TAFIC-Edge: Trust-Aware Federated Intrusion Detection with Calibrated Fusion for Secure Edge-IoT Systems

*

1st Vikas Shukla
Information Technology
Amity Institute of Information Technology
Amity University
Noida, India
vikas.shukla8527@gmail.com

2nd Rekha Agarwal
Information Technology
Amity Institute of Information Technology
Amity University
Noida, India
ragarwal@amity.edu

3rd Rajesh Kumar Tyagi
Computer Science
Amity University,
Gurugram, Haryana
rktyagi@ggn.amity.edu

***Abstract*—Edge-IoT intrusion detection must be accurate, reliable, privacy-preserving, and robust against malicious clients. However, existing IDS methods often focus on overall accuracy while overlooking class imbalance, probability calibration, and client reliability in federated settings. To address these issues, this paper proposes TAFIC-Edge, a Trust-Aware Federated Intrusion and Compliance Framework for secure edge computing. TAFIC-Edge combines Decision Tree, Random Forest, Extra Trees, HistGradientBoosting, LightGBM, and XGBoost using validation-based weighted fusion. Temperature scaling improves calibration, while minority threshold tuning improves balanced attack detection. In the federated setting, client trust is computed using macro-F1, balanced accuracy, LogLoss, ECE, distribution quality, and prediction disagreement, and low-trust clients are filtered before aggregation. Experiments on CICIoT2023 show that the proposed model achieves 0.994821 accuracy, 0.857987 balanced accuracy, and 0.830706 macro-F1. Under 0–30% malicious-client ratios, trust-aware aggregation filters malicious clients and substantially reduces LogLoss and ECE. The results show that TAFIC-Edge improves balanced detection, calibrated reliability, and robustness for secure edge-IoT intrusion detection.**

***Index Terms*—Edge-IoT security, Intrusion detection, Federated learning, Trust-aware aggregation, Probability calibration**

I. Introduction

Edge computing supports low-latency IoT, IIoT, smart-city, healthcare, and cyber-physical applications by processing data near the source. However, this distributed design also increases the attack surface because edge nodes are resource-constrained, geographically dispersed, intermittently connected, and often exposed to untrusted environments. These systems are vulnerable to DDoS, DoS, spoofing, brute-force, reconnaissance, malware, Web-based attacks, poisoning, and malicious-client behavior. Datasets such as CICIoT2023 have enabled systematic evaluation of IoT intrusion detection under realistic attack conditions [1].

Identify applicable funding agency here. If none, delete this.

Recent IoT IDS studies have widely used machine learning, deep learning, ensemble models, and federated learning. Models such as Decision Tree, Random Forest, XGBoost, LightGBM, and CatBoost are popular because they work well on tabular network-flow features and are suitable for lightweight deployment [2]–[7]. Federated learning further supports privacy-preserving distributed IDS by allowing edge clients to train without sharing raw data [8]–[10].

Despite these advances, existing works still face key limitations. High accuracy may hide poor minority-class detection under severe imbalance, and probability calibration is often ignored despite its importance for automated security decisions. Federated IDS also remains vulnerable to non-IID data, noisy clients, poisoning attacks, and unreliable participants [11]–[13]. To address these issues, this paper proposes TAFIC-Edge, which integrates calibrated weighted-fusion IDS with minority threshold tuning and trust-aware federated aggregation. Client trust is computed using macro-F1, balanced accuracy, LogLoss, ECE, distribution quality, and prediction disagreement, enabling low-trust clients to be filtered before aggregation. The main contributions of this paper are summarized as follows:

- A calibrated weighted-fusion IDS is proposed for edge-IoT intrusion detection by combining Decision Tree, Random Forest, Extra Trees, HistGradientBoosting, LightGBM, and XGBoost models using validation-based fusion weights.
- A minority threshold tuning mechanism is introduced to improve balanced accuracy and minority-sensitive detection under imbalanced IoT attack distributions.
- A trust-aware federated intrusion detection framework is developed for edge environments, where client trust is computed using detection quality, calibration reliability, data-distribution quality, and prediction disagreement.
- A malicious-client filtering mechanism is incorporated

to reduce the influence of noisy or poisoned clients in federated edge learning.

The remainder of this paper is organized as follows. Section II reviews related work. Section III presents the proposed TAFIC-Edge framework. Section IV describes the experimental setup. Section V discusses the results, and Section VI concludes the paper with future directions.

II. Related Work

Intrusion detection in IoT and edge computing has become important due to the growth of distributed and resource-constrained systems. Realistic datasets such as CICIoT2023 support evaluation under diverse attacks including DDoS, DoS, brute-force, spoofing, reconnaissance, Web-based attacks, and Mirai [1]. Recent studies have used CICIoT2023 and related datasets with clustering, machine learning, and gradient boosting methods [14]–[16]. Machine learning and ensemble models such as Decision Tree, Random Forest, XGBoost, LightGBM, and MLP are widely used for IoT IDS because they perform well on tabular flow features and are suitable for lightweight deployment [2]–[4], [7], [17]. Deep and lightweight IDS methods have also been explored for industrial IoT and edge settings [5], [6], [18], [19]. Federated IDS further improves privacy by allowing clients to train collaboratively without sharing raw data [8]–[10], [13].

However, many existing works emphasize accuracy while giving less attention to balanced accuracy, macro-F1, minority-class detection, and calibrated reliability. Federated IDS also remains vulnerable to non-IID data, poisoning, noisy clients, and unreliable participants [11], [12]. Although trust-aware and calibration-aware methods exist [20], [21], they rarely combine calibrated local detection, threshold tuning, trust-based client filtering, and malicious-client robustness. To address these gaps, TAFIC-Edge integrates calibrated weighted-fusion IDS with trust-aware federated aggregation for balanced detection, reliable confidence, and malicious-client filtering.

III. Proposed Methodology

The proposed TAFIC-Edge framework has two main components: a calibrated local IDS and a trust-aware federated aggregation module. The local IDS improves detection using calibrated weighted fusion and minority threshold tuning, while the federated module filters low-trust or malicious clients before aggregation.

Let $\mathcal{F} = \{f_1, f_2, \ldots, f_M\}$ be the set of base classifiers, including Decision Tree, Random Forest, Extra Trees, HistGradientBoosting, LightGBM, and XGBoost. For each sample $\mathbf{x}_i$, classifier f_m produces a probability vector:

$$\mathbf{p}_i^{(m)} = [p_{i,1}^{(m)}, p_{i,2}^{(m)}, \ldots, p_{i,C}^{(m)}].$$

Since raw probabilities may be poorly calibrated, temperature scaling is applied as

$$\tilde{p}_{i,c}^{(m)} = \frac{\left(p_{i,c}^{(m)}\right)^{1/T_m}}{\sum_{j=1}^{C}\left(p_{i,j}^{(m)}\right)^{1/T_m}},$$

where T_m is selected by minimizing validation LogLoss. Each classifier is then assigned a validation-based score using Macro-F1 and balanced accuracy:

$$s_m = \lambda \cdot \text{Macro-F1}_m + (1-\lambda)\cdot \text{BA}_m, \qquad w_m = \frac{s_m}{\sum_{j=1}^{M} s_j}.$$

The calibrated weighted-fusion probability is computed as

$$\mathbf{p}_i^{\text{fusion}} = \sum_{m=1}^{M} w_m \tilde{\mathbf{p}}_i^{(m)}.$$

To improve minority-class detection, class-wise threshold tuning is applied. For class c, the bias factor is

$$b_c = \left(\frac{1/n_c}{\frac{1}{C}\sum_{j=1}^{C} 1/n_j}\right)^{\rho},$$

where n_c is the number of validation samples in class c, and ρ is selected on validation data. The adjusted probability is

$$p_{i,c}^{\text{adj}} = \frac{p_{i,c}^{\text{fusion}} b_c}{\sum_{j=1}^{C} p_{i,j}^{\text{fusion}} b_j},$$

and the final local prediction is

$$\hat{y}_i = \arg\max_{c\in\mathcal{Y}} p_{i,c}^{\text{adj}}.$$

In the federated setting, each client k trains a local IDS model on $\mathcal{D}_k$. Since some clients may be noisy or malicious, a trust score is computed before aggregation. Let MF1_k, BA_k, $\mathcal{L}_k$, ECE_k, Q_k, and D_k denote Macro-F1, balanced accuracy, normalized LogLoss, ECE, distribution quality, and prediction disagreement, respectively. The trust score is defined as

$$T_k = [\alpha_1 \text{MF1}_k + \alpha_2 \text{BA}_k + \alpha_3(1-\mathcal{L}_k) + \alpha_4(1-\text{ECE}_k) + \alpha_5 Q_k]\,(1-\gamma D_k),$$

where $\sum_{r=1}^{5}\alpha_r = 1$, and γ controls the disagreement penalty. A client is selected only if

$$\mathcal{S} = \{k \in \mathcal{K} : T_k \geq \tau\}.$$

TABLE I
COMPARATIVE ANALYSIS OF PROMINENT STUDIES RELATED TO IOT/EDGE INTRUSION DETECTION AND FEDERATED IDS

S. No.	Study	Application / Dataset	Method / Model	Learning Setting	Calibration	Trust / Poisoning Defense	Key Limitation
1	Neto et al. [1]	CICIoT2023 / IoT security	Large-scale IoT attack dataset and benchmark	Centralized benchmark	No	No	Dataset only; no trust-aware federated IDS.
2	Gheni and Al-Yaseen [14]	CICIoT2023	Two-step data clustering for IDS	Centralized	No	No	No calibration or federated robustness.
3	Houichi et al. [15]	Smart city / CICIoT2023	Machine learning based IDS	Centralized	No	No	No trust-aware FL or calibration analysis.
4	Almahaqeri et al. [16]	CICIoT2023	Optimized gradient boosting framework	Centralized	No	No	No malicious-client or ECE/LogLoss analysis.
5	Adewole et al. [2]	IoT IDS	Rule induction and ensemble learning	Centralized	No	No	No federated trust or client filtering.
6	Kouassi et al. [3]	IoT IDS	Top-K feature selection using XGBoost, LightGBM, and Random Forest	Centralized	No	No	No calibrated fusion or edge FL setting.
7	Bhutta et al. [19]	WiFi / IoT IDS	Lightweight real-time LightGBM IDS	Centralized / lightweight	No	No	No federated or trust-aware defense.
8	Rashid et al. [8]	Industrial IoT IDS	Federated learning based IDS	Federated	No	Limited	Limited calibration and poisoning analysis.
9	Yang et al. [11]	IoT IDS	Dependable FL against poisoning attacks	Federated	No	Yes	No calibrated weighted-fusion IDS.
10	Sanjalawe et al. [12]	IoT IDS	Adaptive graph attention based FL	Federated	No	Yes	Limited focus on calibration reliability.
11	Wang et al. [13]	Network IDS	Secure aggregation with gradient similarity	Federated	No	Yes	No minority-threshold tuning.
12	Talpini et al. [21]	Distributed ML-based IDS	Federated calibration for IDS	Federated	Yes	No	No malicious-client filtering.

Algorithm 1 TAFIC-Edge: Trust-Aware Federated IDS

Data $\mathcal{D}$, client data $\{\mathcal{D}_k\}_{k=1}^{K}$, base models $\mathcal{F}$, validation set $\mathcal{D}_v$, threshold τ Local prediction $\hat{y}$ and federated prediction $\hat{y}^{\text{fed}}$

Train base models $f_m \in \mathcal{F}$ and calibrate them using validation LogLoss Compute fusion weights w_m from validation Macro-F1 and balanced accuracy Fuse calibrated probabilities and apply minority threshold tuning:

$$\mathbf{p}^{\text{fusion}} = \sum_{m=1}^{M} w_m \tilde{\mathbf{p}}^{(m)}, \qquad \hat{y} = \arg\max_c p_c^{\text{adj}}$$

client $k \in \mathcal{K}$ Train local model on $\mathcal{D}_k$ Compute trust score using MF1_k, BA_k, $\mathcal{L}_k$, ECE_k, Q_k, and D_k

$$T_k = [\alpha_1 \text{MF1}_k + \alpha_2 \text{BA}_k + \alpha_3(1-\mathcal{L}_k) + \alpha_4(1-\text{ECE}_k) + \alpha_5 Q_k](1-\gamma D_k)$$

Select trusted clients:

$$\mathcal{S} = \{k : T_k \geq \tau\}$$

Aggregate trusted predictions and generate final output:

$$\mathbf{p}^{\text{fed}} = \sum_{k \in \mathcal{S}} \frac{T_k}{\sum_{j \in \mathcal{S}} T_j} \mathbf{p}_k, \qquad \hat{y}^{\text{fed}} = \arg\max_c p_c^{\text{fed}}$$

return $\hat{y}$, $\hat{y}^{\text{fed}}$

The final trust-aware federated prediction is obtained as

$$\mathbf{p}^{\text{fed}}(\mathbf{x}) = \sum_{k \in \mathcal{S}} \omega_k \mathbf{p}_k(\mathbf{x}), \qquad \omega_k = \frac{T_k}{\sum_{j \in \mathcal{S}} T_j}.$$

Finally,

$$\hat{y}^{\text{fed}} = \arg\max_{c \in \mathcal{Y}} p_c^{\text{fed}}(\mathbf{x}).$$

This design improves balanced detection locally and reduces the influence of unreliable or malicious clients in the federated edge setting.

A. Computational Complexity

Let N be the number of training samples, d the number of features, M the number of base classifiers, K the number of clients, and C the number of classes. The local training cost depends on the base learners and can be expressed generally as

$$\mathcal{O}\left(\sum_{m=1}^{M} \mathcal{C}(f_m)\right),$$

where $\mathcal{C}(f_m)$ is the training complexity of classifier f_m. Probability fusion requires

$$\mathcal{O}(NMC)$$

operations because probabilities from M classifiers are aggregated over C classes. In the federated stage, trust computation for all clients requires

$$\mathcal{O}(KN_vC),$$

where N_v is the size of the validation set. Trust-aware prediction aggregation requires

$$\mathcal{O}(|\mathcal{S}|C)$$

per sample, where $|\mathcal{S}|$ is the number of selected trusted clients. Since malicious or unreliable clients are filtered before aggregation, $|\mathcal{S}| \leq K$, which reduces unnecessary aggregation overhead and improves robustness.

IV. Experimental Setup

The experiments are conducted on the CICIoT2023 dataset using coarse attack labels: benign, DDoS, DoS, Mirai, spoofing, reconnaissance, brute-force, and Web-based attacks. Since the data are highly imbalanced, balanced accuracy, macro-F1, per-class F1, LogLoss, Expected Calibration Error (ECE), and Brier score are used along with accuracy. The data are preprocessed by removing non-numeric and constant features, imputing missing values, encoding labels, and applying z-score normalization. A stratified 60:20:20 split is used for training, validation, and testing.

The proposed method is compared with Decision Tree, Random Forest, Extra Trees, HistGradientBoosting, Logistic Regression, MLP, LightGBM, and XGBoost. The local TAFIC-Edge IDS combines Decision Tree, Random Forest, Extra Trees, HistGradientBoosting, LightGBM, and XGBoost using calibrated weighted fusion. Temperature scaling is applied using validation LogLoss, fusion weights are computed from validation macro-F1 and balanced accuracy, and minority threshold tuning is applied with $\rho = 0.3$.

For federated evaluation, the training data are distributed across $K = 10$ clients using Dirichlet non-IID partitioning with $\alpha = 1.2$. Experiments are repeated over five seeds: 42, 52, 62, 72, and 82. Malicious-client ratios of 0%, 10%, 20%, and 30% are evaluated using noisy-label corruption. Four settings are considered: centralized proposed IDS, local-only IDS, standard federated ensemble IDS, and trust-aware federated IDS. Clients with trust scores below $\tau = 0.60$ are excluded from aggregation. The implementation is done in Python on Kaggle using pandas, NumPy, scikit-learn, LightGBM, XGBoost, and Matplotlib.

V. Results and Discussion

This section evaluates the proposed TAFIC-Edge framework on CICIoT2023 using balanced accuracy, macro-F1, LogLoss, ECE, and per-class F1, since the dataset is highly imbalanced. After preprocessing, the dataset contains 449,886 samples, 41 numerical features, and eight classes: Benign, BruteForce, DDoS, DoS, Mirai, Recon, Spoofing, and WebBased. The data are split into 287,926 training samples, 71,982 validation samples, and 89,978 test samples. Figure 1 shows the Dirichlet non-IID distribution across 10 edge clients, confirming heterogeneous client-wise traffic distribution. Table II compares the proposed IDS with standard baselines. Decision Tree gives the highest macro-F1 among individual models, but its LogLoss is high. XGBoost gives the lowest LogLoss, but its balanced accuracy is limited. The proposed threshold-tuned fusion achieves the best balanced accuracy of 0.857987 and competitive macro-F1 of 0.830706, showing improved minority-sensitive detection. Figure 2 shows that the proposed model remains close to the best macro-F1 baseline while improving balanced accuracy and reliability. Table III presents the fusion and ablation analysis. Temperature calibration improves macro-F1 from 0.781822 to 0.823761 and reduces LogLoss from 0.021671 to 0.016747. The proposed threshold-tuned fusion gives the best balanced accuracy and macro-F1 among fusion variants, confirming the benefit of minority threshold tuning. Figure 3 confirms that threshold tuning gives the largest balanced accuracy improvement.

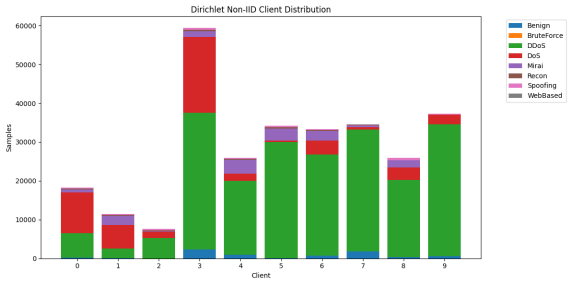

Fig. 1. Dirichlet non-IID client distribution across edge clients.

TABLE II
Baseline comparison with the proposed local IDS

Model	Acc.	Bal. Acc.	Macro-F1	LogLoss
Decision Tree	0.992998	0.832332	**0.831346**	0.193464
Random Forest	0.993999	0.738983	0.777868	0.024841
Extra Trees	0.992331	0.719412	0.764626	0.042483
HistGradientBoosting	0.992476	0.769671	0.736794	0.033227
XGBoost	0.994632	0.709668	0.723692	**0.015386**
LightGBM	0.993976	0.695628	0.700035	0.021632
MLP	0.989175	0.663589	0.681256	0.032288
Logistic Regression	0.711374	0.647203	0.508099	0.537407
Proposed IDS	**0.994821**	**0.857987**	0.830706	0.021570

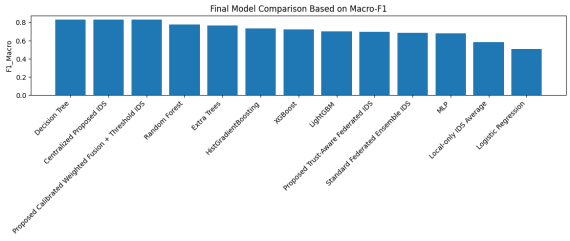

Fig. 2. Final model comparison based on macro-F1.

TABLE III
Fusion and ablation analysis

Variant	Acc.	Bal. Acc.	Macro-F1	LogLoss
Raw Equal Fusion	0.995143	0.747599	0.781822	0.021671
Calibrated Equal Fusion	**0.995332**	0.781483	0.823761	**0.016747**
Calibrated Weighted Fusion	0.995310	0.781473	0.823635	0.016836
Proposed Threshold-Tuned Fusion	0.994821	**0.857987**	**0.830706**	0.021570
Logistic Stacking Fusion	0.993243	0.827792	0.739773	0.023132

Table IV shows the per-class results of the proposed IDS. DDoS, DoS, and Mirai achieve near-perfect F1-scores. Benign, Recon, and Spoofing are also detected effectively. BruteForce and WebBased remain difficult because they have very low

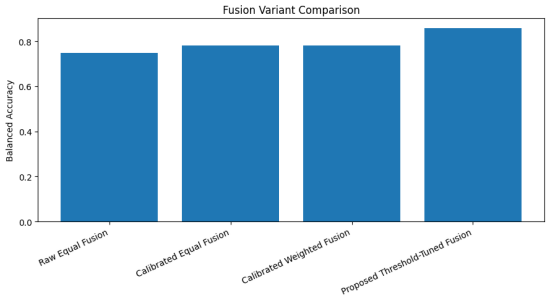

Fig. 3. Fusion variant comparison based on balanced accuracy.

support, but their recall values indicate improved minority-class sensitivity.

TABLE IV
PER-CLASS PERFORMANCE OF THE PROPOSED LOCAL IDS

Class	Precision	Recall	F1	Support
Benign	0.913569	0.949721	0.931294	2148
BruteForce	0.483871	0.652174	0.555556	23
DDoS	0.999954	0.999542	0.999748	65478
DoS	0.999360	0.999616	0.999488	15631
Mirai	0.999801	1.000000	0.999901	5030
Recon	0.839428	0.771930	0.804265	684
Spoofing	0.877996	0.845750	0.861571	953
WebBased	0.400000	0.645161	0.493827	31
Macro Avg.	0.814247	0.857987	0.830706	89978

Figure 4 shows that standard federated aggregation becomes unreliable as the malicious-client ratio increases, while trust-aware aggregation keeps LogLoss stable.

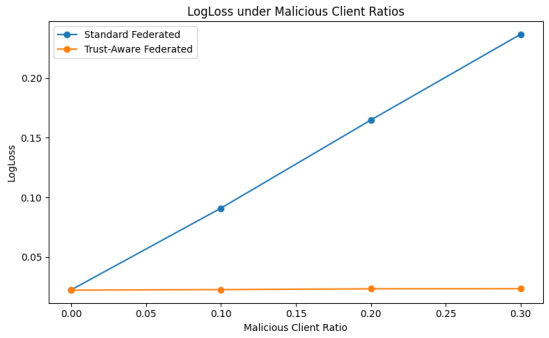

Fig. 4. LogLoss under malicious-client ratios.

Table V summarizes trust-based malicious-client filtering. At 10%, 20%, and 30% malicious-client ratios, all malicious clients are filtered and no benign clients are removed. Figure 5 further shows that malicious clients fall below the trust threshold, validating the proposed trust score. Overall, TAFIC-Edge improves balanced detection and reliability in edge-IoT IDS. The local IDS achieves the highest balanced accuracy with competitive macro-F1, while trust-aware aggregation provides modest macro-F1 gains and strong LogLoss/ECE reduction under malicious-client attacks. Therefore, the framework is best interpreted as a reliability- and robustness-oriented IDS rather than a model claiming superiority in every metric.

TABLE V
TRUST-BASED MALICIOUS-CLIENT FILTERING

Ratio	Malicious	Selected	Filtered	Rate
0.0	0	50	0	0.0
0.1	5	45	5	1.0
0.2	10	40	10	1.0
0.3	15	35	15	1.0

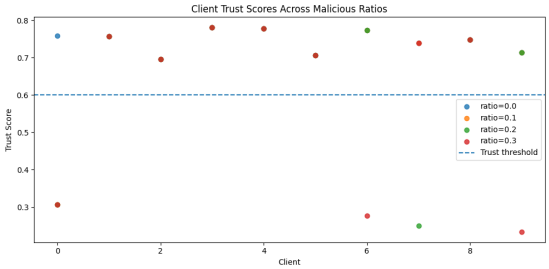

Fig. 5. Client trust scores across malicious-client ratios.

VI. CONCLUSION AND FUTURE DIRECTIONS

This paper presented TAFIC-Edge, a trust-aware federated IDS for secure edge-IoT environments. It combines calibrated weighted fusion of Decision Tree, Random Forest, Extra Trees, HistGradientBoosting, LightGBM, and XGBoost with minority threshold tuning to improve balanced detection under class imbalance. On CICIoT2023, the proposed model achieved 0.994821 accuracy, 0.857987 balanced accuracy, and 0.830706 macro-F1, showing improved balanced detection with competitive macro-F1. In federated experiments, trust-aware filtering reduced the impact of malicious clients and substantially improved LogLoss and ECE, although macro-F1 gains were modest. Future work will consider stronger attacks, adaptive trust thresholds, secure aggregation, differential privacy, blockchain-based audit logging, real-time edge deployment, and validation on additional IoT, IIoT, healthcare, and smart-city datasets.

REFERENCES

[1] E. C. P. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu, and A. A. Ghorbani, "Ciciot2023: A real-time dataset and benchmark for large-scale attacks in iot environment," *Sensors*, vol. 23, no. 13, p. 5941, 2023.

[2] K. S. Adewole, A. Jacobsson, and P. Davidsson, "Intrusion detection framework for internet of things with rule induction for model explanation," *Sensors*, vol. 25, no. 6, p. 1845, 2025.

[3] B. M. Kouassi, A. B. Ballo, K. J. Ayikpa, D. Mamadou, and M. Z. J. Coulibaly, "Top-k feature selection for iot intrusion detection: Contributions of xgboost, lightgbm, and random forest," *Future Internet*, vol. 17, no. 11, p. 529, 2025.

[4] M. A. O. Ahmed, Y. Abdelsatar, R. Alotaibi, and O. Reyad, "Enhancing internet of things security using performance gradient boosting for network intrusion detection systems," *Alexandria Engineering Journal*, vol. 116, pp. 472–482, 2025.

[5] Y. Deng, "Design of industrial iot intrusion security detection system based on lightgbm feature algorithm and multi-layer perception network," *Journal of Cyber Security and Mobility*, vol. 13, no. 2, pp. 327–347, 2024.

[6] D. Zhang, D. Huang, Y. Chen, S. Lin, and C. Li, "A lightweight iot intrusion detection method based on two-stage feature selection and bayesian optimization." *AIMS Electronics & Electrical Engineering*, vol. 9, no. 3, 2025.

[7] A. Javed, A. Ehtsham, M. Jawad, M. N. Awais, A.-u.-H. Qureshi, and H. Larijani, "Implementation of lightweight machine learning-based intrusion detection system on iot devices of smart homes," *Future Internet*, vol. 16, no. 6, p. 200, 2024.

[8] M. M. Rashid, S. U. Khan, F. Eusufzai, M. A. Redwan, S. R. Sabuj, and M. Elsharief, "A federated learning-based approach for improving intrusion detection in industrial internet of things networks," *Network*, vol. 3, no. 1, pp. 158–179, 2023.

[9] R. Bensaid, N. Labraoui, A. A. A. Ari, H. Saidi, J. H. M. Emati, and L. Maglaras, "Sa-flids: secure and authenticated federated learning-based intelligent network intrusion detection system for smart healthcare," *PeerJ Computer Science*, vol. 10, p. e2414, 2024.

[10] 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 Networks*, vol. 162, p. 103540, 2024.

[11] R. Yang, H. He, Y. Wang, Y. Qu, and W. Zhang, "Dependable federated learning for iot intrusion detection against poisoning attacks," *Computers & Security*, vol. 132, p. 103381, 2023.

[12] Y. Sanjalawe, T. Alqurashi, Z. H. Alharbi, S. N. Makhadmeh, M. Alsharaiah *et al.*, "Adaptive graph attention-based federated learning for iot intrusion detection: mitigating poisoning attacks," *PeerJ Computer Science*, vol. 11, p. e3281, 2025.

[13] J. Wang, K. Yang, and M. Li, "Nids-fgpa: A federated learning network intrusion detection algorithm based on secure aggregation of gradient similarity models," *PloS one*, vol. 19, no. 10, p. e0308639, 2024.

[14] H. Q. Gheni and W. L. Al-Yaseen, "Two-step data clustering for improved intrusion detection system using ciciot2023 dataset," *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, vol. 9, p. 100673, 2024.

[15] M. Houichi, F. Jaidi, and A. Bouhoula, "Enhancing smart city security: An intrusion detection system using machine learning methods with the unb cic iot 2023 dataset," *IET Smart Cities*, vol. 7, no. 1, p. e70014, 2025.

[16] S. A. Almahaqeri, M. H. Almourish, A. A. Nasser, A. S. A. Alghawli, A. A. Elsayed, and A. N. Alhejoj, "An optimized gradient boosting framework for iot intrusion detection: a comprehensive evaluation on the ciciot2023 dataset," *Scientific Reports*, 2026.

[17] M. Nassef, "Boosting intrusion detection against ddos attacks using a feature engineering-based fine-tuned xgboost model," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 21, no. 1, pp. 1–39, 2025.

[18] S. Soliman, W. Oudah, and A. Aljuhani, "Deep learning-based intrusion detection approach for securing industrial internet of things," *Alexandria Engineering Journal*, vol. 81, pp. 371–383, 2023.

[19] A. A. Bhutta, M. u. Nisa, and A. N. Mian, "Lightweight real-time wifi-based intrusion detection system using lightgbm," *Wireless networks*, vol. 30, no. 2, pp. 749–761, 2024.

[20] S. T. Ahmed, A. S. Fathima, M. A. Halwani, A. Almusharraf, and A. Albuali, "A trust-centric federated edge learning paradigm in healthcare for decentralized threat intelligence sharing," *Transactions on Emerging Telecommunications Technologies*, vol. 37, no. 4, p. e70401, 2026.

[21] J. Talpini, N. Civiero, F. Sartori, and M. Savi, "A federated approach to enhance calibration of distributed ml-based intrusion detection systems." in *ICAART (2)*, 2025, pp. 840–848.