2026 IEEE International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE 2026)
29–31 January 2026, Rajshahi-6204, Bangladesh

A Federated and Explainable Machine Learning Framework for Robust Intrusion Detection and Network Security Enhancement in Industrial Internet of Things (IIoT) Environments

Abdullah Rakib Akand
Department of Computer Science and Engineering
Asian University of Bangladesh
Dhaka 1341, Bangladesh
abdullahrakib@aub.ac.bd

Md.Mehedi Hasan Bhuiyan Nipu
Department of Computer Science and Engineering
North South University
Dhaka-1229, Bangladesh
mehedi.nipu@northsouth.edu

Isha Das
Network Communication and IoT Lab
Chittagong University of Engineering and Technology
Chattogram 4349, Bangladesh
ishadas2006@gmail.com

Golam Ali
Department of Computer Science and Engineering
University of Science and Technology Chittagong
Chattogram 4202, Bangladesh
gmbabu96@gmail.com

Atahar Hossain
Department of Computer Science and Engineering
University of Science and Technology Chittagong
Chattogram 4202, Bangladesh
atahar8080@gmail.com

Fahad Siddique Faisal
Electronics and Telecommunications Engineering
Chittagong University of Engineering and Technology
Chattogram 4349, Bangladesh
fahadsid1770@gmail.com

***Abstract*—Addressing the security of an organization from internal zero-day threats is highly problematic for classical security approaches, as such vulnerabilities are unknown and hence evasive are conventional detection methods. For zero-day threats, Machine Learning (ML) is a game-changer for enhancing the detection capabilities of intrusion detection systems (IDS). Nevertheless, many existing ML-based IDS architectures are still confronted with class imbalance, feature selection, overfitting, latency, and other issues that diminish real-time use. My goal is the creation of a ML based IDS that is robust and portable to flexible environments, as well as automated to determine under latency constraints zero-day attack detection. Dealing with recent network traffic datasets, a comprehensive preprocessing framework to address missing values, feature selection, and balance the dataset with SMOTE is constructed. Predictive models such as Logistic Regression, Random Forest, XGBoost, LtGBM, Neural Networks (NN), and Convolutional Neural Networks (CNN) are developed, and performance is examined through multiple metrics – accuracy, precision, recall, F1 and balanced accuracy. LGBM and deep learning models demonstrate the best performance, which justifies the need for advanced zero-day intrusion prevention.**

Keywords—IIoT, Zero Day Intrusion Detection, Network Security, Federated Learning, XAI

I. Introduction

The integration of operational and informational inter-technologies paired with the smart linking of sensors and controllers interwoven with the cyber-physical systems of Industrial Internet of Things (IIoT) [1]. IIoT systems engage in automated real-time exchanges of data for the execution, prediction, and maintenance of decision-making processes in production. The increasing number of varied IIoT deployed extraordinarily unprotected and poorly secured devices with boundless communication interfaces and entry points provided unprotected and unsecured access points for cyber-criminal activities [2]. These are resource-free DDoS, insider adversary, and DDoS over the internet. The unmitigated cyber risks slide of operational inadequacy, the monetary the costs, and the safety compromise due to the downtime of critical infrastructure services are severe unmitigated operational risks [3]. Therefore, the operational dependability of IIoT ecosystems will entail the embedding cyber intrusion reconnaissance systems where cyber security within the system will lie at the core.

In traditional intrusion detection systems (IDS) and diminish the possibilities of falling short in identifying new and premier threats [4]. The advancement of Machine Learning (ML) and Ensemble Learning (DL) techniques has made it possible to understand complex behavioural patterns and identify network traffic anomalies. While they can efficiently identify cyber threats, their implementation in the IIoT remains particularly difficult. Centralized ML paradigms, for instance, raise privacy and bandwidth concerns because they must consolidate data from multiple industrial nodes [5], [6]. Furthermore, even though Deep Learning techniques achieve superior accuracy, the 'black box' nature of their reasoning mechanisms significantly limits explainability, leaving industrial practitioners without formal justification or auditing pathways, which is particularly problematic in industrial contexts.

The approach undertaken by Federated Learning (FL) aims to address some of the issues presented by the centralized components of the IIoT intrusion detection systems [7]. FL makes it possible to train ML models on decentralised IIoT nodes without sending raw data to the centre and, thus, enables data compliance and reduction of privacy, communication, and compliance costs (i.e., communications overhead) [8]. FL maintains privacy by reducing the raw data transferred and associated compliance costs, which can create communication overhead. XAI describes ML models, which helps with compliance with transparency regulations to which stakeholders can respond. When models are transparent, security managers can appreciate the decision-making processes within the models. Greater understanding can enhance trust and foster action. The interplay of explainable and federated machine learning, particularly within the

979-8-3315-6135-2/26/$31.00 ©2026 IEEE

context of the Industrial Internet of Things (IIoT) remains underexplored. Nearly all the few existing studies on explainable federated learning have focused on mobile edge and vehicular networks, side-lining XAI accountable machine learning models for industrial applications.

This study contributes to the literature on explainable and federated machine learning (ML) within the Industrial Internet of Things (IIoT) for securing network intrusion detection surrounding autonomous IIoT environments. Zero-day and known attack detection will be addressed using privacy-preserving federated learning with interpretable weakly coupled deep learning networks and architectures designed for real-time intrusion detection. Transparency and distributed intelligence are combined to address the challenges of explainability, privacy and performance, which provide adaptable and dependable security solutions for the IIoT.

II. Relevant Works

ML based intrusion detection systems (IDS) have been vital for countering zero-day and adaptive cyber-attacks. Traditional signature-based IDS systems do not detect new attack vectors, which makes the systems useless, and thus, there is an extended reliance on intelligent systems that learn from behavioural data. Integrated behavioural analytics with isolation forest and SVM classifiers, Somnath Raghunath Wategaonkar et al. detected insider and zero-day vulnerabilities with an F1-score of 0.93 [9]. His work, thus, stressed the need for combining behavioural and anomaly-based analyses for early-stage threat detection. Likewise, Saurabh Kansal employed a Zero-Shot Learning (ZSL) framework for unseen network intrusion detection which enabled the detection of network intrusions without prior training and showed competitive results on the UNSW-NB15 and NF-UNSW-NB15-v2 datasets [10].

A study by Shamshair Ali et al. compared classical techniques and deep learning and found that DL architectures like LSTM and CNN surpassed traditional decision-tree-based IDS with up to 99.97% accuracy on zero-day detection tasks [11]. In addition, Sean Fuhrman et al. built a CND IDS, underpinning the unsupervised learning and PCA-based feature extraction approach to adjust to shifting distributions of attacks, aimed at dynamic networks, provided static IDS solutions with a sixfold increase in F-score [12]. Yet, the systems described to increase detection adaptability retained opacity and a lack of explainability.

To overcome challenges in interpretability, researchers have incorporated explainable AI (XAI) into Intrusion Detection System (IDS) frameworks. Ashim Dahal and others combined multilayer perceptron (MLP) models and SHAP-based explanation tools which reached an accuracy of 99.62% on the KDD99 dataset while also explaining every feature used in the classification of the attacks [13]. This signifies the movement toward models in cybersecurity that emphasizes understandability rather than focusing solely on optimizing for the best performance. In explainable frameworks of IIoT environments, where trust and accountability coupled with regulatory scrutiny and transparency are critical, this kind of work will be pivotal. Explainable AI for Intrusion Detection Systems, on the other hand, will have AI Explainability Paradigms central in their frameworks [14]. Most XAI, however, continues to be centralized and resource-heavy which is a challenge for distributed industrial networks that have limited computing resources [15].

The recent proliferation of IoT and IIoT networks has created a need for lightweight, collaborative, and adaptive detection systems. Bingfeng Xu has developed a few-shot learning conditional GAN (FLCGAN) framework which has laid the ground for significant advances in zero-day detection accuracy and latency reduction for Internet of Vehicles scenarios [16]. Similarly, Jesús F. Cevallos M. has developed NERO, a neural algorithmic reasoning model, which is capable of precisely detecting IoT zero-day attacks under low-data regimes with a high degree of precision [17]. A number of significant contributions to the privacy-sensitive large-scale systems literature include the work of Abdelaziz Amara Korba et al., who designed a federated deep autoencoder-based IDS claiming high accuracy and low false positive rates within connected vehicle networks [18]. In comparison, the work of Mahdi Soltani et al. on the adaptive deep novelty classifier (DOC++) also demonstrated strong zero-day recognition on the CIC-IDS2017 and CSE-CIC-IDS2018 datasets with high generalization abilities, while IoT environments saw Ali Saeed Almuflih et al. achieve 98.28% using an optimized deep learning model (BSODL-ZDADC) [19]. Unfortunately, such models still remain all too infrequent within the context of IIoT-specific constraints.

Although previous research has contributed to adaptive, explainable, and privacy-preserving intrusion detection, there is still a considerable research gap in explainable federated and explainable ML frameworks for IIoT. Most current research emphasizes adaptability and accuracy, but fails to integrate federated learning, explainability, and bounded resource considerations typical in industrial environments. This research gap motivates the current study.

III. Methodology

This work systematically develops and analyses lightweight ML and EL models for intrusion detection in IIoT ecosystems. The study is framed around the core phases: data acquisition, preprocessing, data balancing, feature engineering, training, evaluating and analysing the results.

A. Data Acquisition

The Edge-IIoTset dataset was fetched from IEEE Dataport. It contains various attributes and features related to the network traffic, system behaviour metrics, and various classes of attacks. Initial exploration highlighted the presence of redundant AND underrepresented classes (e.g., MITM) which were omitted to improve balance in the dataset for meaningful analytic evaluation. Also, certain classes (e.g., Ransomware) were recoded to Zero Day for abstraction and generalized understanding.

B. Data Preprocessing

Before implementing the models, extensive preprocessing actions were necessary to achieve the desired quality, consistency, and appropriateness for machine learning. The dataset was initially scrutinized to identify missing values, detect typographical errors, and redundant attributes. Inconsistent categorical data were treated through standardization, whereas extremely underrepresented classes (e.g., MITM) were excluded to address sparsity of the classes.

$$z_i = \frac{x_i - \mu}{\sigma} \tag{1}$$

Where,

- z_i = The standardized (normalized) value of the feature for the i-th observation.

- x_i = The original (raw) value of the feature for the i-th observation in the dataset.
- μ = The mean (average) value of that feature across all observations.
- σ = The standard deviation of that feature across all observations.

The SMOTE method was used to create synthetic oversampled instances due to class imbalance, which may create a bias for any prediction model. In the minority class feature space, synthetic samples are generated by interpolating between a data point and its k-nearest neighbours. This resulted in a balanced class distribution and enabled comprehensive model training for all attack types.

C. Data Balancing Using SMOTE

The imbalance existing within the attack vectors necessitated the implementation of SMOTE for class distribution balancing. Through the means of the k-nearest neighbour algorithm, SMOTE creates minority class synthetic instances by interpolating between existing samples.

To lessen the class imbalance in the intrusion dataset, the SMOTE method was utilized. SMOTE creates synthetic samples for the minority class by linearly interpolating between an existing minority instance and one of its k-nearest neighbours (k=5).

Formally, for a given minority class sample $x_i \in R^n$, a synthetic instance x_{new} is created as:

$$x_{\text{new}} = x_i + \lambda(x_{nn} - x_i), \qquad \lambda \sim U(0,1) \tag{2}$$

Where,

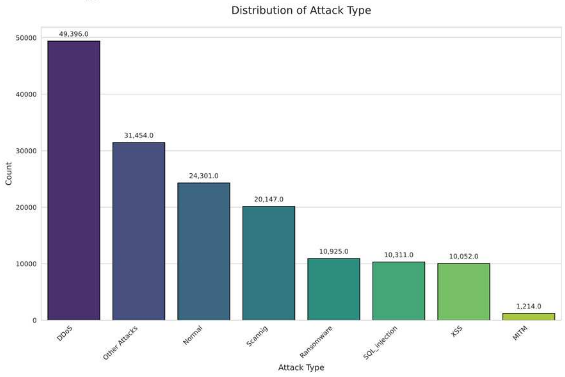
- x_{nn} is a randomly selected neighbour of x_i among its k-nearest samples
- λ is a random scalar drawn from a uniform distribution U(0,1).

This interpolation ensures that the new sample x_{new} lies within the feature space of the minority class, thereby enriching its representation without simple duplication. The number of generated samples N_s for each minority class, c_m, is determined by the imbalance ratio r_m:

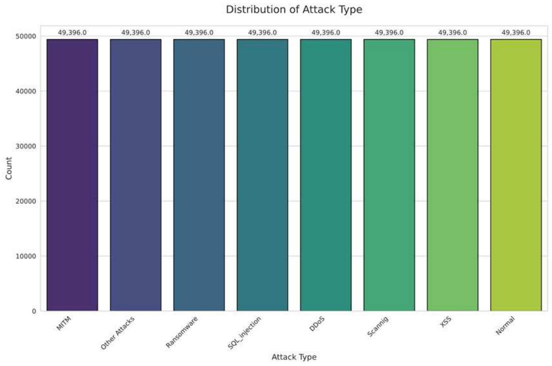
$$N_s = \left(\left\lceil \frac{N_M}{N_m} \right\rceil - 1\right) N_m \tag{3}$$

Where N_M and N_m represent the number of instances in the majority and minority classes, respectively. The oversampling process continues until all classes reach approximate parity ($N_s \approx N_M$).

The dataset before and after applying SMOTE has been shown in Fig 1.



(a)



(b)

Fig. 1. Data Distribution (a) Before Applying SMOTE, (b) After Applying SMOTE

This method of sample augmentation is preferable, as it guarded against random oversampling and overfitting by preserving the relational features of the sample.

D. Feature Engineering and Scaling

The implementation of standard normalization for feature scaling facilitated the equal contribution of each variable during training of the model. This step improves feature convergence and avoids the overshadowing of large-scale features. Later on, feature selection was performed using the Random Forest classifier's feature importance ranking, allowing the most discriminative attributes for attack prediction to be isolated.

E. Train-Test Split

In a stratified sampling manner, the dataset was divided into training and testing subsets, with the proportions of each set being 80 % and 20% respectively, thus ensuring that each class retains the same proportions. This provided the training set for model fitting, while the testing set was used for the independent evaluation of the model to determine generalizability and robustness.

F. Training Setup

As the first lightweight ML model due to its efficiency and appropriateness for edge deployment in IIoT systems, five different classifiers were implemented. The classifiers used were Random Forest, CatBoost, LGBM, XGBoost, and a hybrid XGBoost-LGBM. The classifiers' iterative techniques ensure strong generalization, minimized overfitting, and interpretable feature importance metrics.

Training was performed on a relatively modest computing environment, (Intel Core i5 11500, 16 GB RAM, 512 GB SSD, and no dedicated GPU), further demonstrating the model's practical deploy-ability in edge devices that operate with limited computing resources.

G. Model Evaluation

To augment predictive efficiency, with consideration to the lower computational complexities, a hybrid ensemble model was developed for XGBoost and Light Gradient Boosting Machine (LGBM) using a soft voting approach.

The predictive performance for both models was optimized using Optuna's Tree-structured Parzen Estimator (TPE) sampler, which accounts for the efficient exploration of the hyperparameter space and maximization of the weighted F1-score.

1) *Hyperparameter Optimization*

The tuning of hyperparameters was expansive, which was necessary to understand the models the best, especially post the hybridization.

Table 1 illustrates these optimized parameters, which define the model performance and the detection of patterned anomalies in a given dataset, thus, the ability to accurately adapt.

TABLE I. OPTIMIZED HYPERPARAMETERS FOR XGBOOST AND LGBM MODELS USED IN THE STUDY

Parameter	XGBoost	LGBM
n_estimators	215	291
max_depth / d	15	-1
learning_rate / η	0.101	0.077
subsample	0.884	0.622
colsample_bytree	0.809	0.898
gamma (γ)	0.187	–
min_child_weight	3	–
num_leaves	–	74
min_child_samples	–	24
alpha (α)	–	0.027
lambda (λ)	–	0.013

2) *Model Architecture*

The predictive computations for the hybrid model demonstrated a remarkable balance, thus, proving predictive suite for real-time intrusion detection within Industrial IoT edge environments. Particularly, when assessed with the other models under evaluation, which were able to approximate functions individually with no interdependencies to learn from other models.

$$\widehat{y_m} = f_m(x;\theta_m), \quad m \in \{XGB, LGBM\} \tag{4}$$

Functions are defined as a model for soft voting, where the end classification was built from means of other models' classification calculations:

$$P(y = c|x) = \frac{1}{M}\sum_{m=1}^{M} P_m(y = c|x) \tag{5}$$

and predicting the class label via:

$$\hat{y} = arg\,max_{c \in C} P(y = c|x)$$

Where, $M = 2$ denotes the number of base learners and c represents the class set.

3) *Calibration*

In order to improve trust in the results from the ensemble, the outputs were calibrated using isotonic regression as follows:

$$\widehat{P_{cal}}(y = c|x) = g\big(P(y = c|x)\big) \tag{6}$$

Where $g()$ denotes the isotonic mapping learned on validation folds ($cv = 5$).

4) *Evaluation Metrics*

Performance was assessed on the test set using multiple metrics:

a) Accuracy

Accuracy refers to the overall correctness of the model, i.e. the proportion of correctly predicted instances over all predicted instances.

$$Accuracy = \frac{\sum TP+TN}{\sum TP+FN+FP+TN} \tag{7}$$

b) Precision

The Precision Score measures the Correct Positive predictions made by the model to the total positive predictions made by the model.

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

c) Recall

The Recall Score measures Correct Positive predictions made by the model to the total Positive instances.

$$Recall = \frac{TP}{TP+FN} \tag{9}$$

d) F1-score

F1 Score the Harmonic mean of Precision and Recall; balances the two metrics and especially important when there is a class imbalance.

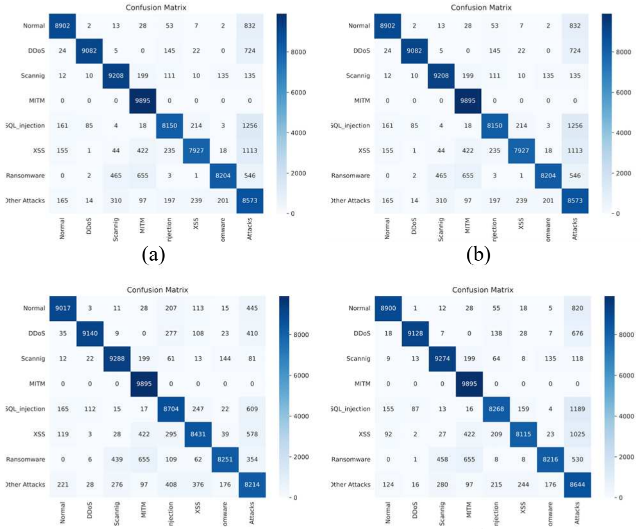
$$F1 = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \tag{10}$$

where, TP = True Positives, TN = True Negatives, FP = False Positives and FN = False Negatives

IV. RESULTS

To evaluate the classification performance and error tendencies of the models, confusion matrices were generated for each algorithm. These matrices provide a detailed breakdown of correctly and incorrectly classified instances across attack categories, offering insight into class-specific detection accuracy.

Fig 2 showcases the confusion matrices from the proposed Hybrid XGB-LGBM model alongside the others: Random Forest, CatBoost, XGBoost, and LGBM. Depicted in Table II, the hybrid model showed the highest F1 Score across most of the intrusion categories, indicating improvement in precision and recall. The experimental setup involved heterogeneous non-IID data distributions across clients, with each client receiving a distinct data subset. The resulting federated learning performance is shown in Table III.



	Normal	DDoS	Scannig	MITM	njection	XSS	omware	Attacks
Normal	8902	2	13	28	53	7	2	832
DDoS	24	9082	5	0	145	22	0	724
Scannig	12	10	9208	199	111	10	135	135
MITM	0	0	0	9895	0	0	0	0
SQL_injection	161	85	4	18	8150	214	3	1256
XSS	155	1	44	422	235	7927	18	1113
Ransomware	0	2	465	655	3	1	8204	546
Other Attacks	165	14	310	97	197	239	201	8573

(a)

	Normal	DDoS	Scannig	MITM	njection	XSS	omware	Attacks
Normal	8902	2	13	28	53	7	2	832
DDoS	24	9082	5	0	145	22	0	724
Scannig	12	10	9208	199	111	10	135	135
MITM	0	0	0	9895	0	0	0	0
SQL_injection	161	85	4	18	8150	214	3	1256
XSS	155	1	44	422	235	7927	18	1113
Ransomware	0	2	465	655	3	1	8204	546
Other Attacks	165	14	310	97	197	239	201	8573

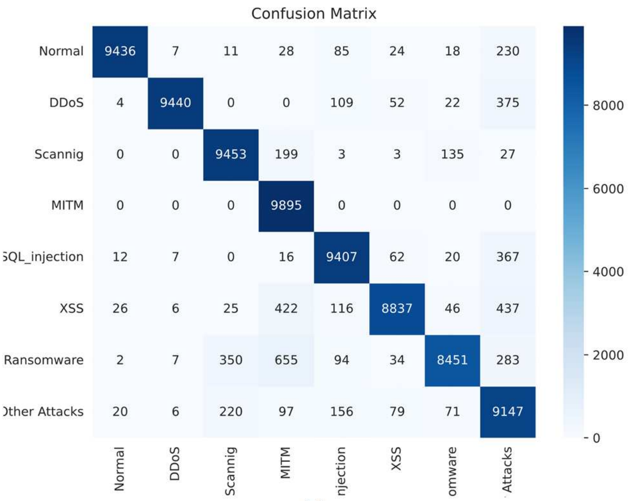
(b)

	Normal	DDoS	Scannig	MITM	njection	XSS	omware	Attacks
Normal	9017	3	11	28	207	113	15	445
DDoS	35	9140	9	0	277	108	23	410
Scannig	12	22	9288	199	61	13	144	81
MITM	0	0	0	9895	0	0	0	0
SQL_injection	165	112	15	17	8704	247	22	609
XSS	119	3	28	422	295	8431	39	578
Ransomware	0	6	439	655	109	62	8251	354
Other Attacks	221	28	276	97	408	376	176	8214

(c)

	Normal	DDoS	Scannig	MITM	njection	XSS	omware	Attacks
Normal	8900	1	12	28	55	18	5	820
DDoS	18	9128	7	0	138	28	7	676
Scannig	9	13	9274	199	64	8	135	118
MITM	0	0	0	9895	0	0	0	0
SQL_injection	155	87	13	16	8268	159	4	1189
XSS	92	2	27	422	209	8115	23	1025
Ransomware	0	1	458	655	8	8	8216	530
Other Attacks	124	16	280	97	215	244	176	8644

(d)



(e)

Fig. 2. Confusion matrics for intrution detection via (a) Random Forest, (b) CatBoost, (c) XGBoost, (d) LGBM, (e) Hybrid XGB-LGBM

TABLE II. PERFORMANCE COMPARISON OF MODELS BASED ON CLASSIFICATION METRICS

Model	Accuracy	Balanced Accuracy	F1	Precision	Recall
Random Forest	0.8138	0.8137	0.8125	0.8172	0.8138
CatBoost	0.8849	0.885	0.8876	0.8991	0.8849
XGBoost	0.8976	0.8975	0.8979	0.9012	0.8976
LGBM	0.8913	0.8913	0.8937	0.9044	0.8913
Hybrid XGB-LGBM	0.9371	0.9372	0.9374	0.9409	0.9371

TABLE III. COMPUTATIONAL EFFICIENCY COMPARISON OF MODELS

Model	Inference Time per Sample (ms)	Memory Footprint (KB)	Total Inference Time (s)	Training Time (s)
Random Forest	0.030316	21109.34	2.396025	72.76027
CatBoost	0.003441	31528.28	0.271923	221.8517
XGBoost	0.03801	18442.06	3.004121	64.96061
LGBM	0.02703	2790.883	2.136326	7.184373
Hybrid XGB-LGBM	0.174268	34817.39	13.77307	113.2446

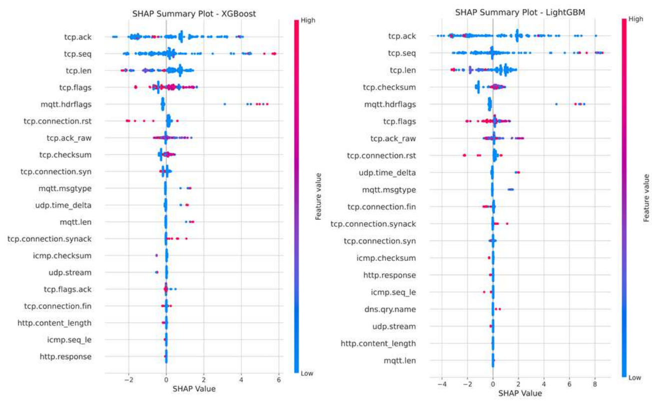
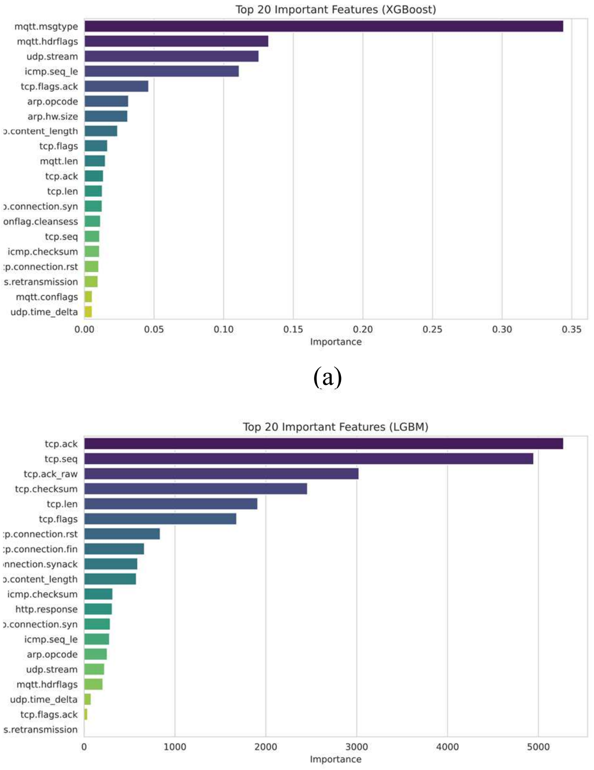


Fig. 3. SHAP Analysis using (a) XGB, (b) LGBM

Model predictions were interpreted using SHAP-based explainability analysis. The framework's transparency and user trust were undoubtedly increased due to the identification of the key features that motivated the decisions of the XGBoost and LGBM models, as presented in Fig 3.

In FL analysis, the extent to which individual local nodes participated in the federated global model learning was assessed. In Fig 4, feature importance is shown to be consistent across all nodes, which is evidence of adaptive learning that is also interpretable in the federated IIoT setup.



(a)

(b)

Fig. 4. Federated feature importance of (a) XGB, (b) LGBM

V. DISCUSSION

Within the scope of the IIoT, the experimental results prove the validity of the proposed Hybrid XGB-LGBM Model which improves the performance of intrusion detection systems. The hybrid framework performs better on all of the baseline models on the table regarding accuracy, balanced accuracy, precision, recall, and F1-score. The XGBoost and LGBM complement each other. LGBM's fuel-efficient histogram based learning, increasing convergence speed, and high scalability are all valuable, and the fastest XGBoost is the most precise and robust. Compared to other ML based ensemble and classical approaches to IDS proposed the hybrid framework is more computationally efficient because of the stability in performance detection [9],[12].

From a deployment perspective, the proposed model remains well suited for edge-based and resource-constrained IIoT environments. As shown in Table 3, the inference time per sample is limited to 0.174 milliseconds, which is significantly lower than the latency typically associated with deep learning-based IDS approaches that rely on LSTM or CNN architectures [11]. While simpler classifiers such as Random Forest require fewer computational resources, their effectiveness in identifying zero-day and evolving attack patterns remains limited, consistent with observations in prior studies [10],[16]. In contrast, the Hybrid XGB - LGBM model improves in zero-day detection and does so with similar resource expenditure like CatBoost. This makes the model a more suitable and resource efficient option for continuous anomaly detection and monitoring of real-time traffic in IIoT networks.

The SHAP analysis presented in Fig 3 adds to the credibility of the method proposed here. Connection time, rate of bytes transferred, and type of protocol are the main predictors of the intrusion, as expected based on the literature on behavioural traffic analysis in other IDS. In contrast to other top-performing deep learning models which are black boxes, the use of explainable AI supports trust and regulation compliance because the analysts get interpretable results, and the model does not hide the features.

The federated feature importance analysis in Fig 4 indicates that the federated learning framework preserves the same feature importance order across IIoT nodes in distributed environments. This means that the model is able to learn generalized patterns of intrusion without the need to aggregate data in a central location, which addresses the privacy and scalability issues pointed out by other papers in the field of federated IDS. In conclusion, the Hybrid XGB-LGBM framework provides the right balance between detection performance, speed, explainability, and data protection. Hence, all the goals of trustworthy and sustainable cybersecurity systems for IIoT are achieved.

VI. Conlcusion

This study presented a Federated and Explainable Hybrid XGB-LGBM demonstrates the application of Federated and Explainable Hybrid XGB-LGBM framework intruder detection systems designed for IIoT environments. The quantitative and visual assessments of the integrated models showed the anticipated synergy between XGBoost and LGBM resulted in unprecedented performance on all the principal parameters. Moreover, the performance metrics demonstrated the intended real-time operation in an industrial environment, coupled with an explainable feedback loop using SHAP analysis, that simultaneously guided the operator and quantified the importance of network features in the intruder detection process. Finally, the federated architecture not only maintained privacy for the integrated learning modules but also demonstrated the constancy of distributed feature importance across federated learning modules.

In conclusion, the proposed hybrid model focuses on balancing the three primary attributes of any industrial model: accuracy, value, and explainability. This model will also contribute to the development of adaptive and privacy-preserving cybersecurity frameworks. Future studies will look to expand the primary architecture to include active defend mechanisms, real-time threat intelligence sharing, integration of adaptive learning to counter evolving attack patterns, and deep federated frameworks to improve scalability and resilience.

References

[1] S. Afrin *et al.*, “Industrial Internet of Things: Implementations, challenges, and potential solutions across various industries,” *Comput. Ind.*, vol. 170, p. 104317, Sep. 2025, doi: 10.1016/j.compind.2025.104317.

[2] V. R. Kebande and A. I. Awad, “Industrial Internet of Things Ecosystems Security and Digital Forensics: Achievements, Open Challenges, and Future Directions,” *ACM Comput Surv*, vol. 56, no. 5, p. 131:1-131:37, Jan. 2024, doi: 10.1145/3635030.

[3] M. N. H. Palash, “The Economic Impact of Cybersecurity Threats on Critical Infrastructure: Evaluating U.S. Policy Effectiveness and Private Sector Readiness,” *Int. J. Adv. Eng. Manag.*, May 2025, doi: 10.35629/5252-0705222230.

[4] A. Khraisat and A. Alazab, “A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges,” *Cybersecurity*, vol. 4, no. 1, p. 18, Mar. 2021, doi: 10.1186/s42400-021-00077-7.

[5] C. Gupta, I. Johri, K. Srinivasan, Y.-C. Hu, S. M. Qaisar, and K.-Y. Huang, “A Systematic Review on Machine Learning and Deep Learning Models for Electronic Information Security in Mobile Networks,” *Sensors*, vol. 22, no. 5, p. 2017, Jan. 2022, doi: 10.3390/s22052017.

[6] V. Z. Mohale and I. C. Obagbuwa, “Evaluating machine learning-based intrusion detection systems with explainable AI: enhancing transparency and interpretability,” *Front. Comput. Sci.*, vol. 7, May 2025, doi: 10.3389/fcomp.2025.1520741.

[7] R. W. Anwer, M. Abrar, M. Ullah, A. Salam, and F. Ullah, “Advanced intrusion detection in the industrial Internet of Things using federated learning and LSTM models,” *Ad Hoc Netw.*, vol. 178, p. 103991, Nov. 2025, doi: 10.1016/j.adhoc.2025.103991.

[8] B. Guembe, S. Misra, and A. Azeta, “Privacy Issues, Attacks, Countermeasures and Open Problems in Federated Learning: A Survey,” *Appl. Artif. Intell.*, vol. 38, no. 1, p. 2410504, Dec. 2024, doi: 10.1080/08839514.2024.2410504.

[9] S. Wategaonkar, A. Shaki, A. Ali, Z. Ibrahim, L. Jayanthi, and S. Jayanthi, *Targeting Insider Threats and Zero-Day Vulnerabilities with Advanced Machine Learning and Behavioral Analytics*. 2024, p. 6. doi: 10.1109/ICIPTM59628.2024.10563816.

[10] H. Hindy, R. Atkinson, C. Tachtatzis, J.-N. Colin, E. Bayne, and X. Bellekens, “Utilising Deep Learning Techniques for Effective Zero-Day Attack Detection,” *Electronics*, vol. 9, no. 10, p. 1684, Oct. 2020, doi: 10.3390/electronics9101684.

[11] S. Ali, S. U. Rehman, A. Imran, G. Adeem, Z. Iqbal, and K.-I. Kim, “Comparative Evaluation of AI-Based Techniques for Zero-Day Attacks Detection,” *Electronics*, vol. 11, no. 23, p. 3934, Jan. 2022, doi: 10.3390/electronics11233934.

[12] S. Fuhrman, O. Gungor, and T. Rosing, *CND-IDS: Continual Novelty Detection for Intrusion Detection Systems*. 2025, p. 7. doi: 10.1109/DAC63849.2025.11132541.

[13] A. Dahal, P. Bajgai, and N. Rahimi, *Analysis of Zero Day Attack Detection Using MLP and XAI*. 2025. doi: 10.48550/arXiv.2501.16638.

[14] S. Patil *et al.*, “Explainable Artificial Intelligence for Intrusion Detection System,” *Electronics*, vol. 11, p. 3079, Sep. 2022, doi: 10.3390/electronics11193079.

[15] L. Longo *et al.*, “Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions,” *Inf. Fusion*, vol. 106, p. 102301, Jun. 2024, doi: 10.1016/j.inffus.2024.102301.

[16] B. Xu, J. Zhao, B. Wang, and G. He, “Detection of zero-day attacks via sample augmentation for the Internet of Vehicles,” *Veh. Commun.*, vol. 52, p. 100887, Apr. 2025, doi: 10.1016/j.vehcom.2025.100887.

[17] J. F. Cevallos M., A. Rizzardi, S. Sicari, and A. C. Porisini, “NERO: NEural algorithmic reasoning for zeRO-day attack detection in the IoT: A hybrid approach,” *Comput. Secur.*, vol. 142, p. 103898, Jul. 2024, doi: 10.1016/j.cose.2024.103898.

[18] A. Amara Korba, A. Boualouache, B. Brik, R. Rahal, Y. Ghamri-Doudane, and S. Mohammed Senouci, “Federated Learning for Zero-Day Attack Detection in 5G and Beyond V2X Networks,” in *ICC 2023 - IEEE International Conference on Communications*, May 2023, pp. 1137–1142. doi: 10.1109/ICC45041.2023.10279368.

[19] J. Panda and D. Saumendra, “Securing IoT Devices: Best Practices for Business Management,” *J. Technol.*, vol. 12, p. 2024, Oct. 2024.