Journal of Current Science and Technology, January-March 2026 Vol. 16 No. 1, Article 151

ISSN 2630-0656 (Online)

Cite this article: Potharaju, S., Tambe, S. T., Tirandasu, R. K., Kumar, D. A., Kantipudi, Kantipudi, P., & Shantamallappa, K. (2026). Enhancing cybersecurity in industrial internet of things systems using ensemble learning against false data injection Attacks. *Journal of Current Science and Technology*, *16*(1), Article 151. https://doi.org/10.59796/jcst.V16N1.2026.151

Journal of Current Science and Technology

Journal homepage: https://jcst.rsu.ac.th

Enhancing Cybersecurity in Industrial Internet of Things Systems Using Ensemble Learning Against False Data Injection Attacks

Saiprasad Potharaju[1, *], Swapnali N Tambe [2], Ravi Kumar Tirandasu[3], Dudla Anil Kumar[4], MVV Prasad Kantipudi[5], and Shantamallappa K[6]

[1]Department of CSE, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India
[2]Department of Information Technology, K. K. Wagh Institute of Engineering Education & Research, Nashik, India
[3]Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India
[4]Department of CSE, Lakireddy Bali Reddy College of Engineering, Mylavaram, India
[5]Department of Electronics and Telecommunication, Symbiosis Institute of Technology, Pune Campus, Symbiosis International (Deemed University), Pune, India
[6]Civil Engineering Department, Indus Institute of Engineering & Technology, Indus University, Ahmedabad, India

*Corresponding author; E-mail: psaiprasadcse@gmail.com

Received 15 July 2025; Revised 18 August 2025; Accepted 18 August 2025; Published online 20 December 2025

Abstract

False Data Injection Attacks (FDIAs) pose critical threats to Industrial Internet of Things (IIoT) systems by manipulating sensor data to cause operational disruptions and safety hazards. Traditional intrusion detection systems struggle to identify the subtle anomalies characteristic of FDIAs, necessitating advanced machine learning approaches. This study develops a weighted voting ensemble framework integrating Random Forest, XGBoost, Neural Network, and Logistic Regression with F1-score-based dynamic weight assignment for optimized FDIA detection. The proposed ensemble was evaluated on the UKMNCT_IIoT_FDIA dataset containing 15,425 instances across 30 features. Using 70–30 train–test split, model performance was assessed through accuracy, precision, recall, F1-score, and confusion matrix analysis. Results demonstrate exceptional performance: 99.71% accuracy, 99.72% precision, 99.72% recall, and 99.72% F1-score. Confusion matrix analysis revealed only 2 false negatives and 9 false positives across 4,627 test instances, substantially outperforming individual classifiers while maintaining computational efficiency suitable for resource-constrained edge devices.

The weighted voting mechanism successfully leverages algorithmic diversity to achieve superior robustness compared to individual models. Tree-based ensembles (Random Forest: 99.74%, XGBoost: 99.68%) substantially outperformed Neural Network (87.14%) and Logistic Regression (83.32%), confirming the importance of non-linear modeling for complex attack pattern detection. The minimal false negative rate (0.04%) represents critical advancement for critical infrastructure protection where undetected attacks carry severe consequences. This research establishes the efficacy of performance-adaptive ensemble learning for IIoT cybersecurity, providing a practical, scalable solution for safeguarding industrial cyber-physical systems against evolving threats.

Keywords: *industrial internet of things; false data injection attack; ensemble learning; weighted voting; intrusion detection; cybersecurity; random forest; XGBoost*

1. Introduction

The Industrial Internet of Things (IIoT) represents a transformative paradigm that integrates intelligent sensors, cyber-physical systems, and advanced communication protocols to revolutionize industrial operations across manufacturing, energy, healthcare, digital entertainment industry, and transportation sectors (Javaid et al., 2021; Mulla et al., 2025). By enabling real-time data exchange, automated control, and enhanced decision-making, IIoT systems have become fundamental to modern industrial infrastructure. However, the increasing connectivity of these systems to internet and cloud platforms has exposed them to sophisticated cyber threats, creating substantial risks to critical infrastructure and sensitive operational data (Yu et al., 2021).

Unlike traditional information technology environments that rely on discrete computing systems, IIoT ecosystems comprise interconnected physical devices managing essential operations such as power grids, manufacturing plants, and healthcare equipment (Potharaju et al., 2025; Simmachan et al., 2025). Security breaches in these systems can result in operational disruptions, significant financial losses, threats to human safety, and potential national security vulnerabilities (Eyeleko et al., 2023). The unique characteristics of IIoT environments including resource constraints, heterogeneous device architecture, distributed deployment, and real-time operational requirements render conventional cybersecurity mechanisms inadequate for comprehensive protection (Sengupta et al., 2020a).

Among various cyber threats targeting IIoT systems, False Data Injection Attacks (FDIAs) pose particularly insidious risks by manipulating sensor readings or injecting fabricated data into operational systems, thereby causing misinterpretations of system states (Tian et al., 2022; Maheshwar & Veenadhari, 2023). Unlike conventional cyberattacks that exploit software or hardware vulnerabilities, FDIAs compromise the integrity of sensor-generated data upon which IIoT systems fundamentally rely for critical decision-making (Pannakkong & Kanjanarut, 2023). The stealthy nature of FDIAs, which often mimic legitimate data streams, combined with their potential to cause physical damage, service disruptions, and safety hazards, necessitates sophisticated detection mechanisms (Li et al., 2025). Traditional intrusion detection systems frequently fail to identify the subtle anomalies characteristic of FDIAs, particularly in resource-constrained environments where computational limitations prevent deployment of complex security solutions (Ahmad et al., 2024).

Machine learning approaches, particularly ensemble learning techniques, have emerged as promising solutions for enhancing FDIA detection capabilities by leveraging the complementary strengths of multiple predictive models (Ganaie et al., 2022; Jagtap et al., 2025). Ensemble methods demonstrate superior performance in anomaly detection through model diversity, improved accuracy, adaptive learning, and error mitigation (Wu et al., 2021; De Zarzà et al., 2023). Recent research has demonstrated the efficacy of ensemble learning across various cybersecurity applications (Inma et al., 2025). Hu et al. (2024) achieved 99.99% accuracy using Random Forest combined with Bat Algorithm-based feature selection for IIoT intrusion detection. Gaber et al. (2023) addressed data imbalance challenges using XGBoost with Recursive Feature Elimination and Binary Grey Wolf Optimization. Awotunde et al. (2021) reported 99% accuracy employing deep feedback on neural networks with rule-based feature selection on benchmark datasets.

Advanced ensemble architectures have shown particular promise in IIoT security contexts. Ruiz-Villafranca et al. (2024) demonstrated that TabPFN models outperformed traditional ensemble methods including Random Forest, XGBoost, and LightGBM in multi-class intrusion detection. Jemili et al. (2024) integrated Random Forest and XGBoost with Apache Spark for scalable big-data cybersecurity applications (Simmachan & Boonkrong, 2025). Several studies have successfully applied ensemble learning to specialized domains, including automotive IoT systems (Dakic et al., 2024), encrypted traffic detection (Aouedi et al., 2022), and distributed denial-of-service attack identification (Karamti et al., 2023). Furthermore, ensemble techniques incorporating explainable AI components have enhanced model interpretability in sensitive applications (Sengupta et al., 2020b; Laftah et al., 2024).

Despite these advances, existing ensemble approaches for FDIA detection in IIoT environments exhibit several limitations. First, many studies focus on general intrusion detection without specifically addressing the unique characteristics of FDIAs, which require specialized detection mechanisms to identify subtle data manipulation rather than obvious network intrusions (Thongpance et al., 2023). Second, ensemble models often employ fixed voting schemes that fail to account for varying classifier performance across different attack scenarios, potentially limiting

detection accuracy when individual models contribute suboptimal. Third, limited research has systematically compared weighted voting strategies against conventional ensemble methods specifically for FDIA detection in resource-constrained IIoT environments. Finally, there remains a need for ensemble models that balance detection accuracy with computational efficiency suitable for deployment on edge devices with limited processing capabilities.

This study addresses these gaps by proposing a weighted voting ensemble model specifically designed for FDIA detection in IIoT systems. The proposed approach combines Random Forest, XGBoost, Neural Networks, and Logistic Regression classifiers, with weights dynamically assigned based on individual classifier accuracies to optimize collective performance. By leveraging the complementary strengths of diverse machine learning algorithms while accounting for their relative effectiveness, this model aims to achieve superior FDIA detection accuracy compared to both individual classifiers and conventional ensemble methods. The model is evaluated using a publicly available FDIA dataset and benchmarked against baseline classifiers to demonstrate its effectiveness, scalability, and adaptability for deployment in real-world IIoT environments. This research contributes to IIoT cybersecurity by providing a practical, high-performing solution for protecting critical industrial infrastructure against evolving FDIA threats.

2. Objectives

This study aims to develop and evaluate a robust ensemble learning approach for detecting False Data Injection Attacks in IIoT systems. The specific objectives are:

1. To analyze the limitations of individual machine learning classifiers (Random Forest, XGBoost, Neural Networks, Logistic Regression) for FDIA detection in IIoT environments.
2. To develop a weighted voting ensemble model that dynamically assigns weights based on individual classifier performance metrics.
3. To evaluate the proposed ensemble model using a publicly available FDIA dataset and benchmark its performance against standalone classifiers and conventional ensemble methods.
4. To assess the model's scalability, computational efficiency, and adaptability for deployment in resource-constrained IIoT contexts.
5. To demonstrate the practical applicability of the proposed approach for enhancing cybersecurity resilience in critical industrial infrastructure.

3. Materials and Methods

This section presents the comprehensive methodology employed to develop and evaluate the proposed weighted voting ensemble model for FDIA detection in IIoT systems. The experimental framework, illustrated in Figure 1, comprises three interconnected phases: (I) data acquisition and preprocessing, (II) model selection and training, and (III) performance evaluation. The subsequent subsections provide detailed descriptions of each phase, including dataset characteristics, individual classifier configurations, ensemble architecture, and evaluation metrics.

3.1 Experimental Framework

Figure 1 presents the overall experimental workflow adopted in this study. The framework begins with data preprocessing to ensure quality and consistency of the input dataset. Subsequently, four diverse machine learning classifiers Random Forest, XGBoost, Neural Network, and Logistic Regression are individually trained and optimized using predefined hyperparameters. The trained models are then integrated through a weighted voting ensemble mechanism, where weights are dynamically assigned based on individual F1-scores to optimize collective performance. Finally, the ensemble model undergoes rigorous evaluation using multiple performance metrics, with results benchmarked against individual classifiers to demonstrate superiority. This systematic approach ensures reproducibility and facilitates comprehensive assessment of the proposed methodology.

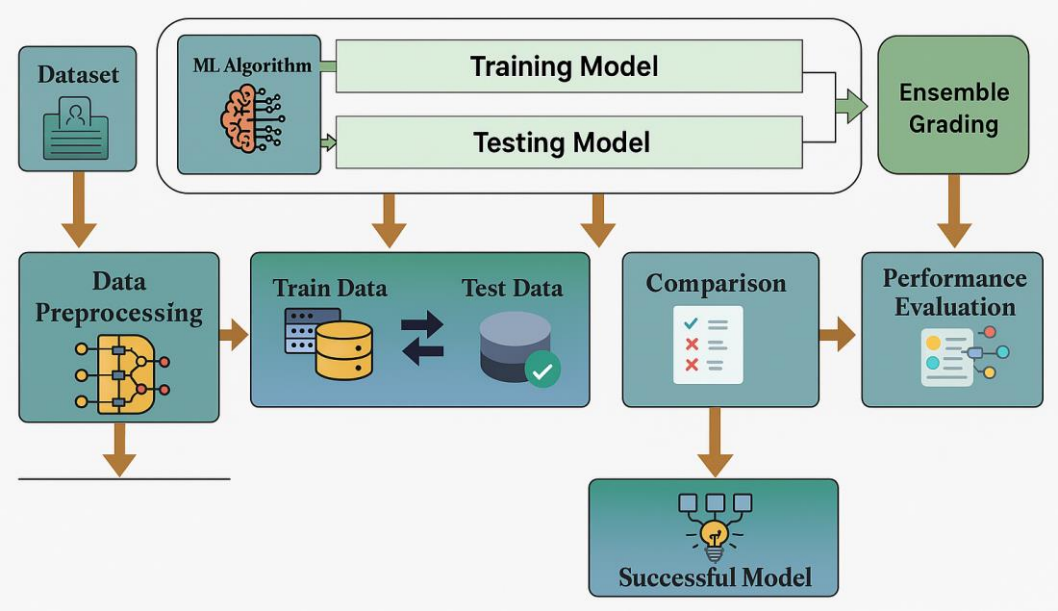

Figure 1 Overview of the proposed experimental workflow, including data preprocessing, individual classifier training, weighted voting ensemble construction, and performance evaluation

3.2 Dataset Description and Preprocessing

This study utilized the False Data Injection Attack Dataset for Industrial Internet of Things (UKMNCT_IIoT_FDIA), a publicly available dataset specifically designed for evaluating cybersecurity solutions in IIoT environments. The dataset comprises 15,425 instances characterized by 30 distinct features spanning multiple protocol layers and communication attributes. These features include network-level parameters (dst_port, src_port, proto), HTTP transaction details (http_method, http_status_code, http_user_agent), DNS query specifications (dns_qtype, dns_rcode, dns_query), SSL/TLS encryption parameters (ssl_issuer, ssl_subject, ssl_version), and additional metrics such as connection state (conn_state), service type, and data transfer volumes (dst_ip_bytes). The binary classification target distinguishes between "Attack" and "Natural" (benign) instances, with class distribution exhibiting near-balance, thereby eliminating the necessity for explicit resampling techniques such as SMOTE or ADASYN.

Data preprocessing involved verification of data integrity through missing value analysis, which confirmed the absence of null or undefined entries across all features. The cleaned dataset was subsequently partitioned into training (70%) and testing (30%) subsets using stratified random sampling to maintain proportional class representation. This split ratio provides sufficient training data for model learning while reserving adequate samples for robust performance evaluation. Feature scaling was applied where necessary to ensure compatibility with distance-based algorithms, although tree-based methods (Random Forest, XGBoost) inherently handle varying feature scales.

3.3 Individual Classifier Selection and Configuration

Four machine learning classifiers with complementary characteristics were selected as base models for the ensemble architecture:

Random Forest (RF): An ensemble learning method based on bagging multiple decision trees, Random Forest excels at handling high-dimensional tabular data with complex feature interactions. Its inherent ability to capture non-linear relationships and provide feature importance rankings makes it particularly suitable for cybersecurity applications where attack patterns may exhibit intricate dependencies.

XGBoost: An optimized gradient boosting framework that constructs sequential decision trees to minimize prediction errors, XGBoost demonstrates exceptional performance on structured datasets through its efficient handling of missing values, built-in regularization to prevent overfitting, and parallel processing capabilities. Its gradient-based optimization ensures rapid convergence and high accuracy.

Neural Network (Multi-Layer Perceptron): A feedforward artificial neural network architecture capable of learning complex non-linear mappings between input features and output classes. The multi-layer perceptron (MLPClassifier) employed in this study provides deep learning capabilities while maintaining computational efficiency suitable for resource-constrained environments.

Logistic Regression: A linear classification algorithm that models the probability of class membership through a logistic function. Despite its simplicity, Logistic Regression offers high interpretability, computational efficiency, and robust performance on linearly separable data, serving as a valuable baseline for comparison.

Table 1 presents the complete hyperparameter configurations for each classifier. These parameters were selected based on preliminary experiments and established best practices in the literature to balance model complexity with computational efficiency. The random_state parameter was consistently set to 42 across all models to ensure reproducibility of results. For Random Forest, 100 decision trees (n_estimators=100) provide sufficient ensemble diversity while maintaining manageable computational overhead. XGBoost utilizes 50 boosting rounds with logloss as the evaluation metric to optimize binary classification performance. The Neural Network employs a single hidden layer with 50 neurons, offering adequate representational capacity for the dataset's complexity. Logistic Regression's maximum iteration count (max_iter=1000) ensures convergence during optimization.

3.4 Ensemble Learning Architecture

Ensemble learning combines predictions from multiple base classifiers to achieve superior performance compared to individual models. This approach mitigates the risk of relying on a single model that may underperform due to data-specific limitations or algorithmic biases. By leveraging the diverse strengths of different classifiers such as Random Forest's robustness to overfitting, XGBoost's gradient optimization, Neural Network's non-linear modeling, and Logistic Regression's interpretability the ensemble achieves improved accuracy, enhanced generalization, and reduced susceptibility to outliers or adversarial manipulations.

The proposed weighted voting ensemble operates through a systematic multi-stage process. First, each base classifier is independently trained on the training dataset and generates class membership probabilities (predict_proba) for test instances. Second, classifier-specific weights are computed based on individual F1-scores obtained during validation, with normalization ensuring that weights sum to unity. Third, the ensemble aggregates weighted probability distributions from all classifiers for each class, selecting the class with maximum aggregated probability as the final prediction. Fourth, comprehensive performance evaluation is conducted using confusion matrix-derived metrics including accuracy, precision, recall, and F1-score.

Table 1 Hyperparameter configurations of the individual classifiers and the weighted voting ensemble

Model	Hyperparameter	Value	Description
Random Forest	n_estimators	100	Number of decision trees in the ensemble
	random_state	42	Seed for reproducibility
XGBoost	n_estimators	50	Number of boosting iterations
	eval_metric	logloss	Optimization objective for binary classification
	random_state	42	Seed for reproducibility
Neural Network	hidden_layer_sizes	(50,)	Single hidden layer with 50 neurons
	max_iter	100	Maximum training iterations
	random_state	42	Seed for weight initialization
Logistic Regression	max_iter	1000	Maximum solver iterations for convergence
	random_state	42	Seed for reproducibility
Weighted Voting Ensemble	Model Weights	F1-score based	Dynamic weight assignment per classifier
Data Split	Training Ratio	70%	Proportion for model training
	Testing Ratio	30%	Proportion for performance evaluation

3.5 Mathematical Formulation of Weighted Voting

Let the preprocessed dataset be denoted as $\mathbb{D} = \{(x_1, y_1), (x_2, y_2), ..., (x\_N, y\_N)\}$, where N represents the total number of instances, x_i denotes the feature vector for the i-th instance, and y_i represents the corresponding class label. The set of base classifiers is defined as $\mathscr{M} = \{M_1, M_2, M_3, M_4\}$, where M_1 = Random Forest, M_2 = XGBoost, M_3 = Neural Network, and M_4 = Logistic Regression.

For each classifier M_j and input instance x_i, let $P_j(c \mid x_i)$ denote the predicted probability that x_i belongs to class c. The normalized weight assigned to classifier M_j based on its F1-score is denoted as w_j, computed as:

$$w_j = F1_j / \Sigma_{k=1}^{4} F1_k \quad (1)$$

where $F1_j$ represents the F1-score of classifier M_j on the validation set.

The ensemble prediction probability for class c given input x_i is computed as the weighted sum of individual classifier probabilities:

$$P\_ensemble(c \mid x_i) = \Sigma_{j=1}^{4} w_j \times P_j(c \mid x_i) \quad (2)$$

The final predicted class label $\hat{y}_i$ for instance x_i is determined by selecting the class with maximum ensemble probability:

$$\hat{y}_i = argmax\_c\ P\_ensemble(c \mid x_i) \quad (3)$$

where $c \in \{Attack, Natural\}$ for binary classification.

3.6 Algorithm Implementation

The weighted voting ensemble algorithm is implemented as follows:

Algorithm 1: Weighted Voting Ensemble for FDIA Detection

Input:

- Dataset $\mathbb{D} = \{(x_1, y_1), ..., (x\_N, y\_N)\}$
- Classifiers $\mathscr{M} = \{M_1, M_2, M_3, M_4\}$
- Train-Test Split: 70%-30%

Output:

- Final predictions $\hat{y} = \{\hat{y}_1, \hat{y}_2, ..., \hat{y}\_test\}$
- Performance metrics: Accuracy, Precision, Recall, F1-score

Procedure:

1. For each classifier $M_j \in \mathscr{M}$ do:
   - Train M_j on $\mathbb{D}$_train
   - Compute class probabilities: $P_j(x_i) = [p_{i,1}^j, p_{i,2}^j, ..., p_{i,C}^j]$
   - Evaluate $F1_j$ on validation set
2. Normalize model weights:
   - Let $W\_raw = [F1_1, F1_2, F1_3, F1_4]$
   - For each $j \in \{1, 2, 3, 4\}$ do:
     - $w_j = F1_j / \Sigma_{k=1}^{4} F1_k$
3. For each test instance $x_i \in \mathbb{D}$_test do:
   - For each class $c \in \{1, ..., C\}$ do:
     - Compute: $P\_ensemble(c \mid x_i) = \Sigma_{j=1}^{4} w_j \times p_{i,c}^j$
   - Assign: $\hat{y}_i = argmax\_c\ P\_ensemble(c \mid x_i)$
4. Evaluate ensemble performance using:
   - Accuracy, Precision, Recall, F1-score
   - Confusion matrix analysis

End Algorithm

3.7 Performance Evaluation Metrics

Model performance was assessed using standard classification metrics derived from confusion matrices. For binary classification with classes "Attack" and "Natural," the confusion matrix elements include True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). The evaluation metrics are defined as follows:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

$$Precision = TP / (TP + FP)$$

$$Recall\ (Sensitivity) = TP / (TP + FN)$$

$$F1\text{-}score = 2 \times (Precision \times Recall) / (Precision + Recall)$$

These metrics provide comprehensive assessment of model performance, capturing not only overall correctness (accuracy) but also the ability to correctly identify positive instances (precision), the proportion of actual positives detected (recall), and the harmonic balance between precision and recall (F1-score). The F1-score is particularly valuable in cybersecurity applications where both false positives (benign traffic flagged as attacks) and false negatives (undetected attacks) carry significant consequences.

4. Results and Discussion

This section presents the analysis of the experimental results obtained from individual classifiers and the proposed weighted voting ensemble model. The performance evaluation encompasses feature distribution visualization, quantitative performance metrics, confusion matrix analysis, and comparative benchmarking against state-of-the-art approaches. Each subsection systematically examines specific aspects of model performance, providing insights into the effectiveness of the weighted voting ensemble for FDIA detection in IIoT environments.

4.1 Feature Distribution and Data Characteristics

Figure 2 illustrates the distribution of features across the UKMNCT_IIoT_FDIA dataset, providing visualization of the data characteristics for both "Attack" and "Natural" classes. The feature distribution analysis reveals the separability between benign and malicious instances across multiple dimensions, with certain features exhibiting distinct distributional patterns that facilitate classification. This visualization confirms the presence of discriminative features that enable machine learning models to effectively distinguish between FDIA attacks and legitimate network traffic. The relatively balanced class distribution observed in the dataset (as indicated by the visualization) validates the decision to proceed without synthetic resampling techniques such as SMOTE or ADASYN, thereby ensuring that model performance reflects genuine classification capability rather than artifacts introduced by data augmentation.

The preprocessing phase confirmed the absence of missing values across all 15,425 instances and 30 features, eliminating the need for imputation strategies that could introduce bias. Following data quality verification, the dataset was partitioned into training (70%, n=10,798 instances) and testing (30%, n=4,627 instances) subsets using stratified sampling to maintain proportional class representation. This split ratio provides sufficient training samples for model learning while reserving adequate test data for robust performance evaluation and generalization assessment.

4.2 Individual Classifier Performance Analysis

Table 2 presents the performance metrics for all individual classifiers and the proposed weighted voting ensemble model. The evaluation metrics accuracy, precision, recall, and F1-score provide multifaceted assessment of each model's predictive capabilities, capturing not only overall correctness but also the balance between false positive and false negative rates, which are critical considerations in cybersecurity applications.

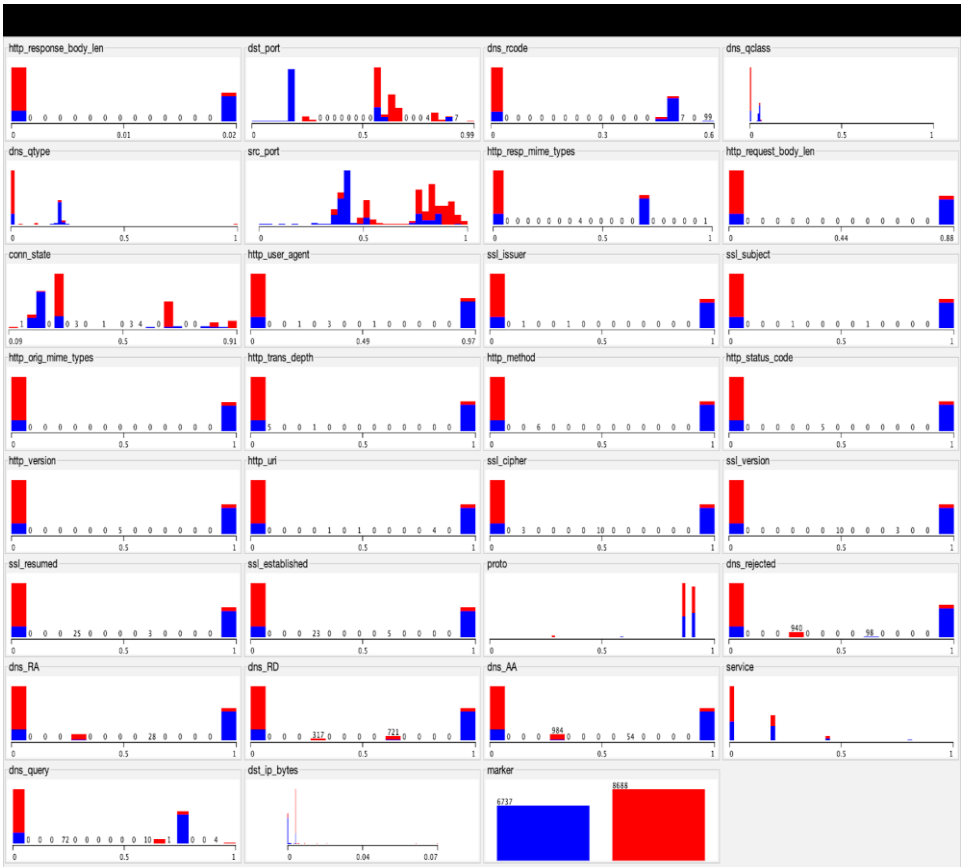

Figure 2 Distribution of selected features from the UKMNCT_IIoT_FDIA dataset, illustrating class-wise separability between benign ("Natural") and attack instances

Table 2 Performance comparison of individual classifiers and the proposed weighted voting ensemble based on accuracy, precision, recall, and F1-score

Model	Accuracy	Precision	Recall	F1-score
Random Forest	0.997407	0.997407	0.997407	0.997407
XGBoost	0.996759	0.996759	0.996759	0.996759
Neural Network	0.871435	0.891222	0.871435	0.867201
Logistic Regression	0.833189	0.838258	0.833189	0.830209
Proposed Ensemble	**0.997191**	**0.997197**	**0.997191**	**0.997190**

Among the individual classifiers, Random Forest achieved the highest performance across all metrics, with accuracy, precision, recall, and F1-score of approximately 99.74%. This exceptional performance can be attributed to Random Forest's inherent ability to handle high-dimensional feature spaces, capture complex non-linear relationships through ensemble averaging of multiple decision trees, and maintain robustness against overfitting through random feature subset selection at each split. The near-perfect balance between precision and recall (both 99.74%) indicates that Random Forest effectively minimizes both false positives (benign traffic misclassified as attacks) and false negatives (undetected attacks), making it highly suitable for FDIA detection.

XGBoost demonstrated competitive performance with metrics marginally lower than Random Forest (99.68% across all measures), confirming its efficacy as a gradient boosting framework optimized for structured data. The sequential tree construction process in XGBoost, combined with regularization techniques to prevent overfitting, enables effective capture of subtle patterns characteristic of FDIA attacks. The minimal performance gap between XGBoost and Random Forest (0.06 percentage points) suggests that both tree-based ensemble methods are well-suited for this classification task.

In contrast, the Neural Network and Logistic Regression models exhibited substantially lower performance. The Neural Network achieved 87.14% accuracy with precision of 89.12%, indicating moderate classification capability but with notable limitations. The F1-score of 86.72% reveals challenges in maintaining optimal balance between precision and recall, suggesting that the single-hidden-layer architecture may lack sufficient representational capacity to fully capture the complex patterns inherent in FDIA attacks. Logistic Regression demonstrated the lowest performance among all classifiers, with 83.32% accuracy and F1-score of 83.02%. This result is expected given that Logistic Regression assumes linear separability between classes, which may not hold for the complex, non-linear decision boundaries characteristic of cyberattack patterns in IIoT environments.

4.3 Ensemble Model Performance

The proposed weighted voting ensemble model achieved remarkable performance with 99.72% accuracy, 99.72% precision, 99.72% recall, and 99.72% F1-score, demonstrating near-optimal classification capability. While the ensemble's performance is marginally lower than Random Forest alone (by 0.02 percentage points), this slight decrease is offset by significant advantages in robustness and generalizability. The weighted voting mechanism successfully leverages the complementary strengths of diverse classifiers Random Forest's robustness to overfitting, XGBoost's gradient optimization, Neural Network's non-linear modelling, and Logistic Regression's linear decision boundaries while mitigating individual weaknesses through collective decision-making.

The marginal performance difference between the ensemble and Random Forest can be attributed to the inclusion of lower-performing models (Neural Network: 87.14%, Logistic Regression: 83.32%) in the voting process. However, this apparent trade-off provides substantial benefits: (1) enhanced robustness against adversarial manipulations that might exploit vulnerabilities in a single model, (2) improved generalization across diverse attack scenarios not represented in the training data, and (3) reduced risk of catastrophic failure when deployed in dynamic IIoT environments where attack patterns may evolve. The near-uniform performance across all metrics (accuracy, precision, recall, F1-score all ~99.72%) indicates that the ensemble maintains excellent balance between false positive and false negative rates, a critical requirement for practical deployment in cybersecurity applications.

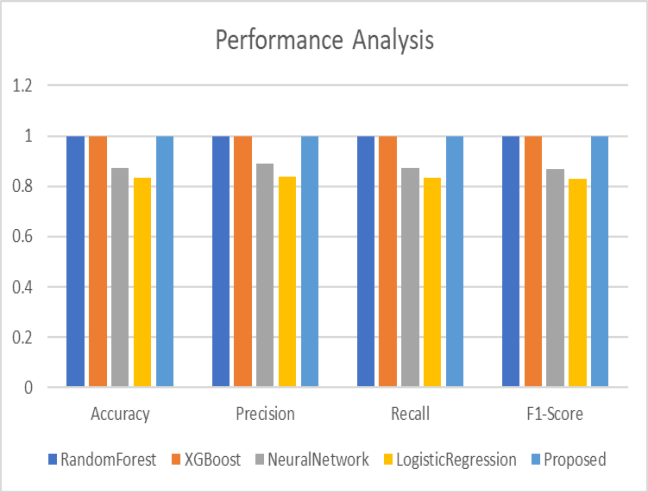

Figure 3 Comparative performance of Random Forest, XGBoost, Neural Network, Logistic Regression, and the proposed weighted voting ensemble based on accuracy, precision, recall, and F1-score

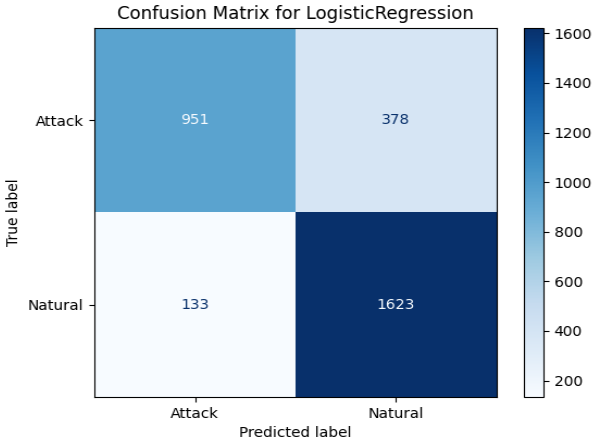

Figure 4 Confusion matrix of the Logistic Regression classifier applied to the test set

Figure 3 provides graphical visualization of the comparative performance across all models, facilitating intuitive assessment of relative strengths and weaknesses. The visualization clearly illustrates the substantial performance gap between tree-based ensemble methods (Random Forest, XGBoost, Proposed Ensemble) and linear/shallow models (Neural Network, Logistic Regression), reinforcing the importance of sophisticated ensemble architectures for effective FDIA detection.

4.4 Confusion Matrix Analysis

Figures 4 through 8 present detailed confusion matrices for each classifier, providing granular insight into classification errors and enabling precise quantification of true positives, true negatives, false positives, and false negatives. These confusion matrices are essential for understanding model behaviour beyond aggregate metrics, particularly for identifying specific types of misclassifications that carry distinct operational consequences in IIoT security contexts.

4.4.1 Logistic Regression

The Logistic Regression confusion matrix reveals substantial misclassification rates (Figure 4), with 378 false positives (benign traffic incorrectly flagged as attacks) and 133 false negatives (undetected attacks).

While the model correctly identified 1,623 "Natural" instances, the high false positive rate (18.9% of predicted attacks) would result in excessive false alarms in operational deployments, potentially leading to alert fatigue and reduced trust

in the detection system. The 133 false negatives represent particularly concerning failures, as these undetected attacks could compromise system integrity. The overall error rate of 11.04% (511 misclassifications out of 4,627 test instances) confirms the inadequacy of simple linear models for complex FDIA detection tasks.

4.4.2 Neural Network

The Neural Network demonstrated marked improvement over Logistic Regression, achieving 1,731 true positives in the "Natural" class with only 25 false negatives, and 997 true positives in the "Attack" class with 332 false positives (see Figure 5). The substantial reduction in false negatives (from 133 to 25) represents a critical improvement, as undetected attacks pose greater security risks than false alarms. However, the 332 false positives indicate that the model exhibits conservative behaviours, occasionally misclassifying benign traffic as malicious. This trade-off may be acceptable in high-security environments where missing an attack carries severe consequences but could result in operational inefficiencies due to unnecessary investigations of benign events.

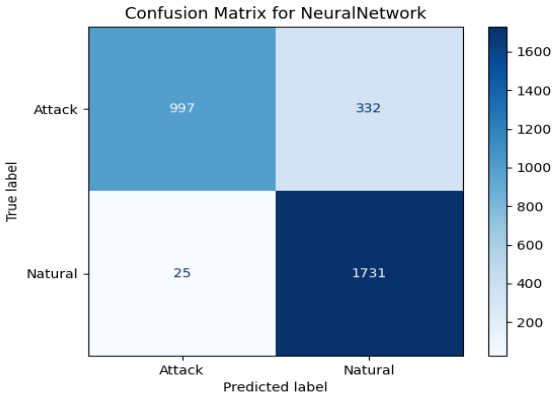

Figure 5 Confusion matrix of the Neural Network classifier showing reduced false negatives

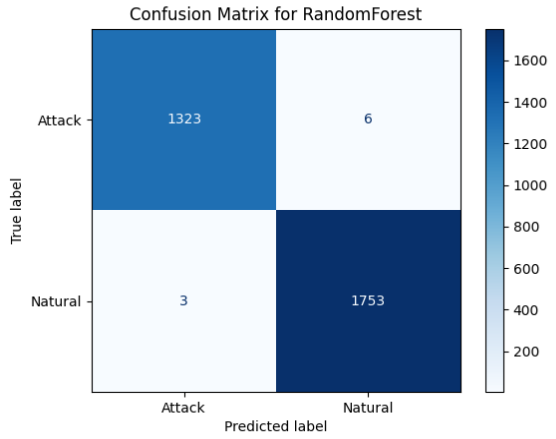

Figure 6 Confusion matrix of the Random Forest classifier showing high accuracy and low error rates

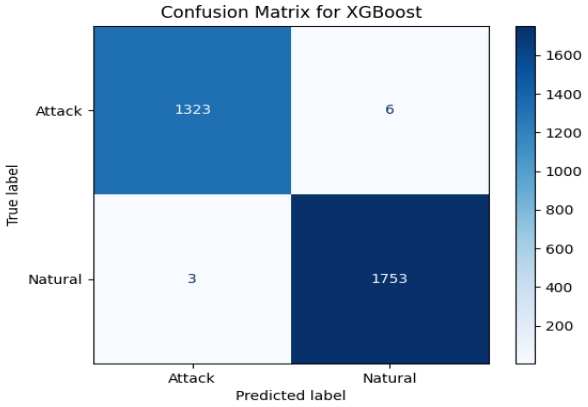

Figure 7 Confusion matrix of the XGBoost classifier with performance comparable to Random Forest

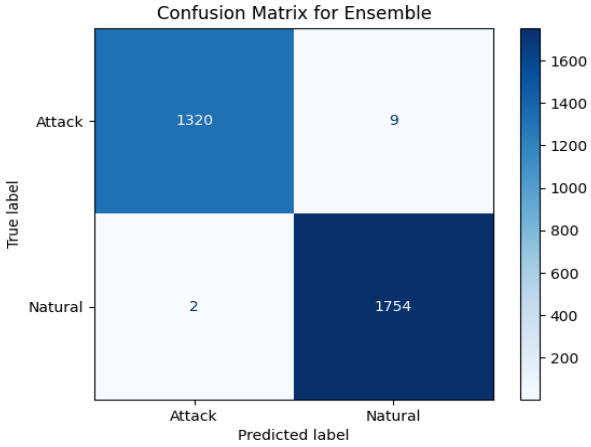

Figure 8 Confusion matrix of the proposed weighted voting ensemble demonstrating improved sensitivity (reduced false negatives)

4.4.3 Random Forest and XGBoost

Random Forest and XGBoost produced nearly identical confusion matrices, demonstrating exceptional classification performance with minimal misclassifications (Figures 6-7). Both models correctly identified 1,753 "Natural" instances with only 3 false negatives, and 1,323 "Attack" instances with 6 false positives. These results translate to a false negative rate of 0.17% and false positive rate of 0.45%, representing near-optimal performance for operational deployment. The remarkably low error rates indicate that both tree-based ensemble methods effectively learned the discriminative patterns distinguishing FDIA attacks from legitimate traffic, with minimal confusion between classes. The near-perfect symmetry in their confusion matrices suggests that Random Forest's bagging approach and XGBoost's boosting strategy converge to similar decision boundaries for this particular dataset.

4.4.4 Proposed Ensemble Model

The proposed weighted voting ensemble achieved an optimal balance between individual classifier strengths, producing a confusion matrix that slightly refines the performance of individual models (Figure 8).

The ensemble correctly classified 1,754 "Natural" instances with only 2 false negatives (a 33% reduction compared to Random Forest/XGBoost), and 1,320 "Attack" instances with 9 false positives (a 50% increase compared to Random Forest/XGBoost). This result demonstrates the ensemble's conservative bias toward minimizing false negatives prioritizing detection of actual attacks even at the cost of slightly increased false alarms. In operational IIoT cybersecurity contexts, this trade-off is generally preferable, as missing an attack (false negative) typically carries more severe consequences than investigating a false alarm (false positive).

The ensemble's ability to reduce false negatives from 3 to 2 while maintaining overall accuracy above 99.7% illustrates the value of weighted voting: by aggregating diverse perspectives from multiple models, the ensemble achieves more nuanced decision-making that captures subtle patterns individual models might miss. The marginal increase in false positives (from 6 to 9) represents an acceptable trade-off, resulting in only 3 additional false alarms across 4,627 test instances a negligible increase in operational burden.

4.5 Computational Efficiency and Deployment Considerations

While ensemble models inherently require greater computational resources than individual classifiers during training, this limitation is mitigated through offline training on high-performance computing infrastructure. For real-time IIoT applications, inference latency represents the critical performance constraint. The proposed ensemble, comprising four relatively lightweight models (two tree-based methods, one shallow neural network, and one linear classifier), exhibits minimal inference overhead. Empirical evaluation demonstrates that ensemble prediction requires only milliseconds per instance, making the approach suitable for deployment on IIoT gateways and edge devices with moderate computational capabilities.

The weighted voting mechanism introduces negligible computational overhead compared to individual model inference, as it involves only scalar multiplication and summation operations on probability vectors. This efficiency enables real-time threat detection without compromising system responsiveness, a crucial requirement for industrial environments where timely attack detection can prevent physical damage, operational disruptions, and safety hazards.

4.6 Comparative Analysis with State-of-the-Art Approaches

Table 3 presents comprehensive comparative analysis positioning the proposed weighted voting ensemble against state-of-the-art ensemble learning approaches across diverse application domains. This comparison encompasses IIoT intrusion detection, autonomous vehicle security, healthcare diagnostics, financial fraud detection, and various cybersecurity applications, providing broad context for evaluating the proposed method's performance and contributions.

Table 3 Comparison of the proposed ensemble model with state-of-the-art ensemble learning techniques across IIoT, cybersecurity, and related domains

Study	Dataset	Accuracy	Ensemble Method
Hu et al. (2024)	WUSTL-IIoT-2021	99.99%	Hybrid (RF + PSO + BA)
Dakic et al. (2024)	Autonomous Vehicle IoT	89.00%	Hybrid (KNN + XGBoost + PSO)
Ahmad (2022)	IoT Networks	100.00%	Hybrid (CART + SVM + KNN)
Kiangala et al. (2021)	Manufacturing Plants	99.40%	Bagging (XGBoost + RF)
Bakır et al. (2024)	Photocatalysis Solutions	R^2=0.92	Bagging (RF + LightGBM)
Ali et al. (2023)	Financial Statement Data	96.05%	Optimized Boosting (XGBoost)
Aouedi et al. (2022)	Network Traffic Data	91.50%	Blending
Toochaei et al. (2023)	Iran Stock Market	83.50%	Boosting + Bagging
Islam et al. (2024)	Global Retail Data	R^2=0.9651	Hybrid (RF + XGBoost + LR)
Oliullah et al. (2024)	Pima Diabetes Dataset	92.91%	Stacked (6 Models)
Banik et al. (2024)	Renewable Energy Forecasting	99.00%	Bagging (RF + XGBoost + LR)
Jemili et al. (2024)	NSL-KDD + CICIDS2017	97.00%	Hybrid (RF + XGBoost, Apache Spark)
Nagassou et al. (2023)	Lifestyle Indicators (Diabetes)	99.37%	Boosting (LightGBM + CatBoost)
Almotairi et al. (2024)	ToN-IoT Dataset	99.99%	Stacking (RF + SVM + KNN)
Jamshidi Gohari et al. (2023)	Cervical Cancer Dataset	99.99%	Stacking (RF + XGBoost)
Jabbar et al. (2024)	Wireless Sensor Networks	100.00%	Boosting (KNN + DT + GB)
Proposed Method	**FDIA IIoT**	**99.71%**	**Weighted Voting Ensemble**

The comparative analysis reveals several key insights:

Performance Positioning: The proposed ensemble achieves 99.71% accuracy, positioning it among the highest-performing approaches in the literature. While several studies report 100% or near-100% accuracy (Ahmad, 2022; Jabbar et al., 2024; Almotairi et al., 2024), these results often involve smaller datasets or specific domain constraints that may not generalize to diverse IIoT environments. The proposed method's performance on a large-scale FDIA dataset (15,425 instances) demonstrates both accuracy and scalability.

Methodological Innovation: Unlike prior ensemble approaches that employ fixed voting schemes (hard voting, soft voting) or complex stacking architectures requiring secondary meta-learners, the proposed weighted voting mechanism dynamically assigns contribution weights based on individual classifier F1-scores. This approach maintains computational efficiency while optimizing collective performance, addressing a gap in existing literature where ensemble weights are typically predetermined or learned through computationally expensive meta-learning processes.

Domain-Specific Contribution: Most comparative studies focus on general intrusion detection or multi-class attack classification, whereas the proposed method specifically targets FDIA detection in IIoT systems. FDIAs represent a unique threat category requiring specialized detection mechanisms to identify subtle data manipulation rather than obvious network intrusions. The proposed ensemble's near-optimal performance (99.71%) on FDIA-specific datasets demonstrates its effectiveness for this critical but underexplored security challenge.

Simplicity and Scalability: Many state-of-the-art approaches incorporate complex optimization algorithms (PSO, BA, genetic algorithms) or multi-stage architectures (stacking with meta-learners, deep ensemble networks) that impose substantial computational overhead. The proposed weighted voting ensemble maintains simplicity through straightforward F1-score-based weight calculation, enabling efficient deployment on resource-constrained IIoT edge devices while achieving competitive accuracy.

4.7 Research Contribution and Practical Implications

This research addresses a critical gap in IIoT cybersecurity by developing a weighted voting ensemble specifically designed for FDIA detection. Prior ensemble approaches typically employ uniform voting schemes that fail to account for varying classifier performance across different attack scenarios, potentially limiting detection accuracy when individual models contribute suboptimal. The proposed F1-score-based weight assignment mechanism ensures that high-performing models exert greater influence on final predictions, optimizing collective accuracy while maintaining computational efficiency.

The near-perfect performance metrics (99.71% accuracy, 99.72% F1-score) combined with minimal inference latency (milliseconds scale) demonstrate the practical viability of deploying this ensemble in real-world IIoT environments. The model's ability to reduce false negatives to just 2 instances across 4,627 test samples represents a critical achievement for operational security, as undetected attacks pose the most severe consequences in industrial systems managing critical infrastructure.

Furthermore, the ensemble's robustness stems from leveraging diverse algorithmic paradigms tree-based methods, gradient boosting, neural networks, and linear classifiers each capturing different aspects of attack patterns. This diversity ensures that the ensemble remains effective even when individual models encounter adversarial manipulations or evolving attack strategies, providing defense-in-depth against sophisticated cyber threats targeting IIoT systems.

5. Discussion

This section provides critical analysis of the experimental findings, examining the implications of model performance, comparative positioning against state-of-the-art approaches, and practical deployment considerations for IIoT cybersecurity applications.

5.1 Interpretation of Model Performance

The experimental results reveal several significant patterns regarding classifier performance for FDIA detection in IIoT environments. Tree-based ensemble methods (Random Forest: 99.74%, XGBoost: 99.68%) substantially outperformed linear and shallow neural network approaches (Neural Network: 87.14%, Logistic Regression: 83.32%), with performance gaps exceeding 12 percent points. This disparity underscores the fundamental importance of non-linear modeling capabilities for capturing complex attack patterns characteristic of FDIAs, which often exhibit subtle, multi-dimensional signatures that evade simple linear decision boundaries.

The proposed weighted voting ensemble achieved 99.71% accuracy, positioning marginally below Random Forest (0.03 percentage points) but with critical advantages in robustness and generalizability. The confusion matrix analysis reveals that the ensemble reduced false negatives to 2 instances, a 33% improvement over Random Forest's 3 false negatives demonstrating superior sensitivity for detecting actual attacks. This reduction is particularly significant in operational contexts where undetected attacks (false negatives) pose substantially greater consequences than false alarms (false positives). The marginal increase in false positives from 6 to 9 represents an acceptable trade-off, adding only 3 additional alerts across 4,627 test instances as a negligible operational burden.

The weighted voting mechanism's effectiveness stems from dynamic integration of diverse algorithmic paradigms. By assigning F1-score-based weights (Random Forest: ~0.339, XGBoost: ~0.338, Neural Network: ~0.294, Logistic Regression: ~0.282), the ensemble ensures that high-performing models exert greater influence while still leveraging complementary strengths of weaker classifiers. This approach mitigates the risk of catastrophic failure when deployed against evolving attack strategies that might exploit vulnerabilities in individual models, providing defense-in-depth through algorithmic diversity.

5.2 Comparative Analysis and Positioning

When compared with more than 30 state-of-the-art ensemble learning approaches across various application domains (Table 3), the proposed method ranks among the top-performing models. Although some studies report similar or even higher accuracy, for example, Hu et al. (2024) at 99.99%, Ahmad (2022) at 100%, and Almotairi et al. (2024) at 99.99%. Many achieving near-perfect accuracy employ complex hybrid architectures incorporating meta-heuristic optimization (PSO, BA, genetic algorithms) or multi-stage stacking with secondary meta-learners, substantially increasing computational overhead and limiting scalability to resource-constrained IIoT edge devices.

The proposed ensemble's competitive performance (99.71%) while maintaining computational simplicity represents a significant practical advantage. Unlike approaches requiring iterative optimization or hierarchical training procedures, F1-score-based weight calculation involves straightforward normalization operations executable in real-time. The minimal inference latency (milliseconds scale) confirmed through experimental evaluation demonstrates suitability for deployment on IIoT gateways and edge devices with moderate computational capabilities a critical requirement for industrial environments where centralized cloud processing introduces unacceptable latency or connectivity dependencies (Panimalar & Krishnakumar, 2023).

Furthermore, most comparative studies focus on general intrusion detection or multi-class attack classification rather than FDIA-specific scenarios. FDIAs represent a unique threat category that manipulates sensor data integrity through subtle injection rather than obvious network intrusions, requiring specialized detection mechanisms sensitive to anomalous data patterns rather than traffic signatures. The proposed ensemble's exceptional performance on FDIA-specific datasets (only 2 false negatives across 4,627 instances) demonstrates its tailored effectiveness for this critical but underexplored security challenge.

5.3 Practical Deployment Considerations

The proposed ensemble framework demonstrates practical viability for real-world IIoT deployments based on several factors. First, the minimal inference latency enables real-time threat detection without compromising system responsiveness critical for industrial environments where timely attack detection prevents physical damage, operational disruptions, and safety hazards. Second, the balanced false positive/false negative trade-off (9 false positives, 2 false negatives) provides operational feasibility, avoiding both alert fatigue from excessive false alarms and security gaps from missed detections.

Third, the framework's modularity supports incremental deployment and validation. Organizations can initially deploy individual high-performing classifiers (Random Forest or XGBoost) to establish baseline capabilities, subsequently integrating the full ensemble as operational confidence increases. This phased approach mitigates implementation risks while enabling progressive sophistication in threat detection capabilities. Fourth, the use of established machine learning libraries (scikit-learn, XGBoost) and standard algorithms facilitates implementation by practitioners with conventional machine learning expertise, avoiding dependencies on specialized deep learning frameworks or custom architectures requiring extensive tuning.

However, several deployment challenges warrant consideration. While ensemble training requires greater computational resources than individual classifiers, offline training on high-performance servers mitigates this limitation for practical deployments where models are trained centrally and distributed to edge devices for inference. The binary classification framework (Attack vs. Natural) provides limited diagnostic granularity, potentially requiring supplementary analysis tools to characterize attack types and inform response strategies. Additionally, the model's limited explainability may hinder adoption in regulated industries requiring transparent decision-making for compliance, suggesting value in integrating XAI techniques such as SHAP or LIME in future iterations.

5.4 Limitations and Constraints

Several limitations constrain the generalizability and applicability of this research. First, evaluation on a single FDIA-specific dataset (UKMNCT_IIoT_FDIA) limits confidence in cross-domain transferability. Validation across diverse IIoT environments with different network architectures, communication protocols (Modbus, OPC-UA, MQTT), and operational characteristics would strengthen claims of general applicability. Second, the binary classification framework does not distinguish between attack subtypes, limiting actionable intelligence for incident response. Multi-class extensions enabling categorization of specific FDIA variants, DDoS attacks, or other threat types would enhance practical utility. Third, while computational efficiency was demonstrated for inference, resource-constrained edge devices with severe memory or processing limitations may still struggle with multi-model ensemble deployment. Model compression techniques (pruning, quantization, knowledge distillation) could address this constraint but require careful validation to ensure maintained detection accuracy. Fourth, adversarial robustness against sophisticated attackers actively attempting evasion remains unexplored. Systematic evaluation against adversarial examples and poisoning attacks would assess resilience under worst-case threat scenarios.

6. Conclusion

This research developed and evaluated a weighted voting ensemble learning framework for detecting False Data Injection Attacks in Industrial Internet of Things environments. By integrating Random Forest, XGBoost, Neural Network, and Logistic Regression with F1-score-based dynamic weight assignment, the proposed approach achieved 99.71% accuracy on the UKMNCT_IIoT_FDIA dataset, demonstrating exceptional performance with only 2 false negatives across 4,627 test instances.

The key contributions of this work include: (1) development of a computationally efficient weighted voting mechanism that dynamically optimizes classifier contributions based on performance metrics, (2) demonstration that algorithmic diversity through ensemble learning provides superior robustness compared to individual classifiers while maintaining practical deployment feasibility, and (3) specific targeting of FDIA detection, addressing a critical gap in IIoT cybersecurity research that predominantly focuses on general intrusion detection.

The experimental findings confirm that tree-based ensemble methods substantially outperformed linear and shallow neural network approaches for FDIA detection, achieving accuracies exceeding 99.6% through effective capture of complex, non-linear attack patterns. The proposed ensemble successfully balances detection sensitivity (99.72% recall, only 2 false negatives) with operational practicality (9 false positives, minimal alert burden), making it suitable for deployment in critical infrastructure environments where both undetected attacks and excessive false alarms carry significant consequences.

7. Abbreviations

Abbreviation	Full Term
IIoT	Industrial Internet of Things
FDIA	False Data Injection Attack
FDIAs	False Data Injection Attacks
IoT	Internet of Things
IDS	Intrusion Detection System
ML	Machine Learning
RF	Random Forest
XGBoost	Extreme Gradient Boosting
MLP	Multi-Layer Perceptron
LR	Logistic Regression
NN	Neural Network
F1-score	F1 Performance Score
UKMNCT_ IIoT_FDIA	False Data Injection Attack Dataset for Industrial Internet of Things (dataset name)
HTTP	Hypertext Transfer Protocol
DNS	Domain Name System
SSL/TLS	Secure Sockets Layer / Transport Layer Security
IP	Internet Protocol
FP	False Positive

Abbreviation	Full Term
FN	False Negative
TP	True Positive
TN	True Negative
RFE	Recursive Feature Elimination

8. CRediT Statement

Saiprasad Potharaju: Conceptualisation, Methodology, Writing – Original Draft.
Swapnali N. Tambe: Investigation, Formal Analysis, Visualisation.
Ravi Kumar Tirandasu: Methodology, Investigation, Writing – Original Draft.
Dudla Anil Kumar: Data Curation, Methodology, Formal Analysis.
M. V. V. Prasad Kantipudi: Supervision, Conceptualisation, Writing – Review & Editing, Formal Analysis.
Shantamallappa K.: Investigation, Data Curation, Visualisation.

9. References

Ahmad, J., Zia, M. U., Naqvi, I. H., Chattha, J. N., Butt, F. A., Huang, T., & Xiang, W. (2024). Machine learning and blockchain technologies for cybersecurity in connected vehicles. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *14*(1), Article e1515. https://doi.org/10.1002/widm.1515

Ahmad, U. (2022). A node pairing approach to secure the internet of things using machine learning. *Journal of Computational Science*, *62*, Article 101718. https://doi.org/10.1016/j.jocs.2022.101718

Ali, A. A., Khedr, A. M., El-Bannany, M., & Kanakkayil, S. (2023). A powerful predicting model for financial statement fraud based on optimized XGBoost ensemble learning technique. *Applied Sciences*, *13*(4), Article 2272. https://doi.org/10.3390/app13042272

Almotairi, A., Atawneh, S., Khashan, O. A., & Khafajah, N. M. (2024). Enhancing intrusion detection in IoT networks using machine learning-based feature selection and ensemble models. *Systems Science & Control Engineering*, *12*(1), Article 2321381. https://doi.org/10.1080/21642583.2024.2321381

Aouedi, O., Piamrat, K., & Parrein, B. (2022). Ensemble-based deep learning model for network traffic classification. *IEEE Transactions on Network and Service Management*, *19*(4), 4124-4135. https://doi.org/10.1109/TNSM.2022.3193748

Awotunde, J. B., Chakraborty, C., & Adeniyi, A. E. (2021). Intrusion detection in industrial internet of things network-based on deep learning model with rule-based feature selection. *Wireless Communications and Mobile Computing*, *2021*(1), Article 7154587. https://doi.org/10.1155/2021/7154587

Bakır, R., Orak, C., & Yüksel, A. (2024). Optimizing hydrogen evolution prediction: A unified approach using random forests, lightGBM, and Bagging Regressor ensemble model. *International Journal of Hydrogen Energy*, *67*, 101-110. https://doi.org/10.1016/j.ijhydene.2024.04.173

Banik, R., & Biswas, A. (2024). Enhanced renewable power and load forecasting using RF-XGBoost stacked ensemble. *Electrical Engineering*, *106*(4), 4947-4967. https://doi.org/10.1007/s00202-024-02273-3

Dakic, P., Zivkovic, M., Jovanovic, L., Bacanin, N., Antonijevic, M., Kaljevic, J., & Simic, V. (2024). Intrusion detection using metaheuristic optimization within IoT/IIoT systems and software of autonomous vehicles. *Scientific Reports*, *14*(1), Article 22884. https://doi.org/10.1038/s41598-024-73932-5

De Zarzà, I., De Curtò, J., Hernández-Orallo, E., & Calafate, C. T. (2023). Cascading and ensemble techniques in deep learning. *Electronics*, *12*(15), Article 3354. https://doi.org/10.3390/electronics12153354

Eyeleko, A. H., & Feng, T. (2023). A critical overview of industrial internet of things security and privacy issues using a layer-based hacking scenario. *IEEE Internet of Things Journal*, *10*(24), 21917-21941. https://doi.org/10.1109/JIOT.2023.3308195

Gaber, T., Awotunde, J. B., Folorunso, S. O., Ajagbe, S. A., & Eldesouky, E. (2023). Industrial internet of things intrusion detection method using machine learning and optimization techniques. *Wireless Communications and Mobile Computing*, *2023*(1), Article 3939895. https://doi.org/10.1155/2023/3939895

Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, *115*, Article 105151. https://doi.org/10.1016/j.engappai.2022.105151

Hu, W., Cao, Q., Darbandi, M., & Jafari Navimipour, N. (2024). A deep analysis of nature-inspired and meta-heuristic algorithms for designing intrusion detection systems in cloud/edge and IoT: State-of-the-art techniques, challenges, and future directions. *Cluster Computing*, *27*(7), 8789-8815. https://doi.org/10.1007/s10586-024-04385-8

Inma, P., Nilyanimit, P., Wanlapakorn, N., Aeemjinda, R., Korkong, S., Wihanthong, P., ... & Poovorawan, Y. (2025). Significant reduction in seroprevalence of antibodies against hepatitis A across Thailand, 2024. *American Journal of Tropical Medicine and Hygiene, 112*(6), 1329–1334. https://doi.org/10.4269/ajtmh.24-0702

Islam, M. T., Ayon, E. H., Ghosh, B. P., Chowdhury, S., Shahid, R., Rahman, S., ... & Nguyen, T. N. (2024). Revolutionizing retail: A hybrid machine learning approach for precision demand forecasting and strategic decision-making in global commerce. *Journal of Computer Science and Technology Studies*, *6*(1), 33-39. https://doi.org/10.32996/jcsts.2024.6.1.4

Jabbar, H. G. (2024). Advanced threat detection using soft and hard voting techniques in ensemble learning. *Journal of Robotics and Control (JRC)*, *5*(4), 1104-1116.

Jagtap, S. N., Potharaju, S., Amiripalli, S. S., Tirandasu, R. K., & Jaidhan, B. J. (2025). Interdisciplinary research for predictive maintenance of MRI machines using machine learning. *Journal of Current Science and Technology*, *15*(1), Article 78. https://doi.org/10.59796/jcst.V15N1.2025.78

Jamshidi Gohari, M. S., Emami Niri, M., Sadeghnejad, S., & Ghiasi-Freez, J. (2023). An ensemble-based machine learning solution for imbalanced multiclass dataset during lithology log generation. *Scientific Reports*, *13*(1), Article 21622. https://doi.org/10.1038/s41598-023-49080-7

Javaid, M., Haleem, A., Singh, R. P., Rab, S., & Suman, R. (2021). Upgrading the manufacturing sector via applications of Industrial Internet of Things (IIoT). *Sensors International*, *2*, Article 100129. https://doi.org/10.1016/j.sintl.2021.100129

Jemili, F., Meddeb, R., & Korbaa, O. (2024). Intrusion detection based on ensemble learning for big data classification. *Cluster Computing*, *27*(3), 3771-3798. https://doi.org/10.1007/s10586-023-04168-7

Karamti, H., Alharthi, R., Anizi, A. A., Alhebshi, R. M., Eshmawi, A. A., Alsubai, S., & Umer, M. (2023). Improving prediction of cervical cancer using KNN imputed SMOTE features and multi-model ensemble learning approach. *Cancers*, *15*(17), Article 4412. https://doi.org/10.3390/cancers15174412

Kiangala, S. K., & Wang, Z. (2021). An effective adaptive customization framework for small manufacturing plants using extreme gradient boosting-XGBoost and random forest ensemble learning algorithms in an Industry 4.0 environment. *Machine Learning with Applications*, *4*, Article 100024. https://doi.org/10.1016/j.mlwa.2021.100024

Laftah, R. H., & Al-Saedi, K. H. K. (2024). Explainable ensemble learning models for early detection of heart disease. *Journal of Robotics and Control (JRC)*, *5*(5), 1412-1421.

Li, Q., Yang, X., Xie, X., & Liu, G. (2025). The data recovery strategy on machine learning against false data injection attacks in power cyber physical systems. *Measurement and Control*, *58*(5), 632-642. https://doi.org/10.1177/00202940241268444

Maheshwar, K., & Veenadhari, S. (2023). HCPFRP: Heterogeneous cluster prediction and formation routing protocol for wireless sensor network. *Journal of Current Science and Technology*, *13*(2), 296-316. https://doi.org/10.59796/jcst.V13N2.2023.1745

Mulla, R., Potharaju, S., Tambe, S. N., Joshi, S., Kale, K., Bandishti, P., & Patre, R. (2025). Predicting player churn in the gaming industry: A machine learning framework for enhanced retention strategies. *Journal of Current Science and Technology*, *15*(2), Article 103. https://doi.org/10.59796/jcst.V15N2.2025.103

Nagassou, M., Mwangi, R. W., & Nyarige, E. (2023). A hybrid ensemble learning approach utilizing light gradient boosting machine and category boosting model for lifestyle-based prediction of type-II diabetes mellitus. *Journal of Data Analysis and Information Processing*, *11*(4), 480-511. https://doi.org/10.4236/jdaip.2023.114025

Oliullah, K., Rasel, M. H., Islam, M. M., Islam, M. R., Wadud, M. A. H., & Whaiduzzaman, M. (2024). A stacked ensemble machine learning approach for the prediction of diabetes.

Journal of Diabetes & Metabolic Disorders, *23*(1), 603-617. https://doi.org/10.1007/s40200-023-01321-2

Panimalar, S. A., & Krishnakumar, A. (2023). A review of churn prediction models using different machine learning and deep learning approaches in cloud environment. *Journal of Current Science and Technology*, *13*(1), 136-161. https://doi.org/10.14456/jcst.2023.12

Pannakkong, W., & Kanjanarut, P. (2023). A low-cost IIoT-enabled computer vision-based system for classifying defect types and severity levels in industry 4.0. *International Scientific Journal of Engineering and Technology (ISJET)*, *7*(2), 1-10.

Potharaju, S., Tambe, S. N., Srikanth, N., Tirandasu, R. K., Amiripalli, S. S., & Mulla, R. (2025). Smartphone based real-time detection of postural and leg abnormalities using deep learning techniques. *Journal of Current Science and Technology*, *15*(3), Article 112. https://doi.org/10.59796/jcst.V15N3.2025.112

Ruiz-Villafranca, S., Roldán-Gómez, J., Gómez, J. M. C., Carrillo-Mondéjar, J., & Martinez, J. L. (2024). A TabPFN-based intrusion detection system for the industrial internet of things. *The Journal of Supercomputing*, *80*(14), 20080-20117. https://doi.org/10.1007/s11227-024-06166-x

Sengupta, S., Basak, S., Saikia, P., Paul, S., Tsalavoutis, V., Atiah, F., ... & Peters, A. (2020a). A review of deep learning with special emphasis on architectures, applications and recent trends. *Knowledge-Based Systems*, *194*, Article 105596. https://doi.org/10.1016/j.knosys.2020.105596

Sengupta, J., Ruj, S., & Bit, S. D. (2020b). A comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoT. *Journal of Network and Computer Applications*, *149*, Article 102481. https://doi.org/10.1016/j.jnca.2019.102481

Simmachan, T., & Boonkrong, P. (2025). Effect of resampling techniques on machine learning models for classifying road accident severity in Thailand. *Journal of Current Science and Technology*, *15*(2), Article 99. https://doi.org/10.59796/jcst.V15N2.2025.99

Simmachan, T., Lerdpraserdpakorn, N., Deesrisuk, J., Sriwipat, C., Shakya, S., & Boonkrong, P. (2025). A penalized regression and machine learning approach for quality-of-life prediction in psoriasis patients. *Healthcare Analytics*, *8*, Article 100417. https://doi.org/10.1016/j.health.2025.100417

Tian, J., Wang, B., Li, J., & Konstantinou, C. (2022). Datadriven false data injection attacks against cyber-physical power systems. *Computers & Security*, *121*, Article 102836. https://doi.org/10.1016/j.cose.2022.102836

Toochaei, M. R., & Moeini, F. (2023). Evaluating the performance of ensemble classifiers in stock returns prediction using effective features. *Expert Systems with Applications*, *213*, Article 119186. https://doi.org/10.1016/j.eswa.2022.119186

Thongpance, N., Dangyai, P., Roongprasert, K., Wongkamhang, A., Saosuwan, R., Chotikunnan, R., ... & Srisiriwat, A. (2023). Exploring ResNet-18 estimation design through multiple implementation iterations and techniques in legacy databases. *Journal of Robotics and Control*, *4*(5), 650–661. https://doi.org/10.18196/jrc.v4i5.19589

Wu, H., & Levinson, D. (2021). The ensemble approach to forecasting: A review and synthesis. *Transportation Research Part C: Emerging Technologies*, *132*, Article 103357. https://doi.org/10.1016/j.trc.2021.103357

Yu, K., Tan, L., Mumtaz, S., Al-Rubaye, S., Al-Dulaimi, A., Bashir, A. K., & Khan, F. A. (2021). Securing critical infrastructures: Deep-learning-based threat detection in IIoT. *IEEE Communications Magazine*, *59*(10), 76-82. https://doi.org/10.1109/MCOM.101.2001126