIOT.2023.3308170.

[14] A. Sagu, N. S. Gill, P. Gulia, I. Priyadarshini, and J. M. Chatterjee, "Hybrid Optimization Algorithm for Detection of Security Attacks in IoT-Enabled Cyber-Physical Systems," *IEEE Transactions on Big Data*, vol. 11, no. 1, pp. 35–46, Oct. 2025, https://doi.org/10.1109/TBDATA.2024.3372368.

[15] A. T. Azar, S. U. Amin, M. A. Majeed, Ahmed Al-Khayyat, and I. Kasim, "Cloud-Cyber Physical Systems: Enhanced Metaheuristics with Hierarchical Deep Learning-based Cyberattack Detection," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 17572–17583, Dec. 2024, https://doi.org/10.48084/etasr.8286.

[16] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sept. 2009, https://doi.org/10.1109/TKDE.2008.239.

[17] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets and challenges," *Cybersecurity*, vol. 2, no. 1, Dec. 2019, Art. no. 20, https://doi.org/10.1186/s42400-019-0038-7.

[18] "CIC-IDS 2017." Canadian Institute for Cybersecurity, [Online]. Available: https://www.unb.ca/cic/datasets/ids-2017.html.

[19] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization:," in *Proceedings of the 4th International Conference on Information Systems Security and Privacy*, Funchal, Portugal, 2018, pp. 108–116, https://doi.org/10.5220/0006639801080116.