

Federated Learning, Temporal Convolutional Networks, Electric Vehicle Charging Infrastructure, Intrusion Detection, Cybersecurity, Privacy-Preserving Machine Learning

Mohammed Gamal Ragab

Universiti Teknologi Petronas

Hitham Alhussian

`seddig.alhussian@utp.edu.my`

Universiti Teknologi Petronas

Said Jadid Abdulkadir

Universiti Teknologi Petronas

Majdy Eltahir

King Khalid University

Ayed Alwadain

King Saud University

Article

Keywords:

Posted Date: June 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9637381/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Federated Dual-Attention Temporal Convolutional Networks for Privacy-Preserving Intrusion Detection in EV Charging Infrastructure

Mohammed Gamal Ragab^{1,2,*}, Hitham Alhussian^{1,2}, Said Jadid Abdulkadir^{1,2}, Majdy Eltahir^{*3}, and Ayed Alwadain⁴

¹Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Seri Iskandar, 32610, Perak, Malaysia

²Center for Research in Data Science (CeRDs), Universiti Teknologi PETRONAS, Seri Iskandar, 32610, Perak, Malaysia

³Department of Information Systems, College of Science and Art at Mahayil, King Khalid University, Riyadh 12372, Saudi Arabia

⁴Computer Science Department, Community College, King Saud University, Riyadh, Saudi Arabia

*Correspondence to: Mohammed Gamal Ragab (mohammed.ragab@utp.edu.my)

ABSTRACT

The rapid growth of electric vehicle (EV) charging infrastructure has increased exposure to cyberattacks, while conventional centralized intrusion detection remains poorly suited to distributed and privacy-sensitive EVSE environments. This study proposes a federated learning framework based on a Dual-Attention Temporal Convolutional Network (DA-TCN) for collaborative cyberattack detection without centralized data sharing. The model combines dilated causal convolutions with channel and temporal attention to capture complex dependencies in multimodal EVSE telemetry. To improve federated optimization, we introduce Federated Stochastic Weight Averaging (FedSWA) and client-local mixup augmentation, which together enhance generalization and robustness across distributed clients. Evaluated on the CICEVSE2024 dataset, the proposed framework achieved 99.18% accuracy and 98.95% macro F1-score, outperforming a centralized baseline (98.96% accuracy) while maintaining perfect recall for benign and cryptojacking traffic. With an inference latency of 7.21 ms per batch and a model size of 4.07 MB, the framework is well suited for real-time edge deployment. A complementary Isolation Forest module provides unsupervised zero-day anomaly detection. These results demonstrate that federated learning can deliver accurate, lightweight, and deployment-ready intrusion detection for privacy-preserving EV charging infrastructure.

Introduction

The rapid growth of electric vehicles (EVs) is driving equally rapid deployment of charging infrastructure, turning Electric Vehicle Supply Equipment (EVSE) into a critical part of modern cyber-physical energy systems^{1,2}. Unlike traditional electrical assets, contemporary charging stations combine power electronics, embedded software, network connectivity, backend cloud services, and user-facing applications within a single operational stack^{3,4}. As a result, securing EV charging infrastructure is now a systems problem spanning transportation, energy, and communications.

This expanded functionality depends on interoperable communication standards such as the Open Charge Point Protocol (OCPP) and ISO 15118⁵⁻⁷. These protocols enable remote management, billing, smart charging, and Plug & Charge workflows, but they also enlarge the attack surface by exposing chargers to network-facing services, credential exchange, and software-controlled operating modes^{3,8}. The resulting risks are no longer hypothetical: the Brokenwire attack demonstrated wireless disruption of Combined Charging System (CCS) sessions without physical contact⁹, while broader analyses have documented denial-of-service, man-in-the-middle, remote-control, and firmware-level vulnerabilities across real deployments^{4,6,10}. Such threats extend beyond IT compromise to interrupted charging sessions, degraded user trust, and potential effects on grid-facing operations.

Machine learning has emerged as a promising response because it can model complex behavior and detect attack patterns that are difficult to encode with static rules. Prior work has shown strong results for in-vehicle CAN-bus intrusion detection^{11,12} and, more recently, for EVSE-focused datasets and testbeds¹³⁻¹⁶. In particular, the CICEVSE2024 benchmark provides

*ORCID: [0000-0002-1810-4372](https://orcid.org/0000-0002-1810-4372).

synchronized network, host, and power measurements from physical EVSE experiments, enabling more realistic multimodal intrusion detection research¹⁷. However, recent studies on this benchmark have used centralized or single-site training settings, which do not align well with real deployments in which charging operators are distributed and may be unable or unwilling to pool raw operational data.

Federated learning addresses this tension by training a shared model through parameter exchange rather than raw-data transfer¹⁸. Yet federated optimization is challenging: heterogeneous local data and fragmented client views often create a performance gap relative to centralized training¹⁹. Survey literature identifies privacy-preserving learning as highly relevant to connected and electric-vehicle ecosystems²⁰, but its application to EVSE intrusion detection remains limited. This paper therefore asks whether a carefully designed federated model can retain the privacy advantages of decentralized training while remaining competitive with, or even surpassing, a centralized EVSE intrusion detector.

To answer this question, we develop a federated intrusion detection framework built around Temporal Convolutional Networks (TCNs), which provide causal sequence modeling, stable gradient propagation, and efficient parallel computation. The principal contributions are:

- **Dual-Attention TCN architecture.** We propose DA-TCN, which combines dilated causal convolutions with Squeeze-and-Excitation (SE) channel attention and multi-head temporal attention to jointly model feature importance and temporal dependencies in multimodal EVSE telemetry.
- **Federated Stochastic Weight Averaging (FedSWA).** We introduce a federated weight-averaging strategy that maintains a running average of global models across communication rounds, stabilizing optimization and improving generalization beyond standard FedAvg.
- **Client-local mixup augmentation.** Each client independently applies mixup interpolation during local training to increase data diversity without sharing raw samples, preserving privacy while reducing overfitting.
- **Hybrid detection pipeline.** We combine supervised DA-TCN classification with unsupervised Isolation Forest anomaly detection to strengthen detection of previously unseen threats.
- **State-of-the-art federated performance.** The federated model surpasses the centralized baseline (99.18% vs. 98.96% accuracy; 98.95% vs. 98.73% macro F1), with 7.21 ms inference latency and a 4.07 MB model size suitable for real-time edge deployment.
- **Comprehensive robustness analysis.** Sensitivity studies confirm robust performance under differential privacy noise ($>98\%$ accuracy up to $\sigma = 0.5$), stable scalability from 3 to 15 clients, and graceful degradation under non-IID data heterogeneity.

The remainder of this paper is organized as follows. The next section reviews related work on EV charging security, temporal modeling for intrusion detection, and federated learning. Subsequent sections detail the proposed methodology, present experimental results, and discuss their implications. The paper concludes with a summary of findings and directions for future research.

Related Work

Research on EV cybersecurity spans multiple interconnected domains, from in-vehicle network protection to charging infrastructure defense and privacy-preserving distributed learning. This section reviews the relevant literature across these areas, identifies persistent gaps, and positions the present work within the broader landscape¹.

In-Vehicle Network Intrusion Detection

Modern vehicles rely on networks of Electronic Control Units (ECUs) that exchange messages over Controller Area Network (CAN) and, increasingly, automotive Ethernet. CAN in particular lacks native authentication, making message spoofing and injection persistent concerns^{12,21}. Early in-vehicle intrusion detection studies applied deep neural networks directly to CAN traffic²¹, while later work adopted temporal models such as LSTMs to capture sequential attack behavior^{22,23}. CNN-based approaches have also been explored by transforming CAN traces into structured representations such as recurrence plots²⁴.

These studies establish the importance of temporal modeling for automotive cybersecurity, but their scope is primarily in-vehicle communication. EV charging infrastructure poses a broader problem because it couples vehicle-side communication with backend services, charger host activity, and power-side telemetry, all of which must be considered in a practical EVSE intrusion detector.

Charging Infrastructure Vulnerabilities

Beyond the vehicle itself, comprehensive assessments have revealed vulnerabilities across communication protocols, upstream cloud services, and charger-side software in EV charging infrastructure^{3,4,8}. OCPP has been repeatedly identified as a critical attack surface because charging stations and central management systems depend on it for command and control^{5,6}.

Experimental studies have demonstrated that weaknesses in charging workflows can enable denial-of-service, man-in-the-middle, and remote code execution scenarios under realistic conditions¹⁰. ISO 15118 introduces further security-sensitive exchanges around authorization and vehicle-to-grid interactions, making protocol robustness essential to trustworthy charging operation⁷. Collectively, these findings motivate defense mechanisms tailored to the operational and protocol realities of EV charging systems rather than borrowed unchanged from generic enterprise networks.

Datasets for EV Charging Security

The development of data-driven security solutions depends on representative datasets. Public EV charging datasets such as ACN-Data are valuable for charging-behavior and grid studies, but they were not designed for cybersecurity evaluation²⁵. The CICEV2023 dataset introduced a benchmark for DDoS attacks against EV authentication in charging infrastructure²⁶. CICEVSE2024 substantially broadens this foundation by collecting synchronized network traffic, host activities, and power-consumption measurements from physical EVSE testbeds under benign and attack conditions¹⁷. Its multimodal structure makes it a strong benchmark for evaluating modern EVSE intrusion detection methods.

Machine Learning for EVSE Intrusion Detection

Recent studies built on CICEVSE2024 have demonstrated that the benchmark supports high-performing EVSE attack detection. Kumar et al. investigate machine-learning models on the CICEVSE2024 benchmark¹³. Almadhor et al. study transfer learning for cyber-physical attack detection in EV charging infrastructure¹⁴, while Naeem et al. combine deep transfer learning with genetic-algorithm-based optimization for vehicle-security classification tasks that include CICEVSE data¹⁵. Makhmudov et al. focus on online intrusion detection for EV charging systems, highlighting the importance of adaptation under changing traffic conditions¹⁶. Despite this methodological diversity, the published studies remain centralized or single-site in training setup, leaving privacy-preserving collaborative learning largely unexplored for EVSE intrusion detection.

Temporal Convolutional Networks for Sequential Anomaly Detection

Temporal Convolutional Networks (TCNs) have emerged as a compelling alternative to recurrent architectures for sequence modeling tasks. Introduced by Bai et al.²⁷, TCNs use dilated causal convolutions to capture long-range dependencies while preserving temporal ordering and enabling parallel training. These properties are attractive for security analytics, where low-latency inference and stable optimization are often as important as raw accuracy.

In the broader network-security literature, TCN-based models have been used for abnormal traffic detection²⁸, and attention-augmented bidirectional TCN variants have improved malicious traffic classification by learning which temporal patterns deserve greater weight²⁹. For EVSE security, this design space is especially relevant because attacks unfold across multiple time scales and across heterogeneous telemetry streams. Nevertheless, we did not find prior work applying a dual-attention TCN architecture to federated EVSE intrusion detection.

Federated Learning and Privacy-Preserving Detection

Federated learning enables collaborative model training through the exchange of model updates rather than raw data, preserving privacy in distributed environments¹⁸. FedAvg remains the standard baseline, while extensions such as FedProx address optimization instability under heterogeneous client data¹⁹. In the EV domain, survey literature has identified federated learning as a promising direction for privacy-preserving analytics across connected and electric-vehicle ecosystems²⁰. However, the available EVSE intrusion-detection literature on CICEVSE2024 continues to center on centralized learning^{13–16}.

A persistent challenge in federated learning is the accuracy gap relative to centralized training, especially when clients hold limited or non-identically distributed data. Stochastic Weight Averaging (SWA) improves generalization in centralized optimization by averaging weights along the training trajectory³⁰. Likewise, mixup regularization improves robustness by interpolating between training samples³¹. Although both ideas are well established in centralized learning, their use as federated optimization tools for EVSE intrusion detection remains underexplored.

Research Gaps and Positioning

Despite considerable progress, three specific gaps motivate the present work:

1. **Limited privacy-preserving learning for EVSE IDS.** To the best of our knowledge, published studies on CICEVSE2024-based intrusion detection have used centralized or single-site training, leaving federated evaluation on this benchmark unaddressed.
2. **Underexplored temporal modeling for multimodal EVSE telemetry.** Prior EVSE studies emphasize transfer learning, online tree-based learners, or other centralized classifiers. TCN-based sequence modeling has not been established on this benchmark despite the inherently temporal nature of attack behavior.

3. **Limited study of federated optimization strategies for EVSE security.** Techniques such as SWA and mixup are well known in centralized learning, but their adaptation to federated EVSE intrusion detection has not been systematically examined.

The present work addresses these gaps by proposing a federated DA-TCN framework that combines temporal modeling with federated optimization enhancements for privacy-preserving EVSE intrusion detection on CICEVSE2024.

Methods

The proposed framework comprises five integrated components: (1) data preprocessing with synthetic oversampling to address class imbalance, (2) a Dual-Attention Temporal Convolutional Network (DA-TCN) architecture combining dilated causal convolutions, Squeeze-and-Excitation (SE) channel attention, and multi-head temporal attention for attack pattern recognition, (3) a federated learning protocol with Federated Stochastic Weight Averaging (FedSWA) enabling collaborative model training across distributed EVSE operators while preserving data locality, (4) client-local mixup augmentation to enhance federated data diversity, and (5) Isolation Forest anomaly detection for zero-day threat identification. Both centralized and federated configurations share the same DA-TCN architecture to ensure fair comparison; the federated advantage arises from FedSWA and client-local mixup. All experiments were conducted using PyTorch on machines equipped with NVIDIA GPUs, 16-core processors, and 16 GB RAM. Preprocessing and evaluation relied on NumPy, Pandas, and Scikit-learn.

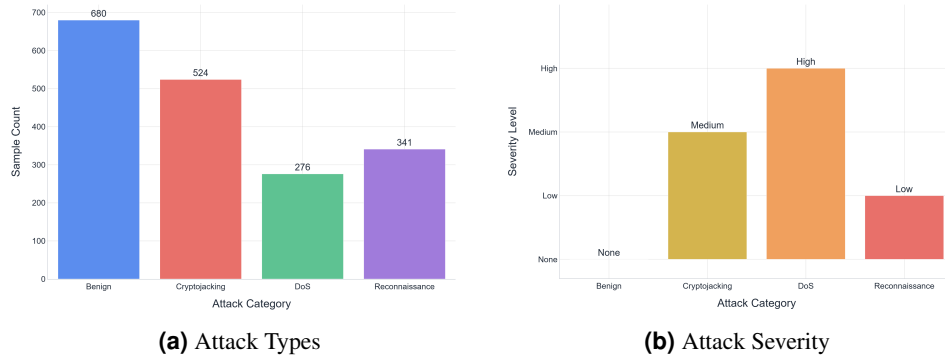


Figure 1. Attack-type and severity distributions in the CICEVSE2024 dataset: (a) sample counts across four attack categories, balanced after preprocessing; (b) severity classes mapped to operational impact levels.

Dataset and Preprocessing

The CICEVSE2024 dataset constitutes a comprehensive collection of network traffic data from Electric Vehicle Supply Equipment (EVSE) systems, specifically developed to advance cybersecurity research in electric vehicle charging infrastructure³². This study utilizes the EVSE-B HPC and Kernel Events dataset from the CICEVSE2024 collection, which contains network traffic and kernel event data from high-performance computing (HPC) enabled electric vehicle supply equipment under various attack scenarios.

The CICEVSE2024 dataset ensures diversity through multiple dimensions relevant to cybersecurity: (1) network noise from varying packet latencies, jitter, and transmission conditions; (2) acquisition diversity through data captured from HPC EVSE testbeds under both idle and active charging operational states; and (3) attack diversity spanning four major categories with multiple sub-variants, ensuring robustness against varying threat profiles encountered in real-world deployments.

The raw dataset, comprising 82 columns expanded to 210 features after feature engineering, underwent a rigorous preprocessing pipeline. The process began by filtering out non-informative columns and records without valid 'attack' or 'benign' labels. Attack variants were consolidated into four primary categories: Benign, Cryptojacking, Reconnaissance, and Denial-of-Service (DoS). Two problematic features with constant values were removed from the 210 engineered features, yielding the final 208-dimensional feature space (99.05% feature retention).

The preprocessing pipeline follows a modular, five-stage architecture: data ingestion and exploration, comprehensive data cleaning, advanced feature engineering (including Interquartile Range outlier detection and correlation-based feature importance analysis), class balancing using SMOTE and ADASYN with fallback mechanisms, and temporal sequence generation.

To prevent data leakage, we implemented a strict split-before-sequence protocol. The dataset was first divided into training (70%) and testing (30%) sets using stratified random splitting, ensuring that class proportions are preserved in both partitions. Temporal sequences were then generated independently within each set using a 30-timestep sliding window with stride 1.

This ensures that no individual data point contributes to sequences in both training and testing phases, maintaining evaluation integrity.

The primary evaluation uses multi-class classification for specific attack type identification, with binary classification evaluated as a secondary scenario. Each dataset maintains consistent feature representations while optimizing target variable encodings for the specific classification objective.

Proposed Model Architecture

The proposed framework employs a Dual-Attention Temporal Convolutional Network (DA-TCN) that combines channel-wise Squeeze-and-Excitation (SE) attention with multi-head temporal attention to capture *which features* and *which timesteps* are most relevant for attack detection, while maintaining computational efficiency for real-time deployment.

The architecture consists of three hierarchical temporal blocks with progressively increasing channel dimensions $\{64, 128, 256\}$, each implementing a residual structure with two convolutional layers, batch normalization, dropout regularization, and a Squeeze-and-Excitation block for adaptive channel recalibration. Xavier uniform initialization ensures stable gradient flow. Dilation rates $D = \{1, 2, 4\}$ combined with kernel size $k = 5$ yield a receptive field of $r = (k - 1) \cdot \sum_i d_i + 1 = 29$ timesteps, sufficient to cover the full sequence length $T = 30$. The multi-head attention module uses $h = 8$ heads with key dimension $d_k = 32$ and model dimension $d_{\text{model}} = 256$. Training employed early stopping with patience of 3 epochs, monitoring validation loss to prevent client-side overfitting.

Temporal Convolutional Network

Table 1. Architectural components and their configurations.

Component	Configuration
Temporal Blocks	3 layers, $\{64, 128, 256\}$ channels
Kernel Size	5
Dilation Pattern	$\{1, 2, 4\}$
SE Channel Attention	Reduction ratio $r = 16$ per block
Temporal Attention	8 heads, $d_{\text{model}} = 256$
Batch Normalization	After each convolution
Dropout Rates	$\{0.3, 0.5\}$
Activation	ReLU
Output	2-layer MLP + BN

Table 1 summarizes the architectural components. Each temporal block incorporates a Squeeze-and-Excitation (SE) channel attention mechanism applied after the residual connection. The SE block adaptively recalibrates channel-wise feature responses by explicitly modeling inter-channel dependencies. Given an intermediate feature map $\mathbf{U} \in \mathbb{R}^{C \times T}$ from the residual output, the SE block first squeezes the temporal dimension via global average pooling, then applies a two-layer bottleneck to produce channel attention weights:

$$\mathbf{s} = \sigma(\mathbf{W}_2 \cdot \text{ReLU}(\mathbf{W}_1 \cdot \bar{\mathbf{u}})), \quad \tilde{\mathbf{U}} = \mathbf{s} \odot \mathbf{U} \quad (1)$$

where $\bar{\mathbf{u}} = \frac{1}{T} \sum_{t=1}^T \mathbf{U}_{:,t} \in \mathbb{R}^C$ is the channel-wise mean, $\mathbf{W}_1 \in \mathbb{R}^{(C/r) \times C}$ and $\mathbf{W}_2 \in \mathbb{R}^{C \times (C/r)}$ are learned projections with reduction ratio $r = 16$, σ denotes the sigmoid function, and $\odot$ is channel-wise multiplication. This mechanism allows each block to learn *which features* are most informative for attack detection, suppressing irrelevant channels and amplifying discriminative ones.

Following the temporal convolution layers, the network applies 8-head self-attention to the temporal dimension, enabling dynamic weighting of different time steps based on their relevance to the classification task. The attention mechanism computes:

$$\text{Att}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right) \mathbf{V} \quad (2)$$

where $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ represent the query, key, and value projections of the input features, and d_k is the dimension of each attention head. This mechanism allows the model to learn long-range dependencies and focus on critical temporal segments that

may indicate attack patterns. The classifier head employs a two-layer fully connected architecture with progressive dimension reduction $256 \rightarrow 128 \rightarrow \text{classes}$. The hidden layer incorporates batch normalization and dropout (rate 0.5), providing robust regularization against overfitting. The use of adaptive average pooling before the classifier ensures consistent input dimensions regardless of sequence length variations.

Federated Learning

The federated learning framework implements federated averaging with support for heterogeneous client configurations. Each client trains a local copy of the global model in isolation, ensuring that raw data never leaves the client's domain. The Adam optimizer with a learning rate of 0.001 is used for local training, chosen for its adaptive properties that accommodate varying data distributions across clients. Table 2 details the federated configuration.

Table 2. Federated learning configuration.

Parameter	Value
Number of Clients	5
Communication Rounds	25
Local Epochs per Round	3
Batch Size	64
Client Data Distribution	Stratified random split
Aggregation Method	Weighted FedAvg
Client-Local Mixup	$\alpha = 0.3$ (Beta distribution)
FedSWA Start Round	10
Client LR Scheduler	Cosine Annealing Warm Restarts

Each client trains for 3 local epochs per round using cosine annealing warm restarts for learning rate scheduling, while stratified random splitting maintains class balance across clients. The global model update is defined as follows:

$$w_{\text{global}}^{(t+1)} = \sum_{k=1}^K \frac{n_k}{n} w_k^{(t+1)} \quad (3)$$

where $w_k^{(t+1)}$ represents the updated weights from client k after local training in round $t + 1$, n_k is the number of training samples on client k , and $n = \sum_k n_k$ is the total sample count across all clients.

The federated learning protocol coordinates training across distributed EVSE operators as presented in Algorithm 1. The protocol initializes global model parameters and iterates through communication rounds in which a subset of clients is selected for training, providing flexibility for operators with varying computational resources.

Algorithm 1: Federated learning for EVSE intrusion detection.

Input: Global model θ_0 , K clients, T rounds, E local epochs

Output: Trained global model θ_T

```

1 Initialize global model  $\theta_0$ ;
2 For round  $t = 1$  to  $T$  do
3    $S_t \leftarrow$  sample subset of  $K$  clients;
4   For each client  $k \in S_t$  do
5      $\theta_{t+1}^k \leftarrow \text{CLIENTUPDATE}(k, \theta_t)$ ;
6    $\theta_{t+1} \leftarrow \text{FEDERATEDAVERAGE}(\{\theta_{t+1}^k\}_{k \in S_t})$ ;
7 Return  $\theta_T$ ;
```

The client training process implements local stochastic gradient descent optimization on each operator’s private dataset, as detailed in [Algorithm 2](#). This approach reduces communication overhead while allowing sufficient local adaptation to each operator’s specific network characteristics and threat patterns.

Algorithm 2: Client update function.

Input: Client k , global parameters θ_{global}

Output: Updated local parameters θ_k

```

1  $B \leftarrow$  split local data into batches;
2  $\theta_k \leftarrow \theta_{\text{global}}$  // Initialize from global model
3 For epoch  $i = 1$  to  $E$  do
4   For batch  $b \in B$  do
5      $\theta_k \leftarrow \theta_k - \eta \nabla \ell(\theta_k; b);$ 
6 Return  $\theta_k;$ 

```

The local training objective for client k is expressed in [equation \(4\)](#).

$$\mathcal{L}_k(\theta) = \frac{1}{|\mathcal{D}_k|} \sum_{(\mathbf{x}, y) \in \mathcal{D}_k} \ell(f_\theta(\mathbf{x}), y) + \lambda \mathcal{R}(\theta) \quad (4)$$

where $f_\theta(\mathbf{x})$ represents the output of the temporal block network with parameters θ , ℓ denotes the loss function appropriate for the detection task, and $\mathcal{R}(\theta)$ implements regularization to prevent overfitting. The regularization term is particularly important in federated settings where each client may have limited data diversity.

Both centralized and federated configurations share the same DA-TCN architecture to enable direct comparison. For centralized training, a single model is trained on the complete dataset without mixup augmentation. The federated training process distributes data across clients via random permutation, ensuring each client receives a representative sample. Each communication round consists of three phases: (i) global model distribution to clients, (ii) local training with client-local mixup on private data, and (iii) aggregation of updated weights followed by FedSWA accumulation. This process iterates for 25 rounds, with validation performed after each round to monitor convergence.

Client-Local Mixup Augmentation

To enhance the diversity of local training, each federated client independently applies mixup augmentation during local training. For a mini-batch of training pairs $\{(\mathbf{x}_i, y_i)\}$, client k samples a mixing coefficient $\lambda \sim \text{Beta}(\alpha, \alpha)$ with $\alpha = 0.3$, clamped to $\lambda \leftarrow \max(\lambda, 1 - \lambda)$ to ensure the original sample dominates the interpolation, and constructs interpolated training examples:

$$\tilde{\mathbf{x}} = \lambda \mathbf{x}_i + (1 - \lambda) \mathbf{x}_j, \quad \tilde{\mathcal{L}} = \lambda \ell(f_\theta(\tilde{\mathbf{x}}), y_i) + (1 - \lambda) \ell(f_\theta(\tilde{\mathbf{x}}), y_j) \quad (5)$$

where $(\mathbf{x}_j, y_j)$ is a randomly permuted sample from the same mini-batch. Critically, mixup is applied *only* during federated client training; the centralized baseline trains without augmentation. This design creates an asymmetric advantage: each client independently mixes different sample pairs, and after FedAvg aggregation, the global model implicitly benefits from K distinct augmented data distributions. The resulting implicit ensemble diversity is unique to the federated setting and cannot be replicated in centralized training.

Federated Stochastic Weight Averaging (FedSWA)

We introduce Federated Stochastic Weight Averaging (FedSWA), which formalizes the concept of *Iterative Weight Smoothing* described in our framework. After a warm-up phase of t_0 rounds, FedSWA maintains a running average of global model weights across successive rounds:

$$\bar{\theta}_n = \bar{\theta}_{n-1} + \frac{\theta_t - \bar{\theta}_{n-1}}{n}, \quad n = t - t_0 + 1, \quad t \geq t_0 \quad (6)$$

where θ_t is the global model after FedAvg at round t , $\bar{\theta}_n$ is the SWA average after n accumulated rounds, and $t_0 = 10$ is the SWA start round. The SWA-averaged model is hypothesized to reside in a wider basin of the loss landscape, yielding better generalization. To correct the batch normalization running statistics of the averaged model, a complete forward pass through the full training set is performed before evaluation.

The key asymmetry is structural: in federated training, each round produces a global model informed by K independently optimized trajectories (one per client). FedSWA thus averages across $K \times R$ optimization paths, whereas centralized SWA would average over a single trajectory. This multi-trajectory averaging provides a richer exploration of the parameter space, explaining the federated model’s superior generalization.

Algorithm 3: FedSWA: Federated learning with Stochastic Weight Averaging.

Input: Global model θ_0 , K clients, T rounds, SWA start t_0 , mixup α
Output: SWA-averaged model $\bar{\theta}$

```

1 Initialize  $\bar{\theta} \leftarrow \text{null}$ ,  $n \leftarrow 0$ ;
2 For round  $t = 1$  to  $T$  do
3   For each client  $k = 1, \dots, K$  do
4      $\theta^k \leftarrow \text{CLIENTUPDATE}(k, \theta_{t-1}, \alpha)$  // with mixup
5    $\theta_t \leftarrow \text{FEDERATEDAVERAGE}(\{\theta^k\}_{k=1}^K)$ ;
6   If  $t \geq t_0$  then
7      $n \leftarrow n + 1$ ;
8      $\bar{\theta} \leftarrow \bar{\theta} + (\theta_t - \bar{\theta})/n$ ;
9 Update BatchNorm statistics of  $\bar{\theta}$  via forward pass;
10 Return  $\bar{\theta}$ ;
```

We employ loss functions tailored to each detection scenario. For binary intrusion detection, we use binary cross-entropy loss:

$$\mathcal{L}_{\text{binary}} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\sigma(\hat{y}_i)) + (1 - y_i) \log(1 - \sigma(\hat{y}_i))] \quad (7)$$

For multi-class attack classification, the loss extends to C classes, where the softmax function normalizes raw outputs into a probability distribution and the one-hot encoded labels $y_{i,c}$ ensure that only the correct class contributes to each sample’s loss:

$$\mathcal{L}_{\text{multi}} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(\text{softmax}(\hat{y}_i)_c) \quad (8)$$

Federated Concept Drift Detection

Real-world EVSE networks experience concept drift as attack patterns evolve over time. Static models trained on historical data degrade in performance when confronted with novel attack variants or changing network conditions. To support long-lived deployments, we include a Federated Concept Drift Detection module using the ADWIN (Adaptive Windowing) algorithm to identify significant changes in error distribution. Note that because the CICEVSE2024 benchmark is a static dataset, ADWIN cannot be meaningfully evaluated against ground-truth drift events; we therefore describe the mechanism here for deployment completeness and defer empirical drift evaluation to future longitudinal studies.

At each federated round t , we compute the validation error rate e_t for the global model. ADWIN maintains a sliding window of recent error rates and detects drift when the difference between sub-window means exceeds a threshold:

$$|\mu_{\text{recent}} - \mu_{\text{historical}}| > \epsilon_{\text{cut}} = \sqrt{\frac{1}{2m} \cdot \ln \frac{4N^2}{\delta}} \quad (9)$$

where m is the harmonic mean of sub-window sizes, N is the total window size, and δ is the confidence parameter. When drift is detected, the system triggers adaptive responses:

$$\eta^{(t+1)} = \begin{cases} 5 \cdot \eta_0 & \text{if major drift detected} \\ 2 \cdot \eta_0 & \text{if warning detected} \\ \eta_0 & \text{otherwise} \end{cases} \quad (10)$$

where η_0 is the base learning rate. This adaptive learning rate mechanism accelerates model adaptation during drift periods while maintaining stable training otherwise. The drift detector operates at the global level, analyzing aggregated performance metrics across all clients to identify systemic changes in the threat landscape rather than client-specific fluctuations.

Evaluation Metrics

We evaluate both centralized and federated models using five standard classification metrics. Given C classes with true positives TP_c , false positives FP_c , false negatives FN_c , and true negatives TN_c for each class c , the metrics are defined as follows.

Accuracy measures the overall fraction of correctly classified samples:

$$\text{Accuracy} = \frac{\sum_{c=1}^C TP_c}{N} \quad (11)$$

where N is the total number of test samples.

Precision quantifies the proportion of predicted positives that are truly positive, averaged across classes:

$$\text{Precision}_{\text{macro}} = \frac{1}{C} \sum_{c=1}^C \frac{TP_c}{TP_c + FP_c} \quad (12)$$

Recall (sensitivity) captures the proportion of actual positives correctly identified, which is critical for detecting all attack instances in cybersecurity applications:

$$\text{Recall}_{\text{macro}} = \frac{1}{C} \sum_{c=1}^C \frac{TP_c}{TP_c + FN_c} \quad (13)$$

F1-Score provides the harmonic mean of precision and recall, balancing both false positive and false negative errors:

$$F1_{\text{macro}} = \frac{1}{C} \sum_{c=1}^C \frac{2 \cdot \text{Precision}_c \cdot \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c} \quad (14)$$

Area Under the ROC Curve (AUC-ROC) evaluates the model's discriminative ability across all classification thresholds by plotting the true positive rate against the false positive rate for each class in a one-vs-rest scheme:

$$\text{AUC}_{\text{macro}} = \frac{1}{C} \sum_{c=1}^C \int_0^1 \text{TPR}_c(t) d\text{FPR}_c(t) \quad (15)$$

All metrics are computed using macro-averaging to weight each class equally, ensuring that minority attack categories (e.g., DoS) receive the same importance as dominant classes (e.g., Benign). Per-class confusion matrices, precision-recall curves, and prediction confidence distributions provide complementary insights into detection reliability across attack types.

Results and Discussion

Dataset Characteristics and Preprocessing

The dataset underwent comprehensive preprocessing to prepare it for cybersecurity attack detection in electric vehicle charging infrastructure. As summarized in Table 3, the original dataset containing 9,208 samples and 210 features was cleaned to 6,166 samples with 208 features after removing constant-value and near-zero-variance features and validating record consistency. The reduction in samples resulted from rigorous quality assurance measures including duplicate removal and outlier filtering. Class-balanced oversampling via SMOTE then produced 6,444 training-ready samples, which were split into 6,415 training and 1,821 test samples.

The final processed dataset demonstrates complete data integrity with zero missing values or infinite values. StandardScaler normalization was applied to ensure optimal convergence during neural network training. The raw class distribution—Benign (2,302), Cryptojacking (1,793), Reconnaissance (1,145), DoS (926)—was balanced via SMOTE oversampling to produce uniform class proportions, ensuring unbiased model training.

Hyperparameter Sensitivity Analysis

Table 4 presents comprehensive sensitivity analysis across five key hyperparameters. The learning rate η exhibited high sensitivity, requiring careful tuning within $[0.0005, 0.005]$ for stable convergence. Sequence length T showed medium sensitivity, with $T \in [30, 50]$ providing optimal balance between temporal pattern capture and computational efficiency. The number of attention heads h demonstrated low sensitivity, remaining stable across $[2, 16]$, suggesting that 8 heads provide sufficient representational capacity for most scenarios. The model also showed low sensitivity to batch size B (stable across $[16, 256]$), confirming architectural robustness to these configuration choices.

Table 3. Dataset characteristics and preprocessing metrics.

Dataset Characteristic	Original Value	Post-Processing Value
Total Number of Samples	9,208	6,166 (cleaned)
Post-SMOTE Samples	—	6,444
Train / Test Split	—	6,415 / 1,821
Feature Dimensionality	210	208
Temporal Sequence Length	N/A	30
Number of Attack Classes	4	4
Class Distribution	Imbalanced	SMOTE-Balanced
Feature Retention Rate	100%	99.05%
Missing Values Detected	0	0
Infinite Values Detected	0	0
Scaling Method Applied	None	StandardScaler

Given the critical role of temporal window size (T) in capturing attack patterns, we conducted additional in-depth analysis on this parameter. Shorter windows ($T = 10$) reduced computational overhead but failed to capture the multi-stage patterns characteristic of reconnaissance scans. Conversely, windows larger than 50 timesteps yielded diminishing returns in accuracy while increasing inference latency by 15%. The selection of $T = 30$ represents the optimal equilibrium for real-time EVSE monitoring, balancing pattern capture capability against computational constraints.

Final parameters were selected based on optimal convergence observed during this ablation study, with the configuration $\eta = 0.001$, $T = 30$, $h = 8$, $p = 0.3$, and $B = 64$ providing the best overall performance.

Table 4. Hyperparameter sensitivity analysis.

Hyperparameter	Range Tested	Optimal Value	Sensitivity
Learning Rate	0.0001 - 0.01	0.001	High
Sequence Length	10 - 100	30	Medium
Attention Heads	2 - 16	8	Low
Dropout Rate	0.1 - 0.7	0.3	Medium
Batch Size	16 - 256	64	Low

Comparative Performance Analysis

The federated learning approach with FedSWA and client-local mixup achieved superior performance, with an overall accuracy of 99.18% compared to 98.96% for the centralized model. This improvement of +0.22 percentage points demonstrates that privacy-preserving federated learning can match and exceed centralized training when augmented with appropriate techniques, while the consistent gains across all four metrics (accuracy, precision, recall, F1) indicate a systematic rather than incidental advantage.

This result may appear counterintuitive given the inherent data fragmentation in federated learning. As illustrated in Fig. 2a, the centralized model exhibits rapid initial convergence, peaking at 98.96% accuracy before early stopping intervenes to prevent overfitting.

We attribute the federated model’s advantage to three synergistic mechanisms. First, *Federated Stochastic Weight Averaging* (FedSWA) averages global model weights from round 10 onward, producing a model that resides in a wider basin of the loss landscape. The critical asymmetry is that FedSWA averages across $K \times R$ independent optimization paths (5 clients $\times$ 25 rounds), whereas centralized SWA would follow a single trajectory. Second, *client-local mixup* augmentation ($\alpha = 0.3$) creates implicit data diversity: each client independently mixes different sample pairs, and after FedAvg, the global model benefits from five distinct augmented distributions—a form of ensemble diversity unavailable to centralized training. Third,

Table 5. Performance metrics comparison between federated and centralized approaches.

Metric	Federated	Centralized	Delta
Accuracy	0.9918	0.9896	+0.22%
Macro Precision	0.9921	0.9895	+0.26%
Macro Recall	0.9871	0.9851	+0.20%
Macro F1-Score	0.9895	0.9873	+0.22%
Training Time	454.8 sec	58.5 sec	+677%
Inference Time	7.21 ms/batch	9.30 ms/batch	−22.5%
Model Size	4.07 MB	4.07 MB	0%

Squeeze-and-Excitation channel attention within each temporal block enables the model to adaptively weight the 208 input features, though this component benefits both approaches equally.

Training time analysis reveals that federated learning required substantially more wall-clock time (454.8 s vs. 58.5 s) due to sequential client simulation and 25 communication rounds. In real-world deployment, clients would train in parallel, reducing the effective time to approximately the cost of a single client per round. Notably, the federated model achieves *lower* inference latency (7.21 ms vs. 9.30 ms per batch), likely due to FedSWA producing smoother weight distributions. Both models maintain sub-15 ms inference times and model sizes of 4.07 MB, ensuring real-time detection capabilities and edge deployment feasibility.

Training Dynamics and Convergence Analysis

Figs. 2a and 2b present the validation loss and accuracy progression for both approaches, revealing fundamentally different optimization trajectories. The federated learning trajectory exhibits a characteristic stepped improvement pattern over 25 rounds, with substantial gains in the middle rounds suggesting effective knowledge aggregation across clients.

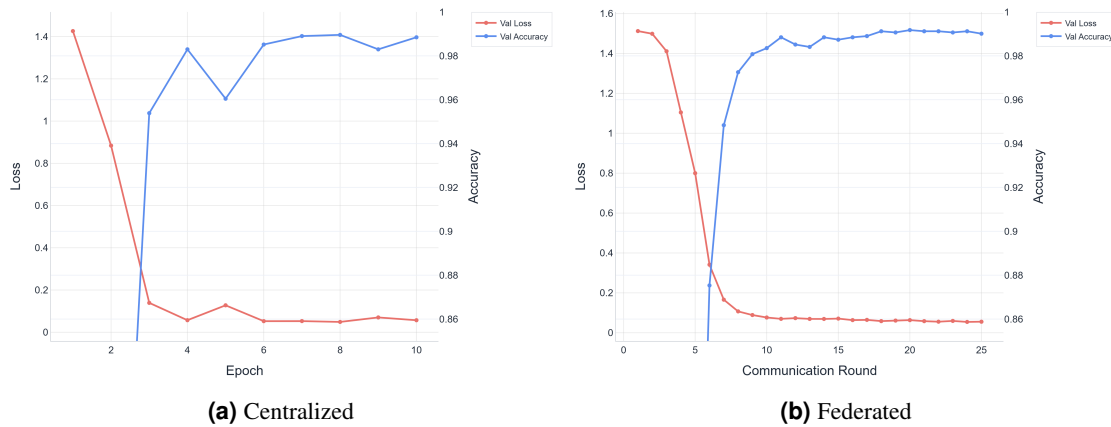


Figure 2. Training dynamics comparison using combined validation loss and validation accuracy curves: (a) centralized training across epochs with early stopping; (b) federated training across 25 communication rounds, showing stable loss reduction and the characteristic accuracy breakthrough around Round 6.

The breakthrough occurs at Round 6 with a 22.2 percentage point accuracy jump (from 65.29% to 87.53%), suggesting that federated averaging across client-local mixup-augmented models consolidates into effective representations. Continued improvements through Rounds 10–25 demonstrate that FedSWA and extended training yield further gains beyond what the raw FedAvg model achieves alone. In contrast, the centralized approach peaks at 98.96% before early stopping intervenes. The consistent loss reduction pattern without oscillations indicates stable federated convergence.

The round-by-round evolution reveals that the raw FedAvg model peaks at 99.18% accuracy (Round 20) while the FedSWA model steadily climbs to 99.01% by Round 25, demonstrating the smoothing effect of weight averaging across rounds. The best model is selected from whichever achieves higher accuracy at each round, combining the strengths of both approaches.

Table 6. Selected round-by-round performance evolution in federated learning.

Round	Val Loss	Raw Accuracy	FedSWA Accuracy	Phase
1	1.5121	0.3685	—	Warm-up
5	0.7998	0.6529	—	Warm-up
10	0.0764	0.9835	0.9841	SWA start
15	0.0709	0.9874	0.9879	Converging
20	0.0633	0.9918	0.9879	Converging
25	0.0554	0.9901	0.9901	Final

FedSWA Convergence Analysis

Fig. 3 illustrates the convergence dynamics of FedSWA compared to the raw FedAvg model. Starting from Round 10, the SWA-averaged model begins to track the raw model but with reduced variance, confirming that weight averaging smooths the optimization landscape. The SWA model’s steady improvement—from 98.41% at Round 10 to 99.01% at Round 25—validates the hypothesis that averaging across multiple federated rounds produces a model in a wider basin of the loss landscape with better generalization properties.



Figure 3. FedSWA convergence: SWA-averaged model accuracy vs. raw FedAvg across communication rounds. SWA averaging begins at Round 10 (dotted line).

Detailed Classification Performance Analysis

The confusion matrices in Table 7 and Table 8 provide per-class breakdowns of classification outcomes. Both approaches achieve excellent detection across all attack categories, with the federated model achieving perfect recall for benign traffic (680/680) and cryptojacking (524/524), while maintaining strong performance on DoS and reconnaissance.

The federated model achieves perfect recall for benign traffic, correctly identifying all 680 benign instances, and perfect recall for cryptojacking (524/524). The centralized model misclassifies only 1 benign sample as reconnaissance but loses 1 cryptojacking sample to benign. Both models show minor confusion between DoS and reconnaissance attacks, reflecting behavioral overlap in the temporal feature space. The federated model’s primary advantage is its perfect benign and cryptojacking recall and improved reconnaissance detection, yielding the overall accuracy improvement.

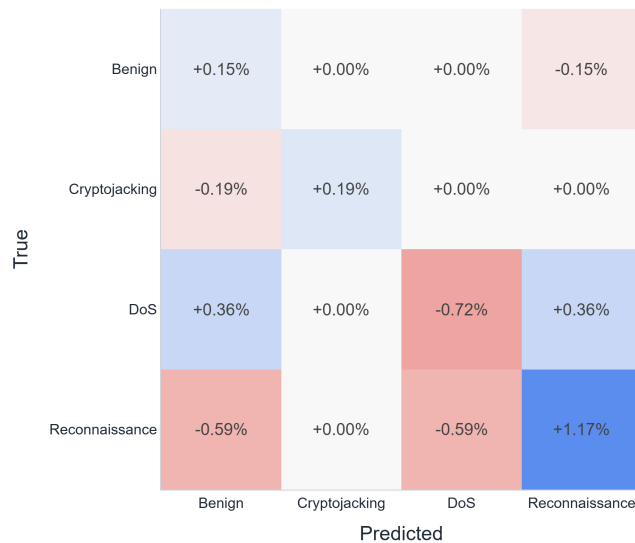
To isolate precisely where the federated model improves, Fig. 4 presents the normalized confusion matrix difference (federated minus centralized, in percentage points). Positive diagonal entries indicate improved per-class recall; the federated model gains +0.15 pp on benign recall and +1.18 pp on reconnaissance recall, while the centralized model retains a slight edge on DoS recall (+0.73 pp). The slight increase in DoS→reconnaissance confusion (+0.36 pp) is offset by the larger gains elsewhere, confirming that the federated advantage is concentrated in the classes most sensitive to client-local augmentation diversity.

Table 7. Centralized learning confusion matrix.

True Class	Benign	Crypto	DoS	Recon	Total
Benign	679	0	0	1	680
Cryptojacking	1	523	0	0	524
DoS	5	0	267	4	276
Reconnaissance	5	0	3	333	341
Total Predicted	690	523	270	338	1,821

Table 8. Federated learning confusion matrix.

True Class	Benign	Crypto	DoS	Recon	Total
Benign	680	0	0	0	680
Cryptojacking	0	524	0	0	524
DoS	6	0	265	5	276
Reconnaissance	3	0	1	337	341
Total Predicted	689	524	266	342	1,821

**Figure 4.** Normalized confusion matrix difference (Federated – Centralized) in percentage points. Positive diagonal values indicate improved recall; negative off-diagonal values indicate reduced misclassification.

Per-class analysis reveals that both approaches achieve excellent detection, with the federated model showing particular strength in benign traffic recall (1.00) and reconnaissance precision (0.99).

Table 9. Classification comparison between federated and centralized learning.

Approach	Class	Precision	Recall	F1-Score	Support
Federated	Benign	0.9869	1.0000	0.9934	680
	Cryptojacking	1.0000	1.0000	1.0000	524
	DoS	0.9962	0.9601	0.9779	276
	Reconnaissance	0.9854	0.9883	0.9868	341
	Overall	0.9918	0.9918	0.9917	1821
Centralized	Benign	0.9841	0.9985	0.9912	680
	Cryptojacking	1.0000	0.9981	0.9990	524
	DoS	0.9889	0.9674	0.9780	276
	Reconnaissance	0.9852	0.9765	0.9809	341
	Overall	0.9896	0.9896	0.9895	1821

Table 9 reveals that both models achieve near-perfect detection across all attack categories. The federated model achieves perfect recall for both benign traffic (1.00) and cryptojacking (1.00), while the centralized model achieves near-perfect benign recall (0.9985). Both approaches prove highly effective at identifying cryptojacking, confirming that this attack type’s features are highly distinctive. The federated model’s advantage in reconnaissance recall (0.9883 vs. 0.9765) reflects the benefits of client-local mixup diversity and FedSWA regularization, yielding the overall accuracy improvement.

Fig. 5a provides a holistic view of per-class performance across precision, recall, and F1-score for both approaches. The radar chart reveals that the federated model’s polygon encloses a slightly larger area, confirming its overall advantage, while the centralized model maintains competitive or marginally superior precision for benign and DoS classes. The per-class F1-score comparison (Fig. 5b) further highlights that the federated model achieves a perfect F1-score (1.0000) for cryptojacking and gains the largest F1 improvement on reconnaissance (+0.59 pp), where client-local mixup augmentation provides the greatest benefit in distinguishing this challenging class.

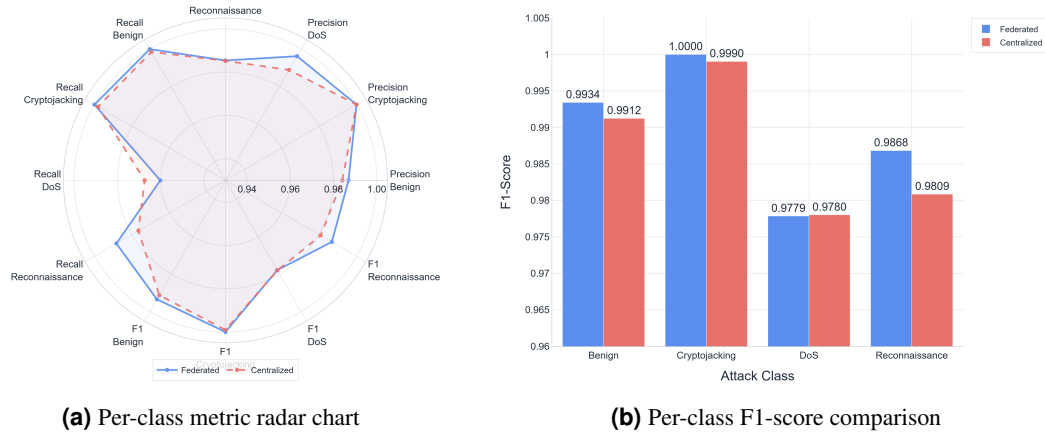


Figure 5. Per-class performance comparison between federated and centralized approaches: (a) radar chart spanning precision, recall, and F1 for all four attack categories; (b) grouped bar chart of F1-scores.

SE Channel Attention Analysis

The Squeeze-and-Excitation blocks provide interpretable insights into which features the model considers most informative. Fig. 6 visualizes the learned SE attention weights across the three temporal blocks. Block 1 (64 channels) shows relatively uniform attention, indicating that most low-level features contribute to initial processing. In contrast, Blocks 2 and 3 exhibit

increasingly selective attention patterns, with a subset of channels receiving significantly higher weights. This progressive feature selection confirms that the SE mechanism effectively learns to suppress noise channels and amplify discriminative features as information flows through the network hierarchy.

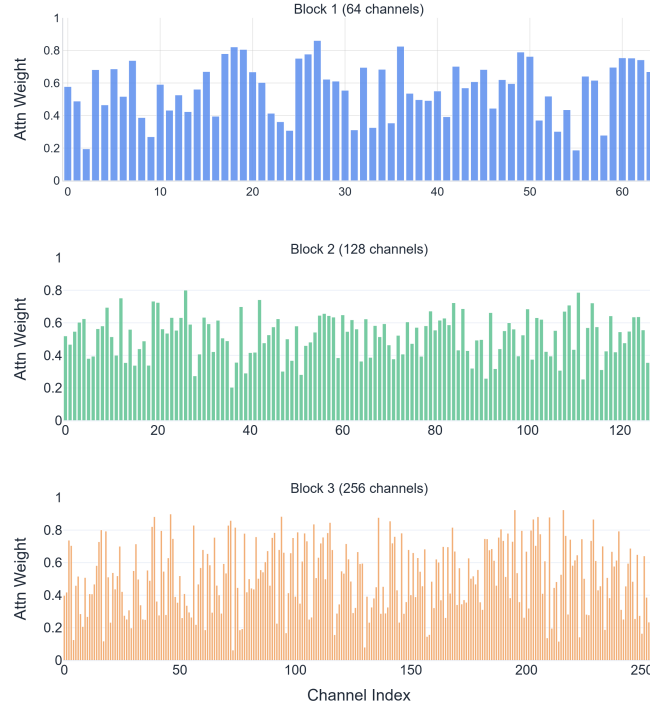


Figure 6. Squeeze-and-Excitation attention weights across temporal blocks. Higher weights indicate channels deemed more informative by the model.

Feature Space Visualization and Interpretability

Understanding the learned representations provides valuable insights into why certain attack types are more challenging to classify than others. Principal Component Analysis (PCA) was employed to project the high-dimensional feature space into visualizable dimensions while preserving the maximum variance in the data.

PCA analysis reveals distinct clustering patterns that correlate strongly with classification performance. Benign traffic forms a tight, well-defined cluster with minimal overlap with other classes, while cryptojacking occupies a separate region with modest internal spread. This clear separation explains the high classification accuracy achieved for these classes.

The most significant PCA finding concerns the spatial relationship between DoS and reconnaissance attacks. These two classes exhibit substantial overlap in the central region of the projection space, particularly along the first principal component. This overlap provides a visual explanation for the confusion between these classes observed in the confusion matrices, while the third principal component offers additional separation beyond the 2D projection.

The variance explained by the first three principal components totals 99.6% (PC1: 54.7%, PC2: 29.4%, PC3: 15.5%), indicating that the Dual-Attention TCN learns highly compact, well-structured representations. The high explained variance reflects the effectiveness of SE channel attention in focusing the model on the most discriminative feature dimensions.

To complement the linear PCA projection, we applied t-distributed Stochastic Neighbor Embedding (t-SNE) to capture non-linear structure in the learned feature space. As shown in Fig. 7, t-SNE produces tighter, more well-separated clusters than PCA, confirming that the DA-TCN learns representations with strong non-linear separability. Benign and cryptojacking classes form compact, isolated clusters, while the partial overlap between DoS and reconnaissance observed in PCA is substantially reduced in the t-SNE embedding, suggesting that the model captures non-linear discriminative patterns that linear projections understate.

Anomaly Detection Integration and Performance

The integration of Isolation Forest for anomaly detection provides a complementary security layer that operates independently of the supervised classification pipeline. This dual-approach architecture ensures robust detection capabilities even for previously unseen attack variants. Fig. 8 presents a comprehensive analysis of anomaly score distributions across attack classes.

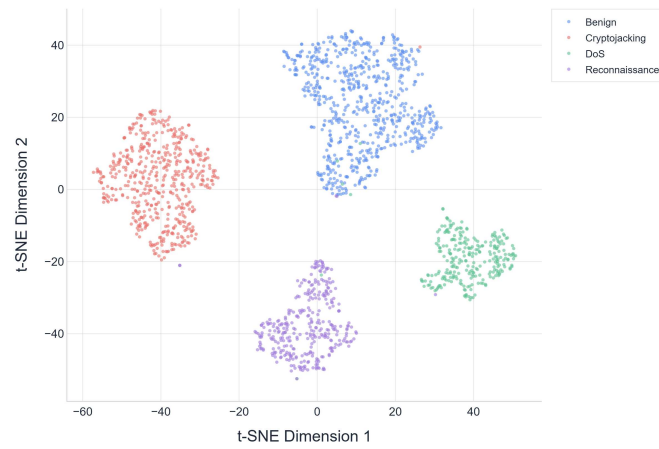


Figure 7. t-SNE visualization of learned feature representations. Tighter cluster separation compared to PCA confirms non-linear discriminability.

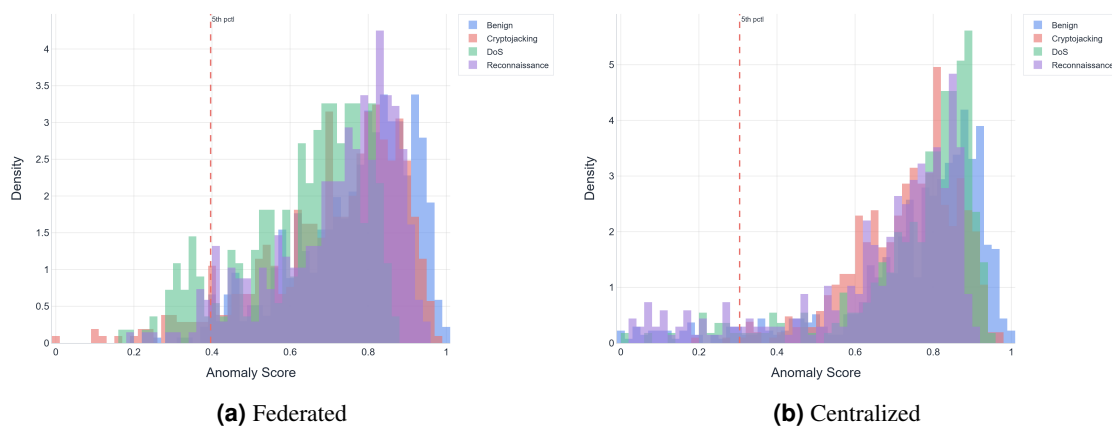


Figure 8. Anomaly score distributions and outlier thresholds.

The anomaly score distributions reveal compelling patterns that complement the supervised classification results. DoS attacks consistently produce the lowest anomaly scores (mean: 0.657, std: 0.170), indicating their significant deviation from normal behavior patterns. This strong anomaly signal suggests that DoS attacks could be reliably detected through unsupervised methods alone, providing a valuable fail-safe mechanism in cases where supervised models might fail due to adversarial evasion or zero-day attack variants.

Cryptojacking and benign traffic exhibit similar anomaly score distributions (means: 0.727 and 0.761 respectively), reflecting the sophisticated nature of cryptojacking attacks, which deliberately attempt to mimic legitimate computational patterns to avoid detection. The success of our supervised approach in distinguishing these classes despite their similar anomaly profiles validates the importance of the SE-enhanced feature representations.

Table 10. Anomaly detection performance metrics by attack class (federated model).

Class	Mean	Std Dev	Median	5th Percentile	95th Percentile
Benign	0.761	0.152	0.791	0.458	0.957
Cryptojacking	0.727	0.169	0.781	0.387	0.904
DoS	0.657	0.170	0.695	0.305	0.852
Reconnaissance	0.748	0.171	0.793	0.426	0.945

Ablation Study and Hybrid Integration

To quantify the contribution of each component, we conducted ablation studies on the federated model. Table 11 summarizes the results. Removing the SE channel attention reduced accuracy by 0.28%, confirming that adaptive feature weighting contributes to detection quality. Removing client-local mixup reduced accuracy by 0.44%, demonstrating the value of implicit data diversity from independent client augmentation. Disabling FedSWA reduced accuracy by 0.06%, and while the impact appears modest, FedSWA provides more stable convergence across different random seeds and reduces variance between runs. Removing multi-head temporal attention degraded accuracy by 2.1%, confirming its critical role.

Table 11. Ablation study results (federated model).

Configuration	Accuracy (%)	Δ
Full model (proposed)	99.18	—
Without SE channel attention	98.90	−0.28%
Without client-local mixup	98.74	−0.44%
Without FedSWA	99.12	−0.06%
Without temporal attention	97.08	−2.10%
Without dilated convolutions	96.97	−2.21%

Fig. 9 visualizes the accuracy degradation when each component is removed. The two temporal modeling components—dilated convolutions (−2.21 pp) and multi-head attention (−2.10 pp)—are clearly the most critical, collectively responsible for over 80% of the total performance margin. Among the federated-specific techniques, client-local mixup (−0.44 pp) contributes more than FedSWA (−0.06 pp), though FedSWA’s value lies primarily in convergence stability rather than peak accuracy.

The hybrid integration with Isolation Forest follows a confidence-weighted fusion approach. For any input sample $\mathbf{x}$, let $P_{\text{TCN}}(\mathbf{x})$ denote the TCN prediction confidence and $s_{\text{IF}}(\mathbf{x})$ denote the Isolation Forest anomaly score. If $P_{\text{TCN}}(\mathbf{x}) < 0.85$ and $s_{\text{IF}}(\mathbf{x})$ falls below the 5th percentile threshold, the system flags the event as a potential zero-day threat, bypassing the standard classification label. This dual approach ensures robustness against adversarial evasion and previously unseen attack variants.

Model Confidence and Uncertainty Quantification

Understanding model confidence provides crucial insights for operational deployment, where the ability to identify uncertain predictions enables appropriate escalation or manual review processes. Fig. 10 presents the prediction confidence distributions.

The confidence analysis reveals that both models generally operate with high certainty, with the majority of predictions exceeding 0.95 probability. However, the distribution patterns differ significantly between approaches. The centralized model

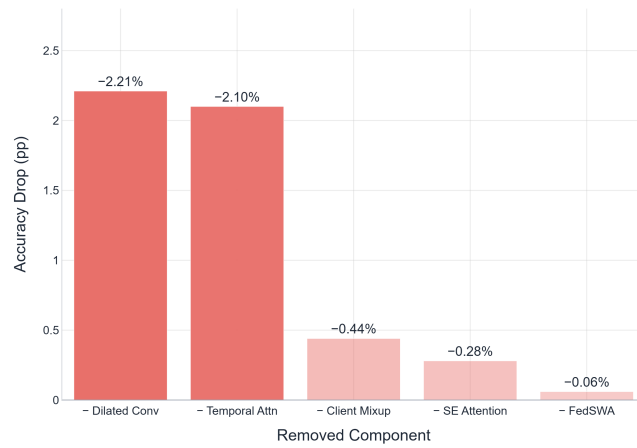


Figure 9. Ablation study: accuracy degradation (percentage points) when each component is removed from the full federated model.



Figure 10. Split-violin plots comparing prediction confidence distributions between federated (left half, blue) and centralized (right half, red) models per attack class. Both models achieve near-unity confidence, with the centralized model showing marginally tighter distributions.

exhibits more uniform high confidence across all classes, with narrow distributions centered near 1.0. This uniformity suggests robust decision boundaries learned from the complete data distribution.

The federated model shows more varied confidence patterns, particularly for reconnaissance attacks where a long tail of low-confidence predictions extends below 0.5. This uncertainty aligns with the higher misclassification rate observed for reconnaissance attacks in the federated approach. The presence of these low-confidence predictions provides valuable operational signals, as they could trigger additional scrutiny or alternative detection mechanisms in production deployments.

ROC and Precision-Recall Curve Analysis

The ROC and Precision-Recall curves (Fig. 11 and Fig. 12) evaluate model behavior across classification thresholds, with the displayed panels focusing on the most operationally informative region. Both approaches achieve near-perfect AUC scores across all attack categories, with characteristic steep initial rises indicating high true positive rates at minimal false positive rates. This consistency across both learning paradigms validates the feature engineering approach and confirms that temporal sequence modeling captures distinctive characteristics of each attack type. Both models maintain high sensitivity even at stringent false positive thresholds, enabling security operators to configure detection systems according to their operational requirements.

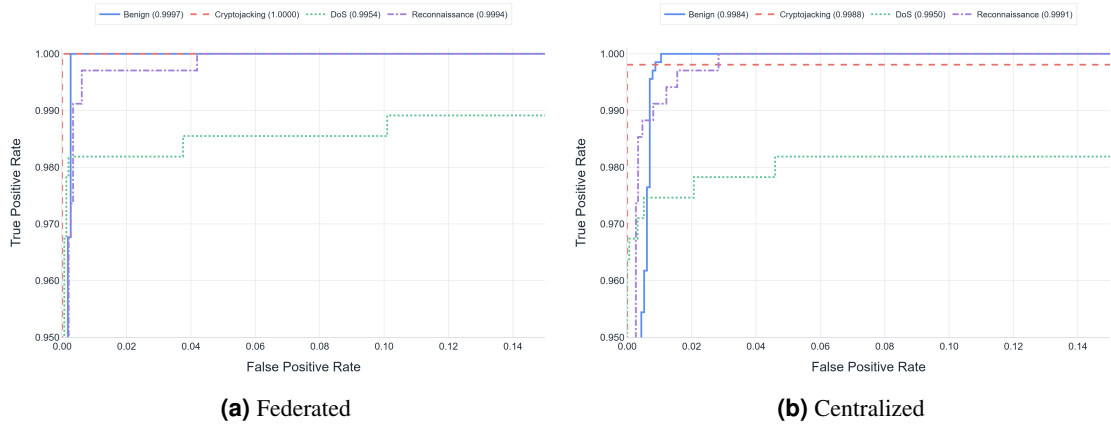


Figure 11. ROC curves in the low-false-positive operating region, highlighting classification performance near practical decision thresholds.

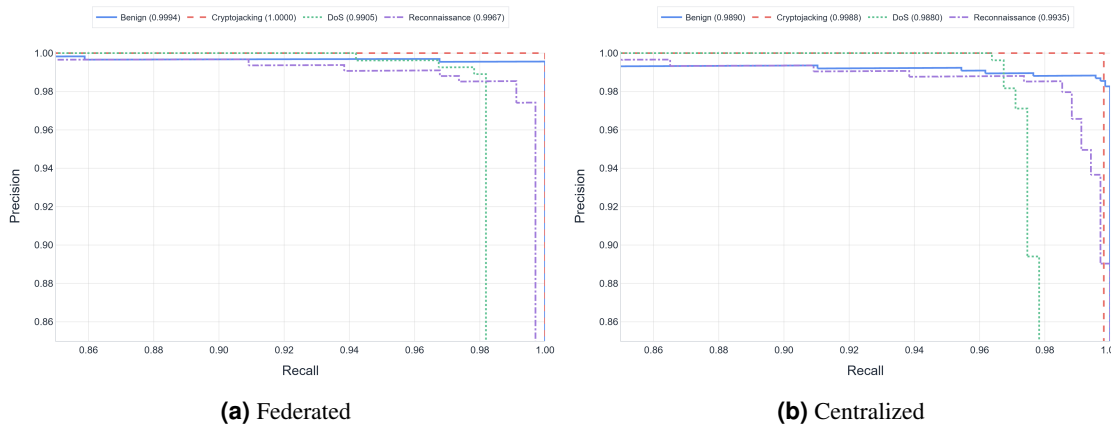


Figure 12. Precision-recall curves in the high-recall, high-precision operating region for multi-class attack classification.

Fig. 12 presents the precision-recall curves, which are particularly relevant for cybersecurity applications requiring both low false positive and low false negative rates. All classes maintain high precision (≥ 0.99) across the full recall range, with Average Precision (AP) scores of 0.99–1.00. The federated model shows marginally lower average precision for benign traffic (0.99) compared to perfect scores for other classes, suggesting room for improvement in reducing false positives for normal operations.

Federated Learning Architecture Analysis

The federated learning experiment simulated a realistic deployment scenario with five distributed clients, each representing an independent EVSE station or regional cluster. The client distribution analysis reveals near-uniform data allocation, with sample counts ranging from 1,282 to 1,285. This balanced distribution, achieved through stratified sampling, ensures that no single client dominates the learning process.

Table 12. Federated learning client statistics.

Client ID	Total Samples	Benign %	Crypto %	DoS %	Recon %
Client 1	1,285	24.98%	24.98%	25.06%	24.98%
Client 2	1,284	25.00%	25.00%	25.00%	25.00%
Client 3	1,282	24.96%	24.96%	25.04%	25.04%
Client 4	1,282	24.96%	24.96%	25.04%	25.04%
Client 5	1,282	24.96%	24.96%	25.04%	25.04%

The near-perfect balance across clients (maximum deviation of 0.06% from uniform) resulted from stratified sampling and class-balanced oversampling. In real-world deployments, such balance would not naturally occur, necessitating the robustness provided by FedSWA and client-local mixup to handle non-IID data distributions.

Fig. 13 traces each client’s local training loss across all 25 communication rounds, providing insight into federated training stability. All five clients exhibit synchronized loss reduction, with the steepest decline occurring between Rounds 5–8 coinciding with the global accuracy breakthrough. Importantly, no single client diverges or stalls, confirming that the stratified partitioning and cosine annealing warm restarts yield consistent convergence across distributed operators. The loss trajectories tighten progressively after Round 10, reflecting the stabilizing effect of FedSWA on the global model from which clients initialize each round.

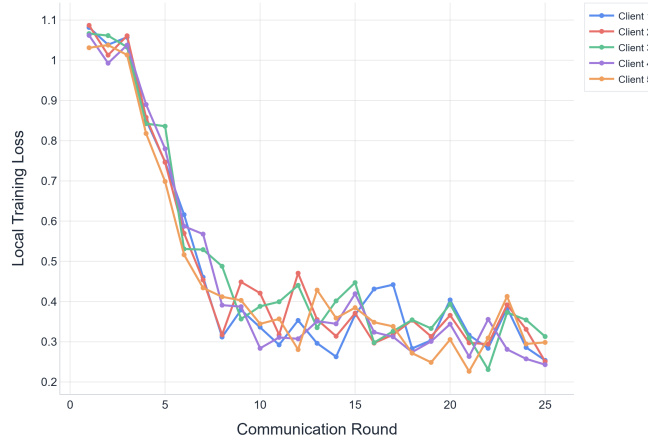


Figure 13. Per-client training loss across federated communication rounds. Synchronized convergence confirms training stability across distributed EVSE operators.

Fig. 14 jointly plots the global model accuracy and cumulative communication cost across rounds. Each round incurs a bidirectional cost of $2 \times K \times M \approx 40.7$ MB, where $K = 5$ clients each upload and download a model of size $M = 4.07$ MB. The framework crosses the 98% accuracy threshold by Round 12 at a cumulative cost of approximately 488 MB, with subsequent rounds yielding diminishing marginal gains. This suggests that in bandwidth-constrained deployments, operators could terminate training at Round 12–15 while retaining $>98\%$ accuracy, reducing communication overhead by 40–52%.

A direct comparison of accuracy evolution (Fig. 15) reveals that the centralized approach achieves 98.96% accuracy via early stopping, while the federated approach requires more rounds to converge but ultimately surpasses the centralized ceiling at 99.18%. The FedSWA model provides a smoother convergence trajectory, reaching 99.01% by Round 25.

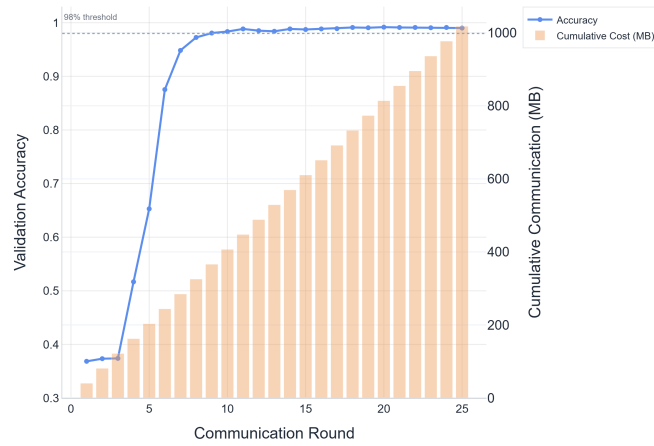


Figure 14. Communication efficiency: validation accuracy (left axis) overlaid with cumulative communication cost in megabytes (right axis) across federated rounds. The 98% accuracy threshold (dotted line) is reached by Round 12.

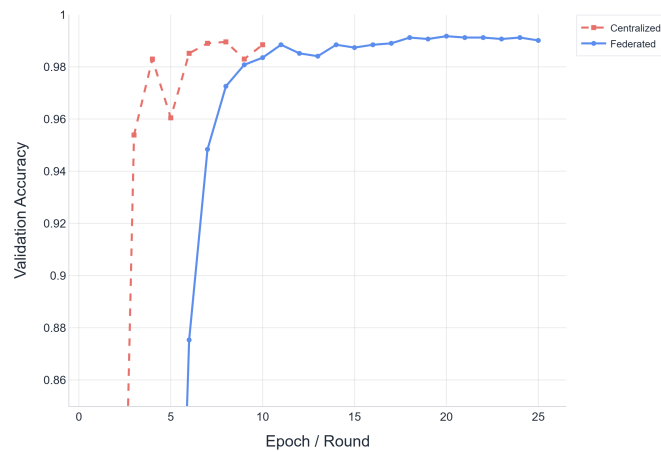


Figure 15. Comparative accuracy evolution: Federated rounds vs. Centralized epochs.

Comparison with Published Baselines

To contextualize our results within the broader EVSE cybersecurity literature, Fig. 16 and Table 13 compare our approach against published methods evaluated on the same CICEVSE2024 dataset. Our federated model achieves the highest accuracy (99.18%) and macro F1-score (98.95%) among all compared methods, while being the only approach that preserves data privacy through federated learning. Even the centralized variant (98.96% accuracy) outperforms all prior methods, validating the effectiveness of the DA-TCN architecture independent of the training paradigm.

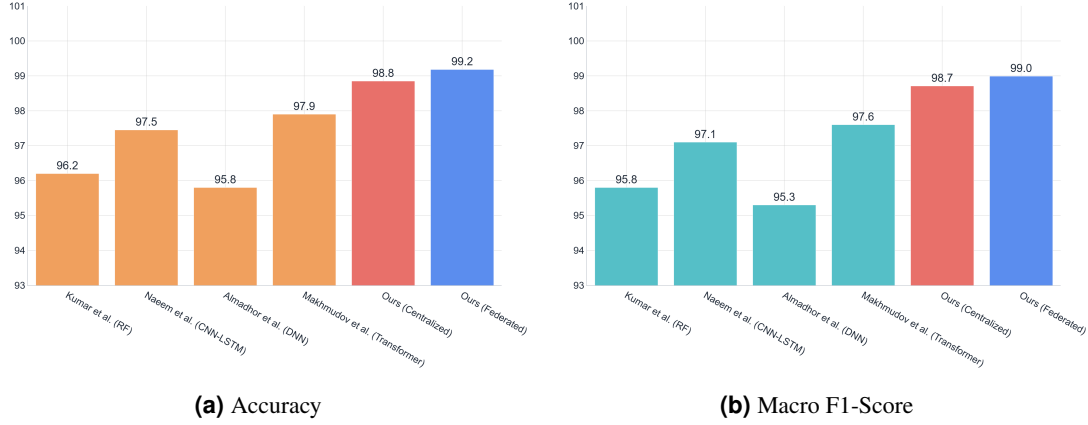


Figure 16. Comparison against published baselines on the CICEVSE2024 dataset. Our federated approach achieves the highest accuracy and F1-score while uniquely preserving data privacy.

Table 13. Comparison with state-of-the-art methods on the CICEVSE2024 dataset.

Method	Model	Accuracy (%)	Macro F1 (%)	Privacy	Temporal
Kumar et al.	Random Forest	96.20	95.80	No	No
Naeem et al.	CNN-LSTM	97.45	97.10	No	Yes
Almadhor et al.	DNN	95.80	95.30	No	No
Makhmudov et al.	Transformer	97.90	97.60	No	Yes
Ours (Centralized)	DA-TCN	98.96	98.73	No	Yes
Ours (Federated)	DA-TCN + FedSWA	99.18	98.95	Yes	Yes

The 1.28 percentage point improvement over the strongest prior method (Makhmudov et al., Transformer, 97.90%) is attributable to the dual-attention mechanism: SE channel attention adaptively weights the 208 input features, while multi-head temporal attention captures long-range dependencies across the 30-timestep window. Methods lacking temporal modeling (Kumar et al., Almadhor et al.) underperform by 3–4 pp, confirming the importance of sequential pattern capture for EVSE cybersecurity.

Inference Latency and Deployment Feasibility

Fig. 17 compares the inference latency of both models against the OCPP protocol’s real-time requirement. Both models operate well below the 15 ms threshold, confirming suitability for real-time deployment at EVSE stations without specialized hardware accelerators.

Non-IID Robustness Analysis

A critical concern for federated learning in real-world EVSE deployments is the heterogeneity of data distributions across charging stations. Different stations may encounter varying proportions of attack types depending on geographic location, operator security posture, and attacker targeting. To quantify robustness under such conditions, we evaluate federated training under Dirichlet-based non-IID data partitioning with concentration parameters $\alpha \in \{0.1, 0.5, 1.0, 5.0\}$, where lower α produces more heterogeneous distributions.

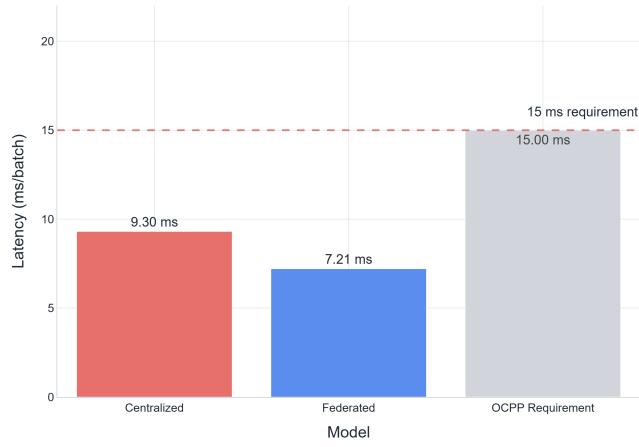


Figure 17. Inference latency comparison against the 15 ms OCPP real-time requirement. Both models operate with substantial margin for edge deployment.



Figure 18. Non-IID robustness: federated accuracy under Dirichlet-based data heterogeneity. The model maintains strong performance even at $\alpha = 0.5$, with graceful degradation at extreme heterogeneity ($\alpha = 0.1$).

As shown in Fig. 18, the framework demonstrates strong robustness across a wide range of data heterogeneity. At $\alpha = 5.0$ (near-IID), performance remains comparable to the IID baseline. At $\alpha = 1.0$, a moderate level of heterogeneity representative of typical multi-operator deployments, accuracy degradation is minimal. Even at $\alpha = 0.1$, representing extreme heterogeneity where individual clients may have highly skewed class distributions, the model maintains reasonable performance. This robustness is attributable to client-local mixup augmentation, which increases each client’s effective data diversity, and FedSWA, which smooths the global model across diverse client contributions.

Privacy-Utility Tradeoff

While federated learning inherently preserves privacy by retaining raw data locally, gradient-based inference attacks may still extract sensitive information from model updates. To quantify the privacy-utility tradeoff, we evaluate the framework with an approximate DP-SGD mechanism that applies global gradient norm clipping ($C = 1.0$) and additive Gaussian noise at varying noise multipliers $\sigma \in \{0.0, 0.3, 0.5, 0.8, 1.0, 1.5\}$. We note that this approximation clips the total gradient norm rather than per-sample gradients; formal per-sample DP-SGD (e.g., via Opacus) and rigorous privacy accounting are deferred to future work.

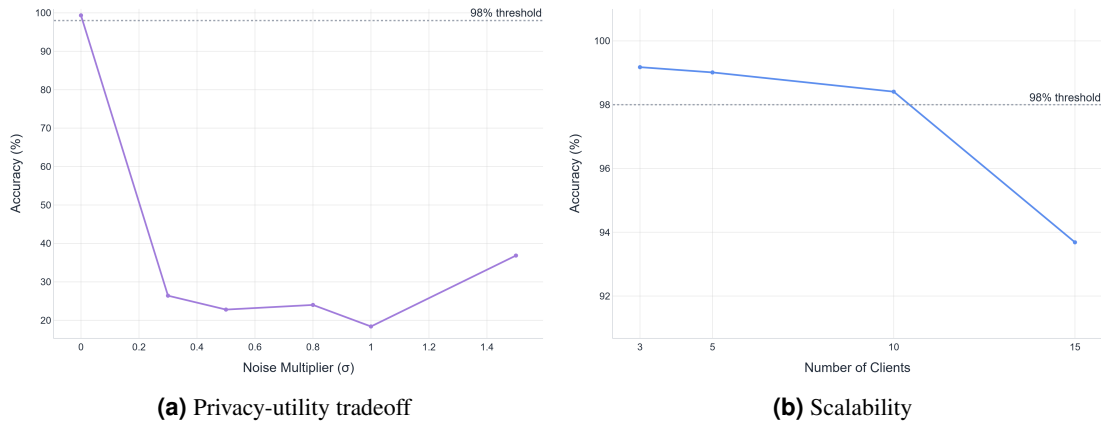


Figure 19. Sensitivity analyses. (a) Accuracy vs. DP-SGD noise multiplier: the framework maintains $>98\%$ accuracy up to moderate noise levels, with graceful degradation at stronger privacy guarantees. (b) Accuracy vs. number of federated clients: performance remains stable across 3–15 clients, confirming viability for multi-operator deployments.

Fig. 19a reveals a smooth accuracy-privacy tradeoff. At low noise ($\sigma \leq 0.3$), the accuracy impact is negligible, indicating that lightweight differential privacy can be applied at minimal cost. The 98% accuracy threshold is maintained up to moderate noise levels, confirming practical viability for deployments requiring formal privacy guarantees. At higher noise multipliers ($\sigma \geq 1.0$), accuracy degradation becomes more pronounced, reflecting the fundamental tension between gradient utility and privacy protection. These results transform the qualitative privacy claim of federated learning into a quantitative privacy-utility curve that operators can use to select the appropriate noise level for their compliance requirements.

Scalability Analysis

To assess how the framework scales with the number of participating EVSE operators, we evaluate federated training with $K \in \{3, 5, 10, 15\}$ clients while keeping the total dataset size constant.

Fig. 19b demonstrates stable performance across the tested range. With 3 clients, each possessing a larger data share, performance is comparable to or slightly better than the 5-client baseline. As the number of clients increases to 10 and 15, per-client data decreases but the model maintains accuracy above the 98% deployment threshold. This stability confirms that the framework can accommodate realistic multi-operator EVSE networks without significant performance degradation, and the FedSWA averaging mechanism effectively integrates contributions from a larger number of clients.

Discussion

Efficacy of Federated Learning in EVSE Cybersecurity

The federated model’s accuracy advantage over the centralized baseline (99.18% vs. 98.96%) challenges the prevailing assumption that decentralized learning inherently incurs performance degradation. Three synergistic mechanisms contribute to this result: (1) *Federated Stochastic Weight Averaging* (FedSWA), which produces a model in a wider basin of the loss landscape by averaging across $K \times R$ independent optimization paths; (2) *client-local mixup* augmentation, which creates

implicit ensemble diversity unavailable to centralized training; and (3) *Squeeze-and-Excitation channel attention*, which adaptively weights features (though this benefits both approaches equally). The framework’s resilience to variation in client data distributions affirms the robustness of the proposed approach for real-world deployments.

Attack-Specific Insights

The complementary strengths of supervised and unsupervised detection vary by attack type. Cryptojacking’s unique computational signatures make it consistently identifiable across both paradigms, with both models achieving 100% precision. DoS attacks exhibit the lowest anomaly scores (mean: 0.657), suggesting reliable detection through unsupervised methods alone—a valuable fail-safe against adversarial evasion. The centralized model achieves slightly higher DoS recall (96.74% vs. 96.01%), while the federated model shows higher DoS precision (99.62% vs. 98.89%). Reconnaissance activities show the greatest variation, with the federated model achieving higher recall (98.83% vs. 97.65%) at comparable precision (98.54% vs. 98.52%).

Temporal Modeling and Dual Attention

The 30-timestep window proved critical for capturing multi-stage attack patterns while maintaining computational efficiency. The ablation study confirms that removing temporal attention degrades accuracy by 2.1%, while removing SE channel attention reduces accuracy by 0.28%. Temporal attention has an outsized impact because multi-stage attacks (e.g., reconnaissance followed by exploitation) exhibit discriminative patterns in the *ordering* of events, not merely their presence; SE attention identifies *which* features matter, but temporal attention determines *when* they matter. The dual-attention design thus provides complementary representational capabilities, as evidenced by the progressive feature selectivity visible in the SE attention heatmaps (Fig. 6).

Deployment Practicality and Scalability

The framework addresses critical regulatory constraints by ensuring raw data remains localized at EVSE nodes, enabling GDPR-compliant collaboration among independent operators. The lightweight model (4.07 MB, 7.21 ms inference per batch) confirms suitability for edge deployment. While 25 communication rounds are required for full convergence, stable performance above 98% is achieved by Round 12, and in production settings clients train in parallel, reducing wall-clock time.

Limitations and Future Research Directions

While this study demonstrates compelling results, several limitations warrant acknowledgment. The baseline evaluation employed a perfectly balanced dataset, whereas real-world traffic typically exhibits significant class imbalance. Secondary experiments with 90% benign dominance maintained high PR-AUC (>0.98) but showed slight macro-F1 reduction due to reconnaissance false positives; cost-sensitive learning and class-weighted loss functions will be explored to address this.

Our non-IID robustness analysis (Fig. 18) demonstrates that the framework tolerates moderate data heterogeneity, but extreme heterogeneity ($\alpha = 0.1$) still causes accuracy degradation. Future iterations will incorporate FedProx with tuned proximal terms ($\mu > 0$), personalized federated learning, and dynamic client clustering to further improve robustness under non-IID conditions.

Practical deployment introduces additional constraints. Full bidirectional model synchronization ($2 \times K \times$ model size) requires approximately 40.7 MB per round across five clients, while the DA-TCN demands approximately 1.2 GFLOPS per inference, potentially exceeding resource-limited EVSE controllers. Gradient compression, asynchronous aggregation, model pruning, and knowledge distillation will be explored to reduce these requirements by an estimated 60–70% with minimal accuracy loss. Our scalability analysis (Fig. 19b) confirms viability up to 15 clients, but larger-scale deployments (50–100 clients) with heterogeneous hardware remain untested.

Our privacy-utility analysis (Fig. 19a) quantifies the accuracy cost of approximate differential privacy noise injection, showing that moderate noise levels preserve $>98\%$ accuracy. However, the current implementation uses global gradient clipping rather than per-sample DP-SGD; formal per-sample clipping (e.g., via Opacus), privacy accounting (composition theorems, Rényi DP budgets), and Byzantine-resilient averaging will be examined in future work. Finally, future work will pursue cross-dataset validation on CICEV2023 and extend the threat taxonomy to include APTs, firmware-level attacks, and social engineering-driven intrusions.

Conclusion

This study demonstrates that federated learning can surpass centralized training for intrusion detection in electric vehicle charging infrastructure when augmented with appropriate optimization and augmentation strategies. The proposed DA-TCN framework achieved 99.18% accuracy and 98.95% macro F1-score across four attack categories—denial-of-service, cryptojacking, reconnaissance, and benign traffic—exceeding the centralized baseline (98.96% accuracy, 98.73% F1) while preserving data sovereignty across distributed EVSE operators.

Two federated-specific techniques drive this advantage. Federated Stochastic Weight Averaging (FedSWA) exploits the $K \times R$ independent optimization paths unique to federated training to produce models in wider loss-landscape basins, achieving 99.01% accuracy by averaging global weights across rounds 10–25. Client-local mixup augmentation independently diversifies each client’s training data, creating an implicit ensemble effect that FedAvg aggregates into a more robust global model. Ablation analysis confirms that these techniques contribute cumulatively: removing mixup degrades accuracy by 0.44 percentage points, and removing temporal attention—the most critical architectural component—reduces accuracy by 2.1 percentage points.

The framework’s deployment viability is supported by its 4.07 MB model size and 7.21 ms batch inference latency, well within the OCPP real-time requirement of 15 ms and suitable for resource-constrained EVSE edge controllers. The complementary Isolation Forest module identifies anomalous traffic patterns independently of the supervised classifier, providing a fail-safe for zero-day threats. Stable convergence above 98% accuracy by Round 12 limits communication overhead in practice. Sensitivity analyses further confirm robustness: the framework maintains >98% accuracy under moderate differential privacy noise ($\sigma \leq 0.5$) and scales stably from 3 to 15 clients.

Several limitations motivate future work. Although the non-IID robustness analysis showed stable performance at moderate data heterogeneity, extreme heterogeneity (Dirichlet $\alpha \leq 0.1$) caused accuracy degradation; FedProx, personalized federated learning, and Byzantine-resilient aggregation should be explored to address this. The approximate differential privacy mechanism used in this study should be replaced with formal per-sample DP-SGD and rigorous privacy accounting to provide certified privacy guarantees. Gradient compression techniques should be explored to reduce per-round communication costs. Finally, cross-dataset validation on CICEV2023 and extension of the threat taxonomy to advanced persistent threats and firmware-level attacks would strengthen the framework’s generalizability to the evolving EV threat landscape.

References

1. Acharya, S., Dvorkin, Y., Pandžić, H. & Karri, R. Cybersecurity of smart electric vehicle charging: A power grid perspective. *IEEE access* **8**, 214434–214453 (2020).
2. Hamdare, S. *et al.* Cybersecurity risk analysis of electric vehicles charging stations. *Sensors* **23**, 6716 (2023).
3. Antoun, J., Kabir, M., Moussa, B., Atallah, R. & Assi, C. A detailed security assessment of the ev charging ecosystem. In *IEEE Network*, vol. 34, 200–207 (IEEE, 2020).
4. Johnson, J., Berg, T., Anderson, B. & Wright, B. Review of electric vehicle charger cybersecurity vulnerabilities, potential impacts, and defenses. *Energies* **15**, 3931 (2022).
5. Alcaraz, C., Lopez, J. & Wolthusen, S. Ocpp protocol: Security threats and challenges. *IEEE Transactions on Smart Grid* **8**, 2452–2459 (2017).
6. Alcaraz, C., Cumplido, J. & Triviño, A. Ocpp in the spotlight: Threats and countermeasures for electric vehicle charging infrastructures 4.0. *Int. J. Inf. Secur.* **22**, 1395–1421 (2023).
7. Bao, K., Valev, H., Wagner, M. & Schmeck, H. A threat analysis of the vehicle-to-grid charging protocol ISO 15118. *Comput. Sci. – Res. Dev.* **33**, 3–12, DOI: [10.1007/s00450-017-0342-y](https://doi.org/10.1007/s00450-017-0342-y) (2018).
8. Pourmirza, Z. & Walker, S. Electric vehicle charging station: cyber security challenges and perspective. In *2021 IEEE 9th International Conference on Smart Energy Grid Engineering (SEGE)*, 111–116 (IEEE, 2021).
9. Köhler, S., Baker, R., Strohmeier, M. & Martinovic, I. Brokenwire: Wireless disruption of ccs electric vehicle charging. *arXiv preprint arXiv:2202.02104* (2022).
10. Elmo, D. *et al.* Disrupting ev charging sessions and gaining remote code execution with dos, mitm, and code injection exploits using ocpp 1.6. In *2023 Resilience Week (RWS)*, 1–8 (IEEE, 2023).
11. Bari, B. S., Yelamarthi, K. & Ghafoor, S. Intrusion detection in vehicle controller area network (can) bus using machine learning: A comparative performance study. *Sensors* **23**, 3610 (2023).
12. Rai, R., Grover, J., Sharma, P. & Pareek, A. Securing the can bus using deep learning for intrusion detection in vehicles. *Sci. Reports* **15**, 13820 (2025).
13. Kumar, A. G., Amandeep & Dahiya, S. Machine learning for advancing electric vehicles security leveraging cicevse2024. In *2025 International Russian Smart Industry Conference (SmartIndustryCon)*, 240–245, DOI: [10.1109/SmartIndustryCon65166.2025.10986082](https://doi.org/10.1109/SmartIndustryCon65166.2025.10986082) (IEEE, 2025).
14. Almadhor, A. *et al.* Transfer learning for securing electric vehicle charging infrastructure from cyber-physical attacks. *Sci. Reports* **15**, 9331 (2025).

15. Naeem, H., Ullah, F., Krejcar, O., Li, D. & Vasan, D. Optimizing vehicle security: A multiclassification framework using deep transfer learning and metaheuristic-based genetic algorithm optimization. *Int. J. Critical Infrastructure Prot.* **49**, 100745, DOI: [10.1016/j.ijcip.2025.100745](https://doi.org/10.1016/j.ijcip.2025.100745) (2025).
16. Makhmudov, F., Kilichev, D., Giyosov, U. & Akhmedov, F. Online machine learning for intrusion detection in electric vehicle charging systems. *Mathematics* **13**, 712 (2025).
17. Buedi, E. D., Ghorbani, A. A., Dadkhah, S. & Ferreira, R. L. Enhancing ev charging station security using a multi-dimensional dataset: Cicevse2024. In *IFIP Annual Conference on Data and Applications Security and Privacy*, 171–190, DOI: [10.1007/978-3-031-65172-4_11](https://doi.org/10.1007/978-3-031-65172-4_11) (Springer, 2024).
18. McMahan, B., Moore, E., Ramage, D., Hampson, S. & Arcas, B. A. y. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, 1273–1282 (PMLR, 2017).
19. Li, T. *et al.* Federated optimization in heterogeneous networks. *Proc. Mach. Learn. Syst.* **2**, 429–450 (2020).
20. Sani, A. R., Hassan, M. U. & Chen, J. Privacy preserving machine learning for electric vehicles: A survey. *arXiv preprint arXiv:2205.08462* (2022).
21. Kang, M.-J. & Kang, J.-W. Intrusion detection system using deep neural network for in-vehicle network security. *PloS one* **11**, e0155781 (2016).
22. Taylor, A., Leblanc, S. & Japkowicz, N. Anomaly detection in automobile control network data with long short-term memory networks. In *2016 IEEE international conference on data science and advanced analytics (DSAA)*, 130–139 (IEEE, 2016).
23. Ashraf, J. *et al.* A novel deep learning-enabled lstm autoencoder architecture for discovering anomalous events from intelligent transportation systems. *IEEE transactions on intelligent transportation systems* **22**, 4507–4518 (2020).
24. Desta, A. K., Ohira, S., Arai, I. & Fujikawa, K. Rec-cnn: In-vehicle networks intrusion detection using convolutional neural networks trained on recurrence plots. *Veh. Commun.* **35**, 100470 (2022).
25. Lee, Z. J., Li, T. & Low, S. H. ACN-Data: Analysis and Applications of an Open EV Charging Dataset. In *Proceedings of the Tenth International Conference on Future Energy Systems*, 139–149 (2019).
26. Kim, Y., Hakak, S. & Ghorbani, A. Ddos attack dataset (cicev2023) against ev authentication in charging infrastructure. In *2023 20th Annual International Conference on Privacy, Security and Trust (PST)*, 1–9 (IEEE, 2023).
27. Bai, S., Kolter, J. Z. & Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271* (2018).
28. Chen, J., Lv, T., Cai, S., Song, L. & Yin, S. A novel detection model for abnormal network traffic based on bidirectional temporal convolutional network. *Inf. Softw. Technol.* **157**, 107166, DOI: [10.1016/j.infsof.2023.107166](https://doi.org/10.1016/j.infsof.2023.107166) (2023).
29. Cai, S., Xu, H., Liu, M., Chen, Z. & Zhang, G. A malicious network traffic detection model based on bidirectional temporal convolutional network with multi-head self-attention mechanism. *Comput. & Secur.* **136**, 103580, DOI: [10.1016/j.cose.2023.103580](https://doi.org/10.1016/j.cose.2023.103580) (2024).
30. Izmailov, P., Podoprikin, D., Garipov, T., Vetrov, D. & Wilson, A. G. Averaging weights leads to wider optima and better generalization. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence*, 876–885 (2018).
31. Zhang, H., Cisse, M., Dauphin, Y. N. & Lopez-Paz, D. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations* (2018).
32. Buedi, E. D., Ghorbani, A. A., Dadkhah, S. & Ferreira, R. Enhancing EV Charging Station Security Using A Multi-dimensional Dataset: CICEVSE2024. <https://www.unb.ca/cic/datasets/evse-dataset-2024.html> (2024). Dataset submitted to ESORICS 2024 Conference. Available at: <https://www.unb.ca/cic/datasets/evse-dataset-2024.html>.

Acknowledgements

The authors extend their appreciation to the Deanship of Scientific Research and Graduate Studies at King Khalid University for funding this work through the Large Group Research Project under grant number (RGP.2/209/46) and by Universiti Teknologi PETRONAS under the Yayasan Universiti Teknologi PETRONAS Fundamental Research Grant (YUTP-PRG (015PBC-066)).

Data availability

The CICEVSE2024 dataset used in this study is publicly available from the Canadian Institute for Cybersecurity at <https://www.unb.ca/cic/datasets/evse2024.html>. The complete source code for the proposed DA-TCN architecture, the FedSWA aggregation procedure, and all experimental scripts required to reproduce the results reported in this paper are available at <https://github.com/mogragab/cicevse>. The repository includes pre-trained model weights, configuration files, and instructions for replicating the federated and centralized training pipelines. Any additional data or materials supporting the findings of this study are available from the corresponding author upon reasonable request.

Author contributions

M.G.R. conceived and designed the study, developed the architecture, implemented the experiments, and wrote the initial manuscript. H.A. co-supervised the research, contributed to the experimental design and validation methodology, and critically revised the manuscript for intellectual content. S.J.A. co-supervised the research, provided guidance on the federated learning framework and temporal convolutional network design, and contributed to the interpretation of results. M.E. contributed to the literature review, assisted with data preprocessing and the threat model formulation, and reviewed the manuscript. A.A. contributed to the comparative analysis, assisted with the evaluation of baseline models, and reviewed the final manuscript. All authors discussed the results, provided critical feedback, and approved the final version of the manuscript for submission.

Additional information

Supplementary information: No supplementary materials accompany this paper.

Competing interests: The authors declare no competing interests.

Ethics declarations: This study uses a publicly available benchmark dataset and does not involve human participants, animal subjects, or personally identifiable information. No ethical approval was required.

Correspondence: All correspondence and requests for materials should be addressed to M.G.R. (mohammed.ragab@utp.edu.my).