IEEE Open Journal of the Computer Society

Received 8 June 2025; revised 15 September 2025; accepted 22 September 2025. Date of publication 10 October 2025; date of current version 13 November 2025. The review of this article was arranged by Associate Editor Laizhong Cui.

Digital Object Identifier 10.1109/OJCS.2025.3616394

FedWGCA: A Federated Learning Based AAV Intrusion Detection With Gradient Clipping and Attention-Based Neural Networks

MD. FAHIM-UL-ISLAM [1], AMITABHA CHAKRABARTY [1], HALIMATON SAADIAH HAKIMI [2], AND SITI SARAH MAIDIN [3]

[1]Department of Computer Science and Engineering, BRAC University, Dhaka 1212, Bangladesh
[2]Computer and Information Sciences Department, University Teknologi PETRONAS, Seri Iskandar 32610, Malaysia
[3]Faculty of Data Science and Information Technology, INTI International University, Nilai 71800, Malaysia

CORRESPONDING AUTHOR: AMITABHA CHAKRABARTY (e-mail: amitabha@bracu.ac.bd).

ABSTRACT Autonomous Aerial Vehicles (AAVs) are progressively employed in various applications, including surveillance and logistics. However, their rising usage is coupled by increased cybersecurity threats. Conventional intrusion detection systems (IDS) frequently inadequately address specific problems presented by AAV networks, including dynamic operational environments, varied data distributions, and severe resource constraints. Federated Learning (FL) has emerged as a viable alternative, offering a decentralized approach to collaborative training of intrusion detection models while preserving data privacy. However, FL is not without its shortcomings, including poisoning and backdoor attacks, which can impair the accuracy and reliability of the models, thereby exposing AAVs to advanced cyber threats. This research attempts to design resilient FL-based frameworks that address these drawbacks, boosting the security and resilience of AAV networks in the face of emerging cyber hazards. We present our proposed weighted gradient clipping aggregation (FedWGCA) framework to mitigate the impact of malicious updates during the model training process. Experimental studies show that our FedWGCA outperforms state-of-the-art methods, surpassing FedAvg, FedNova, FedOpt, and FedProx by up to 7.23% in accuracy, 9.99% in precision, and 10.14% in recall. For enhancing resilience in our FL architecture, we further present our robust attention neural network, ANET, which outperforms XGBoost by 0.90% and DNN by 7.10%, showcasing its superior precision and reduced false positives in local training.

INDEX TERMS Attention neural network, cybersecurity threats, intrusion detection systems (IDS), federated learning (FL), R&D investment, autonomous aerial vehicles (AAVs).

I. INTRODUCTION

Smart cities harness technologies such as IoT, AI, and Big Data to optimize urban efficiency and sustainability [1], [2]. Autonomous Aerial Vehicles (AAVs) play a pivotal role, enabling applications like traffic analysis, infrastructure inspection, and medical deliveries [3]. The integration of IoT with AAVs further supports advanced functionalities, including air quality monitoring, autonomous delivery, and disaster response. However, the growing dependence on AAV networks introduces significant cybersecurity risks. Adversaries can exploit vulnerabilities through attacks such as Denial-of-Service (DoS), replay, Evil Twin, and False Data Injection (FDI), which threaten data integrity, disrupt operations, and jeopardize safety-critical missions [4], as depicted in Fig. 1.

To counter these threats, robust Intrusion Detection Systems (IDS) are vital for monitoring network traffic and device activity in real time to detect anomalies and mitigate risks [5], [6]. Unlike traditional IoT setups, IoT-AAV networks face unique challenges due to lightweight communication protocols, resource-constrained onboard sensors, and dynamic topologies driven by AAV mobility [7], [8]. These factors increase vulnerability to sophisticated attacks, such as model poisoning and adversarial manipulation, necessitating adaptable, multi-layered IDS capable of analyzing diverse data

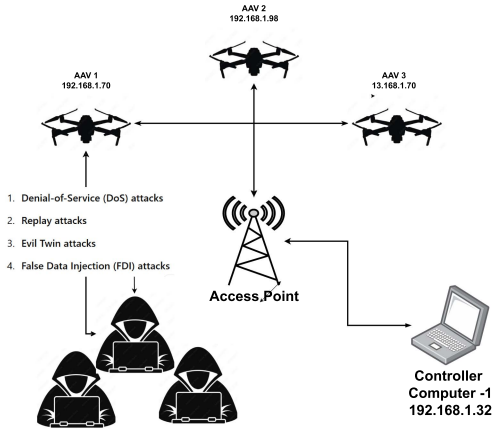

FIGURE 1. Network diagram of AAVs (AAV1, AAV2, AAV3) connected to an access point and controller, illustrating IP addresses and cyber threats like DoS, replay, Evil Twin, and FDI attacks.

sources, from network traffic to sensor signals [9]. However, conventional lightweight IDS solutions often lack the flexibility to address these advanced threats effectively [10], [11].

Federated Learning (FL) offers a promising approach for AAV-IDS deployment by enabling distributed model training across multiple AAVs without sharing raw data [12]. This is critical because AAVs collect sensitive information, such as flight paths, mission objectives, and environmental data, which cannot be centralized due to privacy concerns and regulatory restrictions, such as GDPR or aviation-specific data protection laws. By training models locally and sharing only model updates, FL preserves data privacy while leveraging diverse local observations to enhance detection accuracy. However, FL is susceptible to adversarial attacks, where malicious clients may inject poisoned or backdoored updates, undermining the global model's performance. Robust aggregation strategies are thus essential to ensure reliable federated training.

Implementing Deep Neural Networks (DNNs) as local classifiers in FL-based IDS presents further challenges. While DNNs excel in detecting complex intrusion patterns, they impose significant computational and latency demands, particularly in resource-constrained AAV environments. Hardware accelerators like GPUs, TPUs, and FPGAs can mitigate these issues, but mapping DNN layers to these accelerators is complex due to differences in memory hierarchies, parallelism capabilities, and hardware-specific optimization requirements [13]. Inefficient mapping can lead to suboptimal resource utilization, increased latency, or excessive power consumption, which are critical concerns for AAVs with limited onboard resources. Additionally, pre-trained DNN models, while effective in general scenarios, often struggle in AAV-specific contexts due to distribution shifts caused by dynamic factors like varying altitudes, urban vs. rural environments, or localized attack signatures [10]. For instance, replay or jamming attacks may exhibit unique patterns depending on AAV mobility or signal quality, requiring online training to adapt models to these evolving conditions. Such adaptation reduces false positives and negatives, significantly enhancing mission safety and network resilience.

To address these challenges, we propose a robust FL framework for AAV intrusion detection that integrates advanced aggregation mechanisms and attention-based neural architectures. Our contributions are as follows:

- An **attention-driven neural classifier (ANET)** with multi-head attention and squeeze-and-excitation blocks to prioritize critical intrusion features and enhance detection robustness.
- A novel **Weighted Gradient Clipping Aggregation (FedWGCA)** method to mitigate malicious or noisy updates, improving stability in federated training.
- A **latency-aware accelerator mapping strategy** to optimize DNN component assignment to GPUs, TPUs, and FPGAs, enhancing efficiency in resource-constrained AAV environments.
- A **Neural Architecture Search (NAS)-based optimization framework** to adapt local classifiers to heterogeneous AAV hardware while maintaining high detection accuracy.
- Comprehensive evaluations across multiple public datasets, comparing our framework with state-of-the-art FL aggregation methods and traditional IDS approaches.

By addressing both security and system-level efficiency, our work contributes to the development of a scalable, privacy-preserving, and resilient AAV-IDS framework suitable for real-world smart city deployments.

II. RELATED WORK

Federated Learning (FL) enables multiple clients to collaboratively train a global model (GM) without sharing private data, using a central server and N clients. However, poisoning attacks, where malicious clients upload manipulated local models (LMs), threaten GM integrity [13]. Defense mechanisms include distance-based, performance-based, statistical, and machine learning-based approaches, each with limitations in false positives, scalability, and adaptability [10].

A. DISTANCE-BASED DEFENSE METHODS

Distance-based methods detect poisoned LMs by measuring divergence from benign models [14]. KRUM uses pairwise Euclidean distances to select reliable LMs [15], while MULTI-KRUM enhances robustness by aggregating multiple LMs [21]. FoolsGold identifies Sybil attacks via update pattern similarity [22].

B. PERFORMANCE-BASED DEFENSE METHODS

Performance-Weighted Aggregation methods include FedPA, which balances training time and accuracy with dynamic partial model aggregation [23]. FedHQ optimizes convergence via client-specific weights based on quantization errors [24]. MHAT uses Knowledge Distillation for heterogeneous model aggregation, reducing computational needs [25].

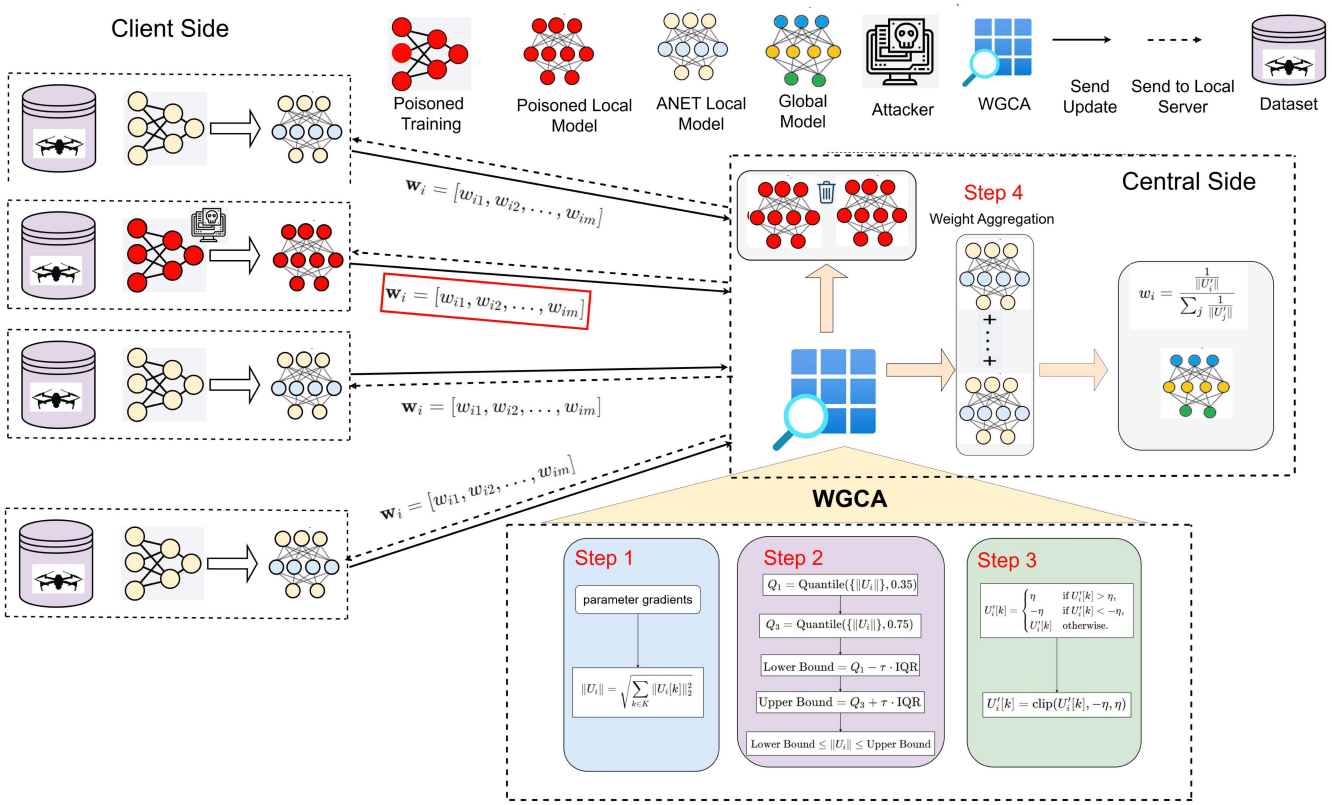

FIGURE 2. Schematic of the FedWGCA defense method, where the central server collects LMs, applies FedWGCA in three steps to filter poisoned models, and aggregates benign LMs to update the GM.

C. STATISTICAL AND ROBUST AGGREGATION METHODS

Statistical methods detect anomalies in model updates or use robust aggregation to mitigate poisoned LMs [33]. Median-based aggregation replaces weighted averages with medians, resilient to outliers [34]. Trimmed Mean discards extreme updates to reduce malicious impact [35].

D. MACHINE LEARNING-BASED DEFENSE METHODS

Machine learning-based methods learn benign and malicious patterns [41]. Lightweight IDS for AAV networks use fuzzy rough sets and shallow deep learning for high-precision detection [42]. Other approaches include a AAV-satellite 5G security model, an RNN-based FANET framework for real-time anomaly detection [43], and a collaborative IDS (AAV-CIDS) with FFCNN for zero-day attack detection [44].

E. LIMITATIONS OF EXISTING METHODS

Current FL defenses face challenges: KRUM methods have high false positive rates, reducing GM accuracy. Static thresholds limit adaptability to dynamic FL environments and evolving attacks. Scalability issues hinder complex methods like deep learning-based or robust aggregation in large-scale FL systems.

III. PROPOSED METHODOLOGY

In Federated Learning (FL), the global model $\mathcal{M}$ is iteratively updated by aggregating local model updates $\{\Delta\mathcal{M}_i\}_{i=1}^{N}$ from N clients. However, these local updates may include noise or adversarial gradients due to non-IID data distributions or the presence of malicious clients. To address these challenges, we propose a novel aggregation strategy called **Weighted Gradient Clipping Aggregation** (FedWGCA), which integrates gradient clipping with an inverse norm weighting scheme to improve the robustness and stability of the global model. The overall schematic of the proposed FedWGCA defense method is shown in Fig. 3.

The core component of FedWGCA is its inverse norm weighting strategy, where each client's contribution is weighted inversely to the L_2-norm of their clipped gradient update. This is defined as:

$$w_i = \frac{1/\|\Delta\mathcal{M}_i'\|}{\sum_{j=1}^{N} 1/\|\Delta\mathcal{M}_j'\|} \tag{1}$$

Here, $\Delta\mathcal{M}_i'$ denotes the clipped version of the local update $\Delta\mathcal{M}_i$. This formulation assigns smaller weights to updates with larger magnitudes, based on the observation that reliable and generalizable gradients often have smaller norms. Conversely, larger gradient norms may indicate overfitting, noisy

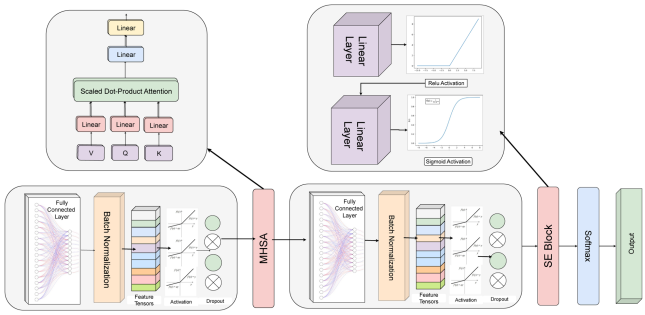

FIGURE 3. Architecture of the proposed ANET model. The model integrates attention mechanisms and squeeze-and-excitation (SE) blocks into a neural network designed for federated learning (FL) environments.

behavior, or adversarial manipulation such as model poisoning. As such, this approach mitigates their influence during aggregation, resembling robust techniques like the Trimmed Mean which discard extreme values to enhance robustness.

Additionally, gradient clipping is employed to bound each local update within a specified range:

$$\Delta \mathcal{M}_i' = \text{clip}(\Delta \mathcal{M}_i, -C, C) \tag{2}$$

This ensures that no individual client can disproportionately influence the global model, enhancing the method's defense against adversarial behaviors. Fig. 2 shows theproposed architecture.

IV. PROBLEM FORMULATION

We consider a federated learning (FL) system with N (where $N > 1$) participating clients, each having its local dataset $D_i = \{(x_{i,j}, y_{i,j})\}_{j=1}^{n_i}$, where $x_{i,j} \in \mathcal{X}$ represents the input space, and $y_{i,j} \in \mathcal{Y}$ denotes the corresponding label. The input space $\mathcal{X}$ may include data from heterogeneous sources, such as images, text, or sensor readings, with each client possessing a unique data distribution $P_i(\mathcal{X}, \mathcal{Y})$. Our goal is to train a global model $f : \mathcal{X} \to \mathcal{Y}$ that generalizes well across the distributions of all participating clients while being robust to adversarial attacks and noisy updates from malicious or faulty clients.

The key challenge arises from the heterogeneity of the data distributions across clients, i.e., $P_i(\mathcal{X}, \mathcal{Y}) \neq P_j(\mathcal{X}, \mathcal{Y})$ for $i \neq j$, and the presence of adversarial updates in the system, which may skew the global model training. Furthermore, communication constraints necessitate efficient aggregation mechanisms to combine the client updates while mitigating the impact of outliers.

V. PROPOSED MODEL ANET AS LOCAL CLASSIFIER

We have utilized ANET to integrate attention mechanisms and squeeze-and-excitation (SE) blocks into a neural network designed for federated learning (FL) environments. This architecture is tailored in local updates for robust local classification on heterogeneous data distributions. Each component is detailed below with mathematical formulations. Fig. 3 depicts the internal component of the model architecture.

A. INPUT LAYER

The model takes as input a feature vector $\mathbf{x} \in \mathbb{R}^d$, where d denotes the dimensionality of the feature space, and this vector encapsulates the raw data, which is directly fed into the first dense layer of the model without any prior modifications. The feature vector $\mathbf{x}$ is composed of individual components represented as $\mathbf{x} = [x_1, x_2, \ldots, x_d]$, where each component corresponds to a specific attribute or measurement in the data.

B. DENSE LAYER 1

The first dense layer applies a linear transformation to project the input into a higher-dimensional space, expressed as $\mathbf{h}_1 = \text{Dropout}(\sigma_1(\text{BatchNorm}(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1)))$, where $\mathbf{W}_1 \in \mathbb{R}^{d\times h}$ and $\mathbf{b}_1 \in \mathbb{R}^h$ are the weights and biases of the layer, BatchNorm$(\cdot)$ normalizes the layer outputs to stabilize training, $\sigma_1(\cdot)$ is the Leaky ReLU activation function defined as $\sigma_1(x) = \begin{cases} x, & \text{if } x \geq 0, \\ \alpha x, & \text{if } x < 0, \end{cases}$ with α as a small positive constant, and Dropout$(\cdot)$ randomly sets a fraction of the activations to zero during training to reduce overfitting.

C. ATTENTION MECHANISM

We utilize the attention mechanism [46] that is designed to focus on the most relevant features, enhancing the model's adaptability to diverse datasets. It employs multi-head attention, where each head processes a subset of features.

a) Query, Key, and Value Transformations: The input $\mathbf{h}_1$ is transformed into query ($\mathbf{Q}$), key ($\mathbf{K}$), and value ($\mathbf{V}$) matrices:

$$\mathbf{Q} = \mathbf{h}_1\mathbf{W}_Q, \quad \mathbf{K} = \mathbf{h}_1\mathbf{W}_K, \quad \mathbf{V} = \mathbf{h}_1\mathbf{W}_V, \tag{3}$$

where $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \in \mathbb{R}^{h\times h}$ are learnable weights.

b) Scaled Dot-Product Attention: The attention output is computed as:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{h}}\right)\mathbf{V}. \tag{4}$$

Here, h represents the hidden dimensionality, and scaling by $\sqrt{h}$ ensures numerical stability.

c) Multi-Head Attention: The outputs from multiple heads are concatenated:

$$\mathbf{h}_{\text{attn}} = \text{Concat}(\text{head}_1, \text{head}_2, \ldots, \text{head}_{n_{\text{heads}}}), \tag{5}$$

where head_i denotes the attention output for the i^{th} head.

D. SQUEEZE-AND-EXCITATION (SE) BLOCK

We further use the Squeeze-and-Excitation (SE) block [47] in our ANET model. It is a computational unit designed to enhance the representational power of a network by explicitly modeling channel-wise dependencies. It introduces a mechanism to adaptively recalibrate feature responses by emphasizing informative features and suppressing less useful

Algorithm 1: Weighted Gradient Clipping Aggregation (FedWGCA).

Require: Global model M, client updates $\{U_1, U_2, \ldots, U_n\}$, clipping threshold λ
Ensure: Updated global model M
for each update U_i **do**
  Compute norm of the update: $\|U_i\| = \sqrt{\sum_p \|U_i[p]\|^2}$
end for
Compute weights: $w_i = \frac{1}{\|U_i\|}, \quad w_i = \frac{w_i}{\sum w_i}$
for each parameter p **do**
  Clip updates: $U_i^{\text{clipped}}[p] = \text{clip}(U_i[p], -\lambda, \lambda)$
  Aggregate clipped updates: $M[p] = \sum_i w_i \cdot U_i^{\text{clipped}}[p]$
end for
Update global model M

ones. This is achieved through a two-step process, which is defined as squeeze and excitation.

E. OUTPUT LAYER

The output layer is responsible for producing class probabilities, enabling the model to make predictions. It applies a softmax function to the transformed features obtained from the previous layer. Mathematically, the output $\mathbf{y} \in \mathbb{R}^c$ is computed as:

$$\mathbf{y} = \text{softmax}(\mathbf{W}_3 \mathbf{h}_{\text{SE}} + \mathbf{b}_3), \tag{6}$$

where $\mathbf{W}_3 \in \mathbb{R}^{h \times c}$ and $\mathbf{b}_3 \in \mathbb{R}^c$ are the weight matrix and bias vector of the output layer, respectively, and c is the number of classes. The softmax function ensures that the output values are normalized to form a valid probability distribution.

F. INTEGRATION IN FEDERATED LEARNING

In federated learning, our ANET model is trained in a decentralized manner across multiple client devices, each holding its local dataset. During local training, each client minimizes the cross-entropy loss, which measures the discrepancy between the predicted class probabilities and the true labels. The cross-entropy loss $\mathcal{L}$ is defined as:

$$\mathcal{L} = -\sum_{i=1}^{c} y_i \log(\hat{y}_i), \tag{7}$$

where y_i is the true label (one-hot encoded) for class i, and $\hat{y}_i$ is the predicted probability for class i. This loss function encourages the model to assign high probabilities to the correct classes.

After local training, the global model is updated by aggregating the model parameters from all participating clients.

G. WEIGHTED GRADIENT CLIPPING AGGREGATION (FEDWGCA) WITH OUTLIER FILTERING

In federated learning, the global model $\mathcal{M}$ is updated by aggregating updates $\{\Delta\mathcal{M}_i\}_{i=1}^N$ received from N client models. To ensure robustness and stability during aggregation, our FedWGCA method is used in the FL architecture, which incorporates outlier filtering and gradient clipping. The process consists of the following steps:

1) COMPUTATION OF NORM OF CLIENT UPDATES

Each client update $\Delta\mathcal{M}_i$ consists of gradients for all model parameters. To quantify the magnitude of each update, the L_2-norm of the update is computed. For the i-th client, the norm is calculated as:

$$\|\Delta\mathcal{M}_i\| = \sqrt{\sum_{j=1}^{d} \left(\|\Delta\mathcal{M}_{i,j}\|^2\right)}, \tag{8}$$

where d is the total number of parameters in the model, and $\|\Delta\mathcal{M}_{i,j}\|$ is the norm of the j-th parameter in the update from client i. This norm serves as a measure of the overall magnitude of the update.

2) OUTLIER FILTERING

To mitigate the impact of potentially harmful updates, outlier filtering is applied based on the norms of the client updates. The first quartile (Q_1) and third quartile (Q_3) of the update norms are computed, and the interquartile range (IQR) is determined as:

$$\text{IQR} = Q_3 - Q_1. \tag{9}$$

Updates whose norms fall outside the range:

$$Q_1 - \lambda \times \text{IQR} \leq \|\Delta\mathcal{M}_i\| \leq Q_3 + \lambda \times \text{IQR} \tag{10}$$

are considered outliers and discarded. Here, λ is a tunable threshold parameter that controls the sensitivity of outlier detection. This step ensures that only reliable updates are included in the aggregation process.

3) WEIGHTED AGGREGATION OF THE FILTERED UPDATES

After filtering outliers, the remaining updates are aggregated using a weighted approach. The weight w_i for each client update $\Delta\mathcal{M}_i$ is inversely proportional to its norm:

$$w_i = \frac{1}{\|\Delta\mathcal{M}_i\|}. \tag{11}$$

To ensure that the weights sum to one, they are normalized as:

$$w_i' = \frac{w_i}{\sum_{i=1}^{N'} w_i}, \tag{12}$$

where N' is the number of non-outlier clients. The global model update $\Delta\mathcal{M}$ is then computed as the weighted sum of the filtered updates:

$$\Delta\mathcal{M} = \sum_{i=1}^{N'} w_i' \cdot \Delta\mathcal{M}_i. \tag{13}$$

This weighting scheme assigns higher importance to updates with smaller magnitudes, which are typically more stable and reliable.

4) GRADIENT CLIPPING

To further improve training stability and robustness against outliers or adversarial client updates, **gradient clipping** is applied to each client update before aggregation. Specifically, the clipped gradient $\Delta\mathcal{M}_i'$ is computed as:

$$\Delta\mathcal{M}_i' = \text{clip}(\Delta\mathcal{M}_i, -C, C), \tag{14}$$

where C is a predefined clipping threshold. The function $\text{clip}(x, -C, C)$ constrains the elements of x within the range $[-C, C]$, thereby preventing any single client's update from disproportionately influencing the global model. This is especially critical in federated learning environments characterized by non-IID data and unreliable or adversarial clients.

5) THEORETICAL INTUITION AND MODEL UPDATE

After clipping, an **inverse norm weighting** strategy is employed during aggregation. The key idea is to reduce the influence of clients whose updates have excessively large norms, which may indicate noisy gradients, overfitting, or adversarial behavior. Let $\|\Delta\mathcal{M}_i'\|$ denote the norm of the clipped update from client i. The aggregation weight for client i is defined as:

$$w_i = \frac{1/\|\Delta\mathcal{M}_i'\|}{\sum_{j=1}^{N} 1/\|\Delta\mathcal{M}_j'\|}, \tag{15}$$

where N is the number of participating clients. The aggregated update is then given by:

$$\Delta\mathcal{M} = \sum_{i=1}^{N} w_i \cdot \Delta\mathcal{M}_i'. \tag{16}$$

Finally, the global model is updated as follows:

$$\mathcal{M}_{t+1} = \mathcal{M}_t + \eta \cdot \Delta\mathcal{M}, \tag{17}$$

where η is the global learning rate, and $\mathcal{M}_t$ is the global model at round t. This update rule facilitates stable convergence and improves robustness against client heterogeneity and adversarial attacks. Algorithm 1 illustrates the overall process, including the filtering and aggregation steps.

VI. EXPERIMENTAL RESULT

A. EXPERIMENTAL SETUP

For global testing with our FL model with our existing works, we compare with FedAvg [36]. FedOpt [37] FedNova [38], FedProx [39] and Multikrum [40] model. Besides, for the local testing of our ANET model, we compare it to the SOTA models, such as DNN [53] and XGBoost [54]. We evaluate our model on the GPU A100, GTX 1650, and RTX 4080 used as accelerators to assess its performance across diverse hardware configurations. We have used the cyber-physical dataset for AAVs under normal operations and cyberattacks, including de-authentication DoS, replay, false data injection, and evil twin attacks [45], [55]. The dataset consists of one CSV file containing cyber (37 features) and physical (16 features) data. The testbed includes a AAV, controller, and data collection tools. For benchmark testing and the reliability of our model framework, we utilize the ECU-IoFT dataset [19] to analyze cyberattacks on Wi-Fi communications in low-end AAVs, focusing on developing defense strategies. Additionally, we leverage the WUSTL-EHMS-2020 dataset [20] to examine man-in-the-middle attacks on healthcare monitoring systems, combining network flow and biometric data for intrusion detection. Data preprocessing includes label encoding for categorical features, standardization using StandardScaler, and an 80-20 train-test split.

TABLE 1. Performance Metrics for ANET, DNN, XGBoost Models in the Local Training

Model	Scenario	Acc.	Prec.	Rec.
XGBoost	Local (Final)	96.21%	95.23%	94.13%
	Local (Best)	96.41%	94.27%	95.16%
	Global (Final)	95.38%	96.11%	92.90%
	Global (Best)	94.30%	92.13%	92.21%
DNN	Local (Final)	88.16%	89.01%	87.12%
	Local (Best)	89.18%	89.03%	89.31%
	Global (Final)	86.61%	86.69%	86.54%
	Global (Best)	86.07%	86.71%	86.61%
ANET	Local (Final)	95.58%	94.61%	94.10%
	Local (Best)	95.35%	96.13%	94.61%
	Global (Final)	95.41%	95.07%	93.14%
	Global (Best)	95.42%	94.05%	94.31%

B. COMPARISON OF LOCAL CLASSIFIER MODELS

We assess the performance of XGBoost, DNN, and ANET under local and global training scenarios. Table 1 summarizes the results of local training of the machine learning models. XGBoost performs best in the Local (Best) setup, achieving 96.41% accuracy, slightly higher than the Local (Final) at 96.21%. In global training, the Global (Final) setup outperforms the Global (Best) in accuracy (95.38%) and precision (96.11%), though recall drops to 92.90%. DNN shows lower performance than XGBoost and ANET. Its Local (Best) setup achieves 89.18% accuracy, with minor improvements over the Local (Final). In global training, performance declines, with Global (Final)accuracy at 86.61% and Global (Best) at 86.07%, indicating its limitations in global setups. Therefore, ANET excels in local training, with Local (Final) and Local (Best) achieving 95.58% and 95.35% accuracy, respectively. The Local (Best) setup records the highest precision (96.13%) and recall (94.61%). In global training, performance remains stable, with both configurations achieving around 95.42% accuracy. Overall, XGBoost dominates local training, ANET remains robust across both setups, and DNN struggles in global scenarios, requiring further optimization.

TABLE 2. The Global Model's Test Accuracy, Precision, Recall Comparison With Other State-of-The-Art Aggregation Strategies With Our WCGA Method

Model	Accuracy (%)	Precision (%)	Recall (%)
FedAvg	95.25 ± 0.15	93.78 ± 0.12	90.14 ± 0.25
FedOpt	91.14 ± 0.20	90.33 ± 0.18	91.42 ± 0.22
FedNova	92.33 ± 0.18	89.13 ± 0.16	92.20 ± 0.20
FedProx	88.23 ± 0.12	86.14 ± 0.10	86.12 ± 0.14
MultiKrum	95.16 ± 0.22	95.10 ± 0.20	96.12 ± 0.18
FedWGCA	95.46 ± 0.25	95.13 ± 0.18	96.26 ± 0.24

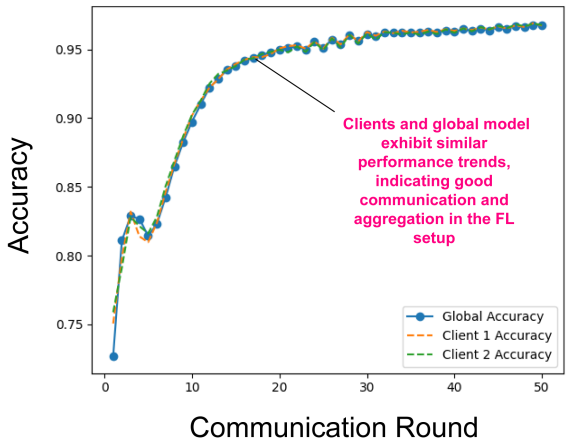

FIGURE 4. Validation Accuracy Curve Performance of FedWGCA training over 50 epochs for Client 1, Client 2, and the global setup.

C. COMPARISON OF AGGREGRATION METHODS

Table 2 compares the global model's test accuracy, precision, and recall across state-of-the-art aggregation strategies. Our proposed FedWGCA method outperforms all, achieving an accuracy of $95.46 \pm 0.25\%$, precision of $95.13 \pm 0.18\%$, and recall of $96.26 \pm 0.24\%$, demonstrating superior robustness in federated learning. FedAvg follows closely with an accuracy of $95.25 \pm 0.15\%$, precision of $93.78 \pm 0.12\%$, and recall of $90.14 \pm 0.25\%$, but its lower recall indicates weaker true positive detection. MultiKrum performs well, with an accuracy of $95.16 \pm 0.22\%$, precision of $95.10 \pm 0.20\%$, and recall of $96.12 \pm 0.18\%$, yet FedWGCA slightly surpasses it in accuracy and recall. FedNova and FedOpt yield moderate results, with accuracies of $92.33 \pm 0.18\%$ and $91.14 \pm 0.20\%$, precisions of $89.13 \pm 0.16\%$ and $90.33 \pm 0.18\%$, and recalls of $92.20 \pm 0.20\%$ and $91.42 \pm 0.22\%$, respectively. FedProx performs worst, with an accuracy of $88.23 \pm 0.12\%$, precision of $86.14 \pm 0.10\%$, and recall of $86.12 \pm 0.14\%$. Fig. 4 shows FedWGCA's validation accuracy curves over 50 epochs, revealing a sharp rise after five epochs, indicating effective client-server communication. Figure 5 and 6 show the confusion matrices of the two clients of our FedWGCA framework and the ROC curve for different classes respectively.

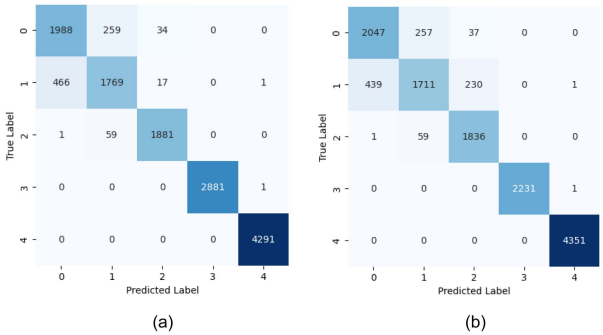

FIGURE 5. Confusion Matrix (a) client 1 and (b) client 2 of our FedWGCA framework.

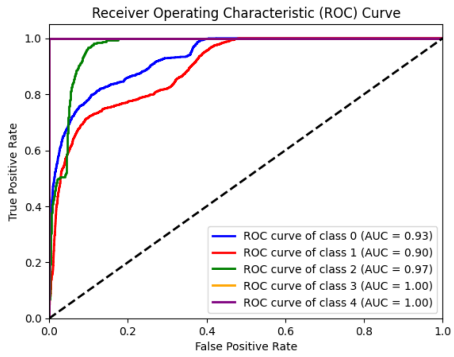

FIGURE 6. ROC Curve.

D. ABLATION STUDIES

1) ANET ABLATION

The ablation study results for the ANET model, presented in Table 3, illustrate the effect of freezing different layers on performance. Freezing the first fully connected layer (fc1) yields the best results, with accuracy 91.06%, precision 91.24%, and recall 91.32%, indicating that retaining fc1's weights preserves substantial learning capacity while allowing other layers to adapt. In contrast, freezing the second fully connected layer (fc2) reduces performance, with accuracy 87.34%, precision 86.38%, and recall 85.32%, suggesting that fc2 is crucial for fine-tuning higher-level features. Freezing the Multi-Head Self-Attention (MHSA) layer also lowers performance, though less drastically, with accuracy 87.44%, precision 84.09%, and recall 83.43%, implying that disrupting the attention mechanism hampers the model's ability to capture complex dependencies. Overall, these findings emphasize the importance of each layer, with freezing different components resulting in varying degrees of performance degradation.

2) FIRST QUARTILE (Q1) THRESHOLD CALCULATION OF FEDWGCA

The results of the ablation study presented in Table 4 demonstrate the impact of varying the first quartile (Q1) threshold on the performance of our FedWGCA method. For all experiments in the ablation study, we set the gradient clipping value to **1.0**, ensuring stability in the FedWGCA method

TABLE 3. Ablation Study of ANET

Model	Accuracy (%)	Precision (%)	Recall (%)
ANET (Layer fc1 Freeze)	91.06	91.24	91.32
ANET (Layer fc2 Freeze)	87.34	86.38	85.32
ANET (MHSA Freeze)	87.44	84.09	83.43

TABLE 4. Ablation Study for Calculating IQR on FedWGCA Method

Model (Q1, Clip Value)	Accuracy (%)
FedWGCA (Q1=0.10, Clip=1.0)	95.66
FedWGCA (Q1=0.15, Clip=1.0)	94.71
FedWGCA (Q1=0.25, Clip=1.0)	95.55
FedWGCA (Q1=0.35, Clip=1.0)	92.46

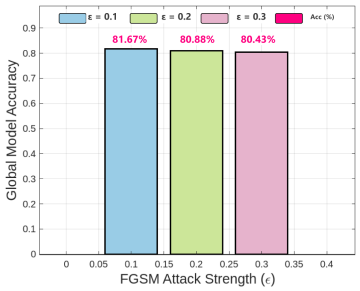

FIGURE 7. FGSM Strength (ε) of applying into our federated aggragraion WSGA method where ANET used as local clasifier.

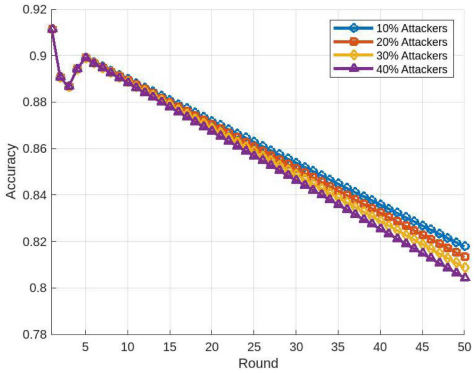

FIGURE 8. Performance Evaluations across multiple number of attackers of the FGSM attack strength.

TABLE 5. Comparison With Privacy-Preserving Methods

Method	Inference Privacy	Outlier Prevention	Adversarial Training
FL-IDS [49]	✗	✗	✗
Danish et al. [50]	✗	✗	✗
Leandro et al. [51]	✗	✗	✗
Syeda et al. [52]	✗	✗	✗
FL-WCGA	✓	✓	✓

while varying the Q1 parameter to analyze its impact on model accuracy. The findings reveal that the choice of Q1 significantly affects the model's accuracy. When Q1 is set to 0.10, FedWGCA achieves the highest accuracy of 95.66%, indicating that a lower threshold for filtering updates captures the most relevant contributions while effectively mitigating the influence of outliers. Conversely, increasing Q1 to 0.15 results in a slight performance drop to 94.71%, suggesting that the stricter inclusion range excludes some beneficial updates. A more conventional Q1 value of 0.25 restores the accuracy to 95.55%, closely matching the optimal performance observed at Q1 = 0.10. However, as Q1 increases further to 0.35, the accuracy declines sharply to 92.46%, likely due to overly relaxed outlier thresholds allowing noisy updates to impact the aggregation. These results highlight the sensitivity of FedWGCA to the Q1 parameter and underscore the importance of careful tuning to balance outlier filtering and model robustness.

E. ATTACK FGSM AND OTHER ATTACK ANALYSIS

In Fig. 7, we asses the robustness of our model by applying the fast gradient sign method (FGSM) [48] attack on the gradients of the model. With an attack strength of $\varepsilon = 0.1$, corresponding to no adversarial perturbation, the model achieves an accuracy of 81.67%. As the attack strength increases to $\varepsilon = 0.2$, the accuracy slightly decreases to 80.88%. Further increasing the attack strength to $\varepsilon = 0.3$ results in a continued decline in accuracy to 80.43%. The observed trend indicates that the model's accuracy diminishes as the strength of the FGSM attack escalates. This degradation in performance underscores the model's vulnerability to adversarial perturbations, highlighting the need for robust defense mechanisms to enhance model resilience against such attacks. Also we have test

Besides in Fig. 8 demonstrates the impact of varying percentages of attackers on the model's accuracy across multiple rounds. As the percentage of attackers increases from 10% to 40%, the model's accuracy shows a consistent decline, starting at approximately 0.92 with 10% attackers and dropping significantly with 40% attackers. This trend highlights the potential effect of malicious actors on the model's performance, with higher attacker proportions leading to greater accuracy degradation. The results emphasize the vulnerability of distributed learning systems to adversarial behavior and underscore the need for robust defense mechanisms to maintain model accuracy in the presence of attackers.

TABLE 6. Comparison of Our Method With Related FL-IDS Approaches on Public Datasets

Method	Accuracy (%)	Precision (%)	Recall (%)
FL-IDS [49]	91.8	90.5	89.7
Danish et al. [50]	92.4	91.2	90.8
Syeda et al. [52]	93.1	92.5	91.4
Ours FedWGCA	**95.5**	**95.1**	**96.3**

TABLE 7. Statistical Significance (p-Values) of FedWGCA Improvements Over Baselines

Model	Accuracy p-value	Precision p-value	Recall p-value
FedAvg	0.0245*	0.0003**	<0.0001**
FedOpt	<0.0001**	<0.0001**	<0.0001**
FedNova	<0.0001**	<0.0001**	<0.0001**
FedProx	<0.0001**	<0.0001**	<0.0001**
MultiKrum	0.1673	0.4097	0.2874

* $p < 0.05$, ** $p < 0.001$ indicate significance levels.

F. STATISTICAL SIGNIFICANCE ANALYSIS

Table 7 shows p-values from paired t-tests comparing FedWGCA with baseline methods. FedWGCA achieves statistically significant improvements in accuracy, precision, and recall over FedAvg, FedOpt, FedNova, and FedProx ($p < 0.05$), with many results highly significant ($p < 0.001$). Differences with MultiKrum are not statistically significant ($p > 0.05$). These findings demonstrate that FedWGCA consistently outperforms most baselines with meaningful improvements.

G. COMPARING RELEVANT SECURITY WORKS

Table 5 presents a comparative analysis of privacy-preserving methods in federated intrusion detection systems (FL-IDS) based on Outlier Prevention and Adversarial Training. Existing methods, including FL-IDS [49], Danish et al. [50], Leandro et al. [51], and Syeda et al. [52], do not incorporate these security measures, making them highly vulnerable to data poisoning and adversarial manipulations. The absence of Outlier Prevention allows malicious client updates to corrupt the model, leading to performance degradation and unreliable decision-making. Similarly, the lack of Adversarial Training leaves these methods defenseless against adversarial perturbations, making them susceptible to evasion attacks. In contrast, our proposed FedWGCA method integrates both Outlier Prevention and Adversarial Training, ensuring a more robust and secure learning framework. By filtering out anomalous updates, Outlier Prevention enhances the reliability of the federated model, while Adversarial Training fortifies it against adversarial threats, improving its resilience in real-world deployment. This dual-layered security strategy makes FedWGCA significantly more robust compared to existing approaches, highlighting the importance of incorporating these mechanisms in federated learning-based IDS frameworks. Besides, as shown in Table 6, our proposed FedWGCA method outperforms existing FL-IDS approaches in terms of accuracy, precision, and recall. This improvement demonstrates the effectiveness of FedWGCA in enhancing intrusion detection performance on public datasets compared to prior methods.

VII. DISCUSSION

A. KEY INSIGHTS

Key findings from the analysis indicate that ANET demonstrates robustness in both local and global federated learning environments, maintaining stable performance. Among the aggregation methods, FedWGCA achieves the highest accuracy (95.46%), precision (95.13%), and recall (96.26%), highlighting its superior effectiveness in federated learning. Ablation studies reveal the critical role of layer freezing in ANET, with optimal performance observed when the *fc1* layer is frozen. Furthermore, adversarial attack analysis identifies the model's vulnerability to FGSM attacks, where accuracy declines as attack strength increases, underscoring the need for robust defense mechanisms.

B. LIMITATIONS AND IMPROVEMENTS

A key limitation of the current model is its vulnerability to FGSM attacks, with accuracy dropping as attack strength increases, showing weak adversarial robustness. Memory efficiency also needs improvement, as methods like FedWGCA require more resources than efficient alternatives like FedOpt. Future work should focus on stronger adversarial defenses—such as adaptive adversarial training, ensemble methods, and detection frameworks—while using lightweight architectures to reduce memory use. To properly preserve privacy, integrating differential privacy and homomorphic encryption is essential. Including these modern baselines or justifying their exclusion will improve the model's robustness, privacy, and efficiency, ensuring secure performance with minimal overhead.

VIII. CONCLUSION

The growing dependence on AAVs in smart city initiatives necessitates resilient and scalable solutions to tackle emerging security challenges. This article presented a federated learning approach aimed at improving the resilience of AAV intrusion detection systems. The framework effectively mitigates adversarial attacks and adapts to dynamic, heterogeneous environments by merging our multi-headed robust attention-based architecture, advanced aggregation techniques, and adversarial training. The integration of these advances guarantees secure and dependable AAV operations, protecting essential services and infrastructure in smart cities. This study highlights the significance of multidisciplinary methods that tackle both security and computational issues, facilitating the development of more resilient and adaptive AAV networks in forthcoming smart city environments. In future endeavors, we intend to augment this federated learning framework by integrating sophisticated defense mechanisms against model poisoning and backdoor attacks, utilizing methodologies such

as differential privacy and homomorphic encryption to bolster security and privacy. Furthermore, investigating self-adaptive federated optimization algorithms that flexibly modify aggregation processes in response to environmental hazards and computational limitations can enhance the model's flexibility.

REFERENCES

[1] S. K. Jagatheesaperumal, S. E. Bibri, J. Huang, J. Rajapandian, and B. Parthiban, "Artificial intelligence of things for smart cities: Advanced solutions for enhancing transportation safety," *Comput. Urban Sci.*, vol. 4, no. 1, 2024, Art. no. 10.

[2] S. E. Bibri, J. Huang, and J. Krogstie, "Artificial intelligence of things for synergizing smarter eco-city brain, metabolism, and platform: Pioneering data-driven environmental governance," *Sustain. Cities Soc.*, vol. 108, 2024, Art. no. 105516.

[3] A. Telikani et al., "Unmanned aerial vehicle-aided intelligent transportation systems: Vision, challenges, and opportunities," *IEEE Commun. Surveys Tut.*, early access, Jan. 20, 2025, doi: 10.1109/COMST.2025.3530913.

[4] Z. Liu, H. Ye, C. Chen, Y. Zheng, and K. -Y. Lam, "Threats, attacks, and defenses in machine unlearning: A survey," *IEEE Open J. Comput. Soc.*, vol. 6, pp. 413–425, 2025.

[5] R. Reyes-Acosta, C. Dominguez-Baez, R. Mendoza-Gonzalez, and M. Vargas Martin, "Analysis of machine learning-based approaches for securing the Internet of Things in the smart industry: A multivocal state of knowledge review," *Int. J. Inf. Secur.*, vol. 24, no. 1, 2025, Art. no. 31.

[6] A. Musa, A. O. Adebayo, P. O. Balogun, and S. Jacob, "Security and privacy of future wireless communication systems," in *Proc. Massive MIMO Future Wireless Commun. Syst., Technol. Appl.*, 2025, pp. 23–57.

[7] S. K. Kim and H. C. Vong, "Secured network architectures based on blockchain technologies: A systematic review," *ACM Comput. Surv.*, vol. 57, no. 7, pp. 1–24, 2025.

[8] C. Dong, S. Pal, S. Chen, F. Jiang, and X. Liu, "A privacy-aware task distribution architecture for UAV communications system using blockchain," *IEEE Internet Things J.*, vol. 12, no. 9, pp. 11233–11243, May 2025.

[9] A. Khan, N. Jhanjhi, D. H. H. Hamid, H. A. H. B. H. Omar, F. Amsaad, and H. Ashraf, "Reshaping cybersecurity for supply chain sustainability in IR 4.0 and 5.0," in *AI Techn. Securing Med. Bus. Practices*. Hershey, PA, USA: IGI Glob. Sci., 2025.

[10] H. Zhao and P. Li, "An efficient class-dependent learning label approach using feature selection to improve multi-label classification algorithms," *Cluster Comput.*, vol. 28, no. 1, 2025, Art. no. 38.

[11] K. Cengiz, S. Lipsa, R. K. Dash, N. Ivković, and M. Konecki, "A novel intrusion detection system based on artificial neural network and genetic algorithm with a new dimensionality reduction technique for UAV communication," *IEEE Access*, vol. 12, pp. 4925–4937, 2024.

[12] A. Qayyum, K. Ahmad, M. A. Ahsan, A. Al-Fuqaha, and J. Qadir, "Collaborative federated learning for healthcare: Multi-modal COVID-19 diagnosis at the edge," *IEEE Open J. Comput. Soc.*, vol. 3, pp. 172–184, 2022.

[13] Y. Cheng, W. Zhang, Z. Zhang, C. Zhang, S. Wang, and S. Mao, "Toward federated large language models: Motivations, methods, and future directions," *IEEE Commun. Surveys Tuts.*, vol. 27, no. 4, pp. 2733–2764, Aug. 2025.

[14] A. Sreevallabh Chivukula, X. Yang, B. Liu, W. Liu, and W. Zhou, "Adversarial defense mechanisms for supervised learning," in *Adversarial Machine Learning: Attack Surfaces, Defence Mechanisms, Learning Theories in Artificial Intelligence*. Cham, Switzerland: Springer, 2022, pp. 151–238.

[15] P. Sun, X. Liu, Z. Wang, and B. Liu, "Byzantine-robust decentralized federated learning via dual-domain clustering and trust bootstrapping," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 24756–24765.

[16] A. Cassimon, S. Mercelis, and K. Mets, "Scalable reinforcement learning-based neural architecture search," *Neural Comput. Appl.*, vol. 37, pp. 231–261, 2024.

[17] C. Pan and X. Yao, "Neural architecture search based on evolutionary algorithms with fitness approximation," in *Proc. Int. Joint Conf. Neural Netw.*, 2021, pp. 1–8.

[18] S. Ali and M. A. Wani, "Gradient-based neural architecture search: A comprehensive evaluation," *Mach. Learn. Knowl. Extraction*, vol. 5, no. 3, pp. 1176–1194, 2023.

[19] M. Ahmed, D. Cox, B. Simpson, and A. Aloufi, "ECU-IoFT: A dataset for analysing cyber-attacks on Internet of Flying Things," *Appl. Sci.*, vol. 12, no. 4, 2022, Art. no. 1990.

[20] A. A. Hady, A. Ghubaish, T. Salman, D. Unal, and R. Jain, "Intrusion detection system for healthcare systems using medical and network data: A comparison study," *IEEE Access*, vol. 8, pp. 106576–106584, 2020.

[21] J. Castillo, P. Rieger, H. Fereidooni, Q. Chen, and A. Sadeghi, "FLEDGE: Ledger-based federated learning resilient to inference and backdoor attacks," in *Proc. 39th Annu. Comput. Secur. Appl. Conf.*, 2023, pp. 647–661.

[22] A. E. Samy and Š. Girdzijauskas, "Mitigating sybil attacks in federated learning," in *Proc. Int. Conf. Inf. Secur. Pract. Experience*, Singapore, 2023, pp. 36–51.

[23] J. Liu, J. H. Wang, C. Rong, Y. Xu, T. Yu, and J. Wang, "FedPa: An adaptively partial model aggregation strategy in federated learning," *Comput. Netw.*, vol. 199, 2021, Art. no. 108468.

[24] S. Chen, C. Shen, L. Zhang, and Y. Tang, "Dynamic aggregation for heterogeneous quantization in federated learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6804–6819, Oct. 2021.

[25] L. Hu, H. Yan, L. Li, Z. Pan, X. Liu, and Z. Zhang, "MHAT: An efficient model-heterogenous aggregation training scheme for federated learning," *Inf. Sci.*, vol. 560, pp. 493–503, 2021.

[26] A. K. Samha, "Strategies for efficient resource management in federated cloud environments supporting infrastructure as a service (IaaS)," *J. Eng. Res.*, vol. 12, no. 2, pp. 101–114, 2024.

[27] J. Sun et al., "An efficient privacy-aware split learning framework for satellite communications," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 12, pp. 3355–3365, Dec. 2024.

[28] Z. Chen, J. Peng, J. Kang, and D. Niyato, "FedGroup-Prune: IoT device amicable and training-efficient federated learning via combined group lasso sparse model pruning," *IEEE Internet Things J.*, vol. 11, no. 24, pp. 40921–40932, Dec. 2024.

[29] W. Guo, Y. Wang, X. Chen, and P. Jiang, "Federated transfer learning for auxiliary classifier generative adversarial networks: Framework and industrial application," *J. Intell. Manuf.*, vol. 35, no. 4, pp. 1439–1454, 2024.

[30] J. Shao, F. Wu, and J. Zhang, "Selective knowledge sharing for privacy-preserving federated distillation without a good teacher," *Nature Commun.*, vol. 15, no. 1, 2024, Art. no. 349.

[31] A. Samanta, I. Hatai, and A. K. Mal, "A survey on hardware accelerator design of deep learning for edge devices," *Wireless Pers. Commun.*, vol. 137, no. 3, pp. 1715–1760, 2024.

[32] E. F. Kfoury, S. Choueiri, A. Mazloum, A. AlSabeh, J. Gomez, and J. Crichigno, "A comprehensive survey on smartNICs: Architectures, development models, applications, and research directions," *IEEE Access*, vol. 12, pp. 107297–107336, 2024.

[33] I. J. Mouri, M. Ridowan, and M. A. Adnan, "Data poisoning attacks and mitigation strategies on federated support vector machines," *SN Comput. Sci.*, vol. 5, no. 2, 2024, Art. no. 241.

[34] D. Yu et al., "Robust federated learning for edge intelligence," in *Handbook of Trustworthy Federated Learning*. Cham, Switzerland: Springer, 2024, pp. 323–366.

[35] A. Song, H. Li, K. Cheng, T. Zhang, A. Sun, and Y. Shen, "GuardFL: An UMAP-assisted robust aggregation for federated learning," *IEEE Internet Things J.*, vol. 11, no. 16, pp. 27557–27570, Aug. 2024.

[36] L. Collins, H. Hassani, A. Mokhtari, and S. Shakkottai, "FedAvg with fine tuning: Local updates lead to representation learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, vol. 35, pp. 10572–10586.

[37] S. T. Ahmed et al., "FedOPT: Federated learning-based heterogeneous resource recommendation and optimization for edge computing," *Soft Comput.*, pp. 1–12, 2024.

[38] J. Wang, Q. Liu, H. Liang, G. Joshi, and H. V. Poor, "Tackling the objective inconsistency problem in heterogeneous federated optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 7611–7623.

[39] X. Yuan and P. Li, "On convergence of FedProx: Local dissimilarity invariant bounds, non-smoothness and beyond," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, vol. 35, pp. 10752–10765.

[40] P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer, "Machine learning with adversaries: Byzantine tolerant gradient descent," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 118–128.
[41] R. S. Bedi and P. S. Rana, "Machine learning based android malware detection: A comprehensive survey," *Turkish Online J. Qualitative Inquiry*, vol. 12, no. 10, 2021, Art. no. 103654.
[42] Y. Wu, L. Yang, L. Zhang, L. Nie, and L. Zheng, "Intrusion detection for unmanned aerial vehicles security: A tiny machine learning model," *IEEE Internet Things J.*, vol. 11, no. 12, pp. 20970–20982, Jun. 2024.
[43] R. A. Ramadan, A. H. Emara, M. Al-Sarem, and M. Elhamahmy, "Internet of drones intrusion detection using deep learning," *Electronics*, vol. 10, no. 21, 2021, Art. no. 2633.
[44] H. J. Hadi, Y. Cao, S. Li, Y. Hu, J. Wang, and S. Wang, "Real-time collaborative intrusion detection system in UAV networks using deep learning," *IEEE Internet Things J.*, vol. 11, no. 20, pp. 33371–33391, Oct. 2024.
[45] S. C. Hassler, U. A. Mughal, and M. Ismail, "Cyber-physical intrusion detection system for unmanned aerial vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 6, pp. 6106–6117, Jun. 2024.
[46] A. Vaswani, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
[47] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
[48] Y. Liu, S. Mao, X. Mei, T. Yang, and X. Zhao, "Sensitivity of adversarial perturbation in fast gradient sign method," in *Proc. IEEE Symp. Ser. Comput. Intell.*, 2019, pp. 433–436.
[49] M. H. Bhavsar, Y. B. Bekele, K. Roy, J. C. Kelly, and D. Limbrick, "FL-IDS: Federated learning-based intrusion detection system using edge devices for transportation IoT," *IEEE Access*, vol. 12, pp. 52215–52226, 2024.
[50] D. Javeed, M. S. Saeed, M. Adil, P. Kumar, and A. Jolfaei, "A federated learning-based zero trust intrusion detection system for Internet of Things," *Ad Hoc Netw.*, vol. 162, 2024, Art. no. 103540.
[51] L. M. Da Silva, I. G. Ferrão, C. Dezan, D. Espes, and K. R. Branco, "Anomaly-based intrusion detection system for in-flight and network security in UAV swarm," in *Proc. Int. Conf. Unmanned Aircr. Syst.*, 2023, pp. 812–819.
[52] 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, 2024, Art. no. 3194.
[53] J. Kim, N. Shin, S. Y. Jo, and S. H. Kim, "Method of intrusion detection using deep neural network," in *Proc. IEEE Int. Conf. Big Data Smart Comput.*, 2017, pp. 313–316.
[54] M. A. M. Hasan, M. Nasser, B. Pal, and S. Ahmad, "Support vector machine and random forest modeling for intrusion detection system (IDS)," *J. Intell. Learn. Syst. Appl.*, vol. 6, no. 1, pp. 45–52, 2014.
[55] Y. Wang, J. Liu, and X. Zhang, "Federated learning-based intrusion detection system for UAV networks using attention mechanisms," *IEEE Internet Things J.*, vol. 10, no. 5, pp. 4321–4335, 2023.

MD. FAHIM-UL-ISLAM received the B.Sc. degree in computer science and engineering from BRAC University, Dhaka, Bangladesh. He is currently working toward the Ph.D. degree in computer science with Arizona State University, Tempe, AZ, USA. He was a Graduate Research Assistant while completing his master's degree. His academic pursuits are complemented by active engagement in research. His research interests include multiple fields, with a particular focus on multi-modal machine learning, computer vision, and adversarial machine learning.

AMITABHA CHAKRABARTY received the M.Sc. degree from the Department of Computer Science and Engineering, University of Rajshahi, Rajshahi, Bangladesh, in 2004, the second M.Sc. degree in telecommunication engineering from Independent University, Dhaka, Bangladesh, and the Ph.D. degree from the Faculty of Engineering and Computing, Dublin City University, Dublin, Ireland, in 2012. He is currently a Professor with the Department of Computer Science and Engineering, BRAC University, Dhaka. He supervises multiple research groups and projects and has published extensively. His research interests include IoT, machine learning, deep learning, vision transformers, embedded systems, and switching theory. He is a TCP member and reviewer for various international journals and conferences and a Senior Judge in national IT competitions.

HALIMATON SAADIAH HAKIMI is currently an Academic and Researcher with expertise in emerging technologies, workforce readiness, and talent development. She holds professional recognition as a Professional Technologist and has contributed extensively to higher education, particularly in advancing technical and vocational education and training. She is actively involved in publications, conferences, and collaborative projects, aiming to bridge the gap between academia and industry while championing knowledge sharing, leadership, and sustainable academic excellence. Her research interests include artificial intelligence applications, digital transformation, and innovation in education to prepare graduates for future industry demands.

SITI SARAH MAIDIN received the B.Sc. degree in computer science from University Teknikal Malaysia Melaka, Durian Tunggal, Malaysia, the M.Sc. degree in information technology from University Teknologi MARA, Shah Alam, Malaysia, and the Ph.D. degree in information and communication technology from Universiti Tenaga Nasional, Kajang, Malaysia. She is currently a Professor with the Department of Data Science and Information Technology, INTI International University, Nilai, Malaysia, and also the Head of Programme and Deputy Director. She has supervised numerous undergraduate and postgraduate students. She also evaluates university and government research grants and audits the MyRA University ranking in Malaysia. Her research interests include software engineering, educational technology, and data science. She is the Editor of the *Journal of Innovation and Technology*.