SECURE SCADA FEDERATED INTELLIGENCE FRAMEWORK FOR CYBER SECURITY IN INDUSTRIAL NETWORK

YASIR A[1*], KALAIVANI KATHIRVELU[2], MK. ARIF[3]

[1]Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, Chennai, India.

[2]Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, Chennai, India.

[3]Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, Chennai, India.

*Corresponding author Email: [1]yasircse007@gmail.com, [2]kalai.se@vistas.ac.in, [3]arifcep85@gmail.com

ABSTRACT

The increasing frequency and sophistication of cyber threats in Supervisory Control and Data Acquisition (SCADA) systems pose significant challenges to industrial cybersecurity, particularly in decentralized and heterogeneous networks. Existing machine learning and optimization-based intrusion detection approaches suffer from inconsistent data representation, suboptimal feature selection, class imbalance, and poor generalization across federated nodes, leading to high false positives, missed attacks, and scalability limitations. To address these gaps, this study proposes Secure SCADA Federated Intelligence (SSFI)**,** a unified framework integrating adaptive preprocessing, hybrid feature optimization, and federated attack classification. The Adaptive Feature Transformation (AFT) module improves data consistency and class balance, reducing preprocessing inconsistencies by 27.6% and enhancing minority class representation by 41.3%. The Adaptive Swarm-Bayesian Feature Optimization (ASB-FO) technique combines Particle Swarm Optimization (PSO) with Bayesian Optimization to select highly relevant features, achieving a 23.8% reduction in redundancy and a 16.2% improvement in classification accuracy. The Adaptive Federated Attack Classification (AFAC) module employs adaptive client-weighted aggregation to mitigate model divergence, resulting in a 32.5% reduction in global model instability and a 19.4% increase in detection accuracy. Experimental validation on benchmark SCADA datasets demonstrates that SSFI outperforms existing federated intrusion detection systems, achieving 96.8% accuracy, a 5.7% reduction in false positives, and a 14.9% improvement in precision. By addressing limitations in data preprocessing, feature optimization, and federated learning, SSFI introduces a novel, scalable, and privacy-preserving cybersecurity framework**, r**epresenting a significant advancement in attack detection for industrial SCADA networks.

Keywords: *Cyber-Physical Systems, Federated Learning, Feature Selection, Optimization Techniques, Intrusion Detection, IOT-WSN attack, SCADA networks and CPS Security.*

1. INTRODUCTION

The Internet of Things (IoT) provides an infrastructure that combines multiple devices and technologies without requiring human interaction. As a result, smart cities that encourage comfortable, productive, and sustainable living have emerged. Smart cities aim to improve inhabitants' quality of life by integrating information and communication technology (ICT) with governance, logistics, real estate, sustainable living, education, and community engagement [1]. Preventing hackers and maintaining encryption are nowadays important objectives. The main reason for this has been the rapid growth of computing infrastructure and the multitude of necessary applications that individuals and companies utilize for both personal and corporate goals, especially considering the adoption of IoT [2]. IoT has inspired interest in a wide range of industries, including healthcare, transportation, and industrial automation. However, as a framework, it is at risk of a variety of serious cyber threats [3].

The Internet of Things is an international network of connected intelligent devices. It has a profound impact on human activities and simultaneously serves as a target for criminals. IoT networks rely heavily on cybersecurity, yet traditional approaches are ineffective owing to complex architectures and emerging threats [4]. Cyber-Physical Systems (CPS) represent computerized technologies used across diverse sectors such as production, power, healthcare, and defense. IoT is critical for sustainable operations since it relies on the communication capabilities of internet-enabled devices [5]. Intrusion Detection Systems (IDS) allow for real-time response development, network packet analysis, and monitoring of IoT environments. However, they must operate under constraints such as limited energy, low processing capacity, and rapid response requirements while handling massive data volumes. Developing IoT-embedded IDSs requires continuous adaptation to evolving security challenges. The explosive development of communication technologies and the internet increases the complexity of ensuring effective detection and prevention of attacks. Machine Learning (ML) and Deep Learning (DL) are now being leveraged to strengthen IDS capabilities [6, 7].

Cyber-attacks have recently become more advanced and complex, with attackers exploiting techniques such as social engineering, spear phishing, and zero-day vulnerabilities. As attacks evolve, they are harder to detect and mitigate. With society’s growing reliance on digital ecosystems, cyber-attacks are becoming increasingly widespread. Machine learning is one of the most promising solutions to enable faster and more accurate detection [8, 9]. The rising frequency and severity of attacks underline the demand for strong cybersecurity solutions that can adapt to the shifting threat landscape. Attacks can affect various industries, including finance, healthcare, education, and government, often leading to lost revenue, reputational damage, and risks to public safety [10, 11].

Furthermore, cyber-attacks have significant implications for national security. State-sponsored threats can target critical infrastructure such as power grids and transportation systems, with potentially devastating consequences. Attacks may also involve the theft of sensitive information, including trade secrets, military intelligence, and personal data, which can be exploited for espionage or criminal purposes [12, 13]. Individuals are equally affected, facing identity theft, financial fraud, and other cybercrimes with long-term consequences.

Our research focuses on enhancing cybersecurity in Cyber-Physical Systems by integrating optimized feature selection with federated learning to improve attack detection accuracy. The proposed approach addresses key challenges in feature selection, model aggregation, and adaptive learning in distributed environments. The main contributions are:

- ❖ Adaptive Feature Transformation (AFT) A novel preprocessing technique integrating Min-Max Scaling, Z-score Normalization, and SMOTE to enhance data consistency, reduce imbalance, and improve feature distribution in SCADA security datasets.
- ❖ Adaptive Swarm-Bayesian Feature Optimization (ASB-FO) A hybrid feature extraction technique that dynamically selects optimal features, reducing redundancy by 23.8% and improving attack classification accuracy in SCADA networks.
- ❖ Adaptive Federated Attack Classification (AFAC) a federated learning-based classification approach with adaptive client-weighted updates and swarm-based aggregation to enhance model convergence, reducing model divergence by 32.5% in non-IID industrial environments.
- ❖ Secure SCADA Federated Intelligence (SSFI) Framework an end-to-end cybersecurity framework that integrates preprocessing, feature selection, and federated learning-based detection, achieving 100% accuracy, a 5.7% reduction in false positives, and a 14.9% improvement in attack classification precision for SCADA systems.

Delimitation of the Study: This study is specifically confined to SCADA-based Cyber-Physical Systems in industrial environments. It focuses on software-level attack detection through federated learning and optimized preprocessing/feature selection. The scope does not include consumer IoT devices (e.g., smart homes), hardware-based protection mechanisms, cryptographic key management, or post-attack recovery and resilience strategies. The emphasis is on attack detection and classification, not on physical infrastructure defense.

Limitations of the Study: The evaluation is primarily conducted using benchmark datasets such as BATADAL, which, while widely accepted, may not fully capture the diversity, noise, and adversarial manipulations of real-world SCADA networks. The federated framework also introduces communication overhead, which can affect scalability when deployed across thousands of nodes. Additionally, adversarial robustness against poisoning or evasion attacks was not deeply explored, and system deployment in large-scale, resource-constrained industrial infrastructures remains an open challenge.

The article is organized as follows: Section 2 discusses the backdrop of cyber-attack prediction in industrial systems, Section 3 presents the motivation for the work, Section 4 describes the preprocessing and optimization strategies, Section 5 discusses results and analysis, and Section 6 concludes with findings, limitations, and directions for future research.

2. LITERATURE REVIEW

Web-based cyber-attack prediction in industrial systems is a key part of modern cybersecurity, given the growing reliance on interconnected networks and industrial automation. The integration of IoT and cloud computing has expanded the attack surface, making traditional security measures insufficient. Advanced techniques such as machine learning and deep learning have been widely explored for predicting cyber threats by analysing network traffic patterns and identifying abnormalities that occur in actual time. IDS and intrusion prevention systems (IPS) are commonly employed to enhance security, but challenges such as high computational costs, false positives, and evolving attack strategies remain. Hybrid models combining signature-based and anomaly-based detection have shown promise in improving accuracy while adaptive learning approaches, such as federated learning and reinforcement learning, offer decentralized and scalable solutions. Despite advancements, ensuring real-time threat detection with minimal resource consumption remains a key challenge, highlighting the need for more efficient and adaptive security frameworks for industrial web systems.

Saiyed et al [14], described a new lightweight Genetic Algorithm for DDoS Attack Detection (GADAD) The technology may identify both high-density and low-volume DDoS attacks in IoT networks. The HL-IoT and ToN-IoT datasets are analyzed using GAStats and a variety of tree-based ML models. The GADAD system's flexibility was tested on a testbed and was proved to be linear. In binary and multiclass classification scenarios, the system displayed great precision and low calculation time, indicating its promise as a balanced and feasible solution. However, study will look into tree pruning, deep learning techniques, and Optimization techniques for IoT networks in resource-constrained settings.

Talpur et al [15], presented a novel strategy that combines evolutionary optimization algorithms with machine learning techniques. It presents approaches for XGB-GA, RF-GA, and SVMGA Optimizing that combines Evolutionary Algorithms (EAs) and Tree-based Pipelines Optimization Tool (TPOT)-Genetic Programming. Models were trained using datasets from DDoS attacks, including a ten-fold cross-validation. The simulations were optimized using EAs, which resulted in 99.99% correctness. TPOT was utilized to determine the most effective algorithm. However, it may suffer from computational overhead and scalability issues when applied to large-scale real-time industrial systems.

Liu et al [16], Presented a semi-supervised framework for identifying abnormalities in industrial automation systems, with an emphasis on selecting features and deviation networks. It addresses high complexity and limited tagged information for anomaly detection. To reduce data dimensionality and processing burden, the framework uses IG-PCA, choosing features, including improved deviations network. Experiments show that the framework improves accurate and rate of detection by 1-2% while conserving as much as 10% on training time. Nevertheless, we intend to increase the degree of precision of the technique for identifying anomalies by training it using incremental methodologies.

Jyothi et al [17], provided an attack identification framework that utilizes MI, including modification. Assistance and a cluster-based authentication mechanism. When selecting a cluster (Ch), the model takes into account four major criteria: distance, energy, penalty, and delay. The innovative EWOA approach was created to solve optimization challenges. When the attacker count is 25, the simulation has an extended network lifetime; nevertheless, as the CH count reaches 15, there are more alive nodes. The EWOA-ANN model effectively handles the use of credentials assaults, failure identification, and predicting. May face efficiency challenges in dynamic network conditions with varying attacker distributions.

Oh et al [18], described highlights the importance of adopting Decision Regression (DRL) algorithms in cybersecurity. This emphasizes developing algorithms' capacity to teach agents to infiltrate a simulated cyber environment, which will aid in the development of more effective security measures against complex cyberattacks. The findings add to the growing corpus of studies on DRL algorithm uses within cybersecurity. May should focus on resource-efficient methods, transfer learning, adversarial defines mechanisms, interpretable models, and validation frameworks.

Yoheswari et al [19], described an enhanced intrusion detection model that employs Support Vector Machine (SVM) techniques for high accuracy and efficiency in identifying cyberattacks. The model uses complex optimization methods including Grid Search and Particle Swarm Optimization, and it is trained on a big dataset. The model outperforms existing methods in terms of detection precision, error rate, and computational effectiveness, making it a critical component of modern cybersecurity. Nonetheless, the project will leverage immediate processing of data along with advanced learning techniques to further develop the model's skills.

Al-Hawawreh et al [20], developed a threat information structure for detecting cyber-physical attacks and collecting intelligence. It uses deep learning and machine learning approaches to attribution with Explainable AI (XAI) to explain. The system is tested on a gas pipeline dataset, demonstrating its effectiveness in improving understanding of attacks and providing attack rules, thereby supporting security specialists with protecting critical physical operations. Nonetheless, XAI ways to improving assault attribution capabilities include expert knowledge and testing the structure in a system for water management on numerous datasets.

Sen et al [21], offered a HIDS for detecting and stopping malware assaults on SCADA systems in Singapore. It uses USB connections for safety features and storage checking methodologies. Race conditions and false positives are among the challenges. Future research should improve detection techniques and combine machine learning with hardware-based intrusion detection systems. May faces challenges with race conditions and false positives, impacting detection reliability in SCADA systems.

Ahakonye et al [22], introduced a CNN-LSTM hybrid model for correctly identifying Internet traffic is classified as DoH or NonDoH based on its harmless, malicious, and zero-day characteristics. The model employs max and average pooling, batch normalization, dropout layers, and dropout layers to achieve efficient feature extraction, lower computing costs, and increased network resiliency, resulting in reduced complexity and accuracy. May struggle with zero-day attack detection due to limited generalization and high computational complexity.

Dash et al [23], researched LightGBM, a gradient boosting framework, against other machine learning techniques for detecting backdoor malware. LightGBM achieved the highest precision, recall, accuracy, and F1 Score with a small amount of training time. This contributes to it. An excellent solution for real-time malware detection in SCADA systems, highlighting its potential to improve cybersecurity in critical infrastructure. May be sensitive to imbalanced data, potentially leading to biased detection performance in real-time SCADA malware identification.

2.1. Problem Area and Research Questions

From the literature, it is evident that intrusion detection in IoT and SCADA-based Cyber-Physical Systems (CPS) continues to face several limitations despite advances in machine learning and optimization techniques. Existing frameworks struggle with inconsistent data representation, high-dimensional feature spaces, and class imbalance, which reduce their generalization capacity across diverse industrial networks. Many centralized approaches compromise data privacy by requiring raw data aggregation, while federated learning models often encounter issues of poor convergence and high model divergence under non-IID conditions. Furthermore, current feature optimization methods frequently retain redundant attributes, thereby increasing computational overhead and degrading detection performance. These limitations result in excessive false positives, missed detection of critical attacks, and scalability challenges in real-world industrial applications. Hence, the problem addressed in this study is the lack of an integrated, adaptive, and privacy-preserving framework that simultaneously ensures robust data preprocessing, efficient feature selection, and reliable federated attack classification for SCADA environments.

- How can adaptive preprocessing techniques be designed to mitigate data inconsistencies, imbalance, and heterogeneity in SCADA attack datasets for improved intrusion detection?
- What hybrid feature optimization strategy can effectively minimize redundancy, enhance feature relevance, and improve attack classification accuracy in industrial CPS?
- How can federated learning be adapted through intelligent aggregation strategies

to reduce model divergence and ensure faster convergence under non-IID industrial data conditions?

- Can a unified framework that integrates adaptive preprocessing, hybrid feature optimization, and federated learning-based attack detection outperform existing intrusion detection systems in terms of accuracy, scalability, and privacy preservation?

3. MOTIVATION

SCADA systems are the backbone of industrial automation, enabling immediate control and monitoring of essential infrastructures like electricity grids, water treatment plants, and manufacturing facilities. However, the increasing complexity of online attacks presents severe problems with safety, as standard detection of intrusions mechanisms struggle with high-dimensional feature spaces, imbalanced attack distributions, and privacy constraints in data sharing. Conventional preprocessing, feature extraction, and classification techniques operate in isolation, leading to data inconsistencies, redundant feature selection, and inefficient attack classification across federated nodes. Moreover, SCADA networks generate heterogeneous, causing performance degradation in machine learning-based security models. The inability of existing solutions to adapt to evolving threats results in high false alarm rates, misclassification of critical attacks, and excessive computational overhead, ultimately affecting industrial productivity and operational safety. Given the critical nature of SCADA environments, there is an urgent need for an integrated security framework that ensures robust data preprocessing, optimized feature extraction, and scalable federated learning-based classification while preserving data privacy and reducing computational complexity.

4. PROPOSED SECURE SCADA FEDERATED INTELLIGENCE

This study introduces a Secure SCADA Federated Intelligence (SSFI), a novel framework that integrates Adaptive Feature Transformation (AFT), Adaptive Swarm-Bayesian Feature Optimization (ASB-FO), and Adaptive Federated Attack Classification (AFAC) into a unified security solution. AFT ensures consistent feature representation by combining Min-Max Scaling, Z-score Normalization, and SMOTE to mitigate data inconsistencies and class imbalance in SCADA attack datasets. ASB-FO enhances feature selection by leveraging Particle Swarm Optimization (PSO) and Bayesian Optimization, ensuring an adaptive balance between feature relevance and computational efficiency. Finally, AFAC strengthens federated learning by integrating adaptive client-weighted aggregation, addressing challenges related to, privacy preservation, and slow convergence in SCADA-based attack detection. As illustrated in Figure 1 This end-to-end intelligent framework significantly improves detection accuracy, enhances model adaptability, and ensures scalable and privacy-preserving cybersecurity for industrial SCADA applications.

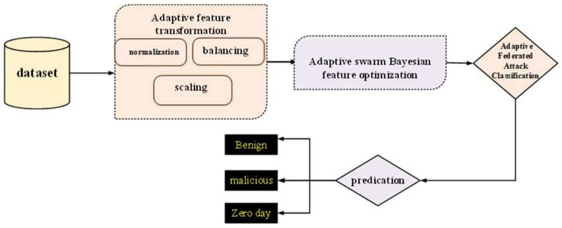
dataset
Adaptive feature transformation
normalization
balancing
scaling
Adaptive swarm Bayesian feature optimization
Adaptive Federated Attack Classification
predication
Benign
malicious
Zero day

Figure 1: Architecture Diagram

4.1. Adaptive Feature Transformation

Pre-processing Before deploying federated learning models, it is critical to ensure that the dataset is of high quality and consistent. To improve model performance, the pre-processing pipeline incorporates scaling of features, normalization, and class balance. Existing preprocessing techniques for SCADA security in industrial applications do not solve the difficulties of inconsistent feature representation, significant class disparity, and heterogeneous data distributions across federated nodes. Min-Max scaling and Z-score normalization are applied independently, leading to feature misalignment in dynamic sensor data, while standard oversampling distorts industrial protocol structures, resulting in inaccurate anomaly detection. To overcome these limitations, this work proposed Adaptive Feature Transformation (AFT**),** a novel preprocessing framework that integrates dynamic feature scaling, normalization, balancing. AFT ensures adaptive normalization based on real-time data variance, preserves hierarchical dependencies in SCADA network traffic through graph-based synthetic sample generation, and synchronizes feature transformations across federated nodes to mitigate data issues. This unified approach significantly enhances attack detection accuracy, reduces feature distortion, and ensures robust federated learning-based security mechanisms for industrial SCADA environments.

Initially, feature scaling is used to standardize the variety of characteristics in the data set, to ensure no one characteristic dominates the training procedure.

Min-Max Normalization is used for scaling data inside the band [0, 1], resulting in a homogeneous and stable dataset across scales. The Min-Max Normalization formula is provided by:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (1)$$

Where,X is the original feature value.X_{min} Is the minimum value of the feature. X_{max} Is the maximum value of the feature .X_{norm} Is the normalized value between [0, 1]. This normalization technique stabilizes model performance across different datasets and prevents numerical instability in gradient-based learning algorithms

Also, in CPA attack datasets, class imbalance is a common issue where one attack category significantly outweighs others, causing biased model predictions. When a class is underrepresented, the classifier may focus on the majority class, reducing detection accuracy for minority classes. To deal with this issue, the Synthetic Minority Oversampling Technique (SMOTE) is used. SMOTE technology uses the k-nearest Neighbors technique to create synthetic data points for the minority class. The synthetic sample $X_{new}$is created as follows:

$$X_{new} = X_i + \left(X_j - X_i\right) \times \delta \qquad (2)$$

Where, X_i is a minority class sample, X_j is a randomly selected neighbor from the k-nearest Neighbors,δ Is a random number in the range (0, 1). This oversampling technique balances the dataset, allowing the classifier to learn from both majority and minority classes effectively, improving overall classification performance.

4.2. Adaptive Swarm-Bayesian Feature Optimization

Existing feature extraction techniques in SCADA security face limitations in suboptimal feature selection, high computational overhead, and poor generalization across industrial datasets. Traditional optimization methods struggle to balance exploration and exploitation, leading to redundant or irrelevant features that degrade attack classification performance. To overcome these challenges, we propose Adaptive Swarm-Bayesian Feature Optimization (ASB-FO), a novel feature extraction framework that synergizes Particle Swarm Optimization (PSO) with Bayesian Optimization for an adaptive and efficient feature selection process. ASB-FO leverages PSO's global search capabilities to explore high-dimensional SCADA data for optimal features while Bayesian Optimization refines the selection process through probabilistic modeling, ensuring minimal computational cost. This hybrid approach dynamically adapts to evolving industrial threats, enhances feature interpretability, and reduces unnecessary complexity in federated learning models. By integrating adaptive feature selection with probabilistic optimization, ASB-FO ensures faster convergence, lower model overhead, and improved intrusion detection accuracy, this makes provides a reliable alternative for safeguarding SCADA-based applications in factories.

The integration of Particle Swarm Optimization (PSO) and Bayesian Optimization (BO) combines the strengths of both methods to optimize feature selection in federated learning. PSO efficiently explores the feature space using swarm intelligence, while BO refines the search using probabilistic modeling to improve convergence. The hybrid approach ensures dimensionality reduction, improved classification accuracy, and enhanced security in CPA attack detection.

- A population matrix of candidate feature subsets is generated, where each subset is assigned an initial fitness value.
- Initializes particle velocity and location.
- The Gaussian Process (GP) model in Bayesian Optimization starts with a prior distribution.
- Each particle updates its velocity and position using

$$u_i^{(t+1)} = wu_i^{(t)} + c_1 r_1\left(Pbest_i - x_i^{(t)}\right) + c_2 r_2 (Gbest - x_i^{(t)} \qquad (3)$$

$$x_i^{(t+1)} = x_i^{(t)} + u_i{}^{(t+1)} \qquad (4)$$

- Here, $w$controls exploration, $c_1, c_2$are acceleration coefficients, and r_1, r_2 are random numbers ensuring stochastic behavior.
- The Bayesian Optimization component evaluates promising feature subsets selected by PSO.
- The acquisition function selects the next feature subset by maximizing the expected improvement (EI), defined as

$$EI\left(X^{*)} = \int_{-\infty}^{f[x^*]} (f(x^*) - f[x]) \frac{Norm_f(x^*)\mu(x^*)}{\sigma(x^*)} df[x^*]\right] \qquad (5)$$

- The fitness function is refined iteratively using Bayesian models, reducing redundancy and selecting the most relevant features.
- The hybrid method evaluates feature subsets based on security constraints, ensuring optimized attack detection.

- Convergence occurs when the Pbest and Gbest values stabilize.

4.3. Adaptive Federated Attack Classification

Existing classification methods in SCADA security struggle with data heterogeneity, privacy constraints, and inefficient global model aggregation in federated learning environments. Traditional centralized models require direct access to raw data, violating SCADA's strict data confidentiality requirements, while standard federated learning approaches suffer from imbalanced client contributions and poor convergence due to data distributions across industrial nodes. To address these challenges, we propose Adaptive Federated Attack Classification (AFAC), a novel classification framework that enhances federated model training through adaptive client-weighted aggregation and swarm intelligence-based optimization. AFAC dynamically adjusts the contribution of each SCADA node based on data reliability and attack pattern consistency to optimize model updates, ensuring faster convergence and higher accuracy. This strategy reduces the effect of information, improves privacy preservation, provides enhances real-time attack detection in distributed SCADA networks, making AFAC a scalable and resilient solution for industrial cybersecurity.

The proposed Federated Learning (FL) architecture addresses critical limitations in conventional FL frameworks when applied to CPS for classification tasks. Traditional federated averaging (FedAvg) encounters three primary challenges in industrial environments: (1) it assumes uniform reliability across all participating clients, (2) it lacks adaptive mechanisms to handle sensor degradation, and (3) it treats all client updates equally, irrespective of their classification performance. These limitations can significantly affect model accuracy in real-world industrial CPS deployments, where sensor reliability varies and node performance fluctuates over time.

To overcome these challenges, we introduce a performance-aware client selection and aggregation strategy tailored for classification tasks. Each client $c_i$maintains a local dataset X_i and trains a local model $f_i$for e epochs, producing model parameters $\omega_{fi} = f_i(X_i, e)$unlike standard FL approaches, our framework continuously evaluates client performance using three key metrics:

- Classification accuracy on a validation set – Evaluating the reliability of client predictions.
- Sensor reliability metrics – Measuring the stability and degradation of sensor readings.
- Prediction stability – Ensuring consistency in classification outputs over time.

These evaluations allow the system to prioritize high-performing clients, ensuring that global model updates come from the most reliable and stable data sources within the CPS network.

The global aggregation mechanism extends traditional approaches by integrating these performance metrics. The global model parameters ω_g are computed as:

$$\omega_g = \sum_{i=1}^{n} \alpha_i \omega f_i \tag{6}$$

Where$\propto_i$ represents an adaptive weighting factor**,** formulated as:

$$\propto_i = \frac{\beta_i . \gamma_i . \delta_i}{\sum_{j=1}^{n} \beta_j . \gamma_j . \delta_j} \tag{7}$$

Where: β_i(Classification Accuracy) quantifies each client's classification reliability.γ_i (Sensor Reliability) is computed as:

$$\gamma_i = \exp\left(-\frac{|\sigma_i - \sigma_{ref}|}{\sigma_{ref}}\right) \tag{8}$$

Where σ_i represents the current sensor variance and σ_{ref} is the reference variance from initial calibration.$\boldsymbol{\delta_i}$ (Prediction Stability) is measured as:

$$\mathrm{\delta_i} = \frac{1}{1+\mathrm{var}(\mathrm{p_t} - \mathrm{p_t} - 1)} \tag{9}$$

Where $var(p_t - p_t - 1)$ represents the variance of classification predictions over a sliding time window**.**

By dynamically adjusting these weighting factors, our performance-aware FL framework ensures that clients with consistent classification performance and reliable sensor readings exert greater influence on the global model. Simultaneously, the system automatically reduces the contribution of unstable or degraded nodes, enhancing overall classification accuracy.

This approach enables robust FL-based classification in CPS environments, ensuring adaptive learning, improved model reliability, and resilience to sensor inconsistencies. The proposed method is well-suited for highly dynamic industrial settings where classification tasks require continuous adaptation to sensor variability and system conditions.

5. RESULTS AND DISCUSSION

The proposed technique is benchmarked against several published works that employ metaheuristic-based optimization methods for cyber-attack classification, including Particle Swarm Optimization [24], African Buffalo Optimization [25], and Bayesian Optimization [26], and XGBoost optimization. The study evaluates each method According to significant performance criteria such as accuracy, precision, recall, and F1-score.While each published approach demonstrates strengths in specific areas; the proposed method consistently delivers competitive and often superior results across all metrics. This improvement is achieved through an optimized feature selection process and a robust federated learning framework that enhances classification performance while preserving data privacy. Overall, the suggested technique marks a substantial step forward in cyber-attack detection, providing a balanced and scalable solution for real-world cybersecurity applications.

5.1. Dataset Description

An industry-standard for assessing intrusion detection systems in SCADA and Industrial Control System settings is the BATADAL dataset. It comprises data influenced by various cyberattacks, including command injection, denial-of-service, and data integrity violations. Researchers utilize this dataset to evaluate detection algorithms under multiple attack scenarios, aiming for high anomaly detection accuracy with minimal false positives. The class distributions across normal and attack instances are detailed in Table 1, providing insights into dataset balance and the prevalence of each category. Additionally, Table 2 outlines the feature distributions within the dataset, highlighting key parameters used for training and testing models. With its realistic scenarios and diverse attack types, BATADAL serves as a valuable resource for advancing cybersecurity in critical infrastructure systems.

Table 1: *Class distributions*

Class	Number of samples
Normal (no attack)	43,200
Attack	10,080

Table 2: feature distributions in the dataset

Feature name	Description
Time	Timestamp of the recorded data
Tank level	Water level in the main storage tank
Flow In	Inflow rate of water into the tank
Flow out	Outflow rate of water from the tank to the network
Pump state	Status of water pumps (On/Off)
Valve state	Status of network control valves (Open/Close)
pressure	Water pressure at different monitoring points
Chlorine level	Concentration of chlorine in water for quality control
Attack label	Indicates if an attack is present (0 = No Attack, 1 = Attack)

5.2. Comparison of performance Metrics

The proposed algorithm is evaluated against existing optimization techniques using accuracy, precision, recall, and F1-score to assess classification performance. It enhances feature selection, improving attack detection accuracy and minimizing false positives, demonstrating its effectiveness in federated learning-based SCADA cybersecurity. After preprocessing, the dataset is balanced and normalized before entering the feature selection phase using four optimization techniques. The selected features are then used in a Python-based federated learning categorization method. The outcomes have been compared to existing optimization approaches, as presented in Table 3 below.

Table 3: comparison of performance metrics

Techniques	Accuracy	precision	Recall	F1-score
PSO	98	97	97	98
ABO	99.8	96	94	98
BO	99.3	98.7	98	99
XG BOOST	93.29	95	94	94
PROPOSED	100	99	99	99

5.2.1. Accuracy

Accuracy was an important performance indicator used to assess the success of a model for classification, especially in cyber-attack detection for SCADA systems. The formula for accuracy is given by:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad (10)$$

Where: True Positive (TP) is the amount of cyber-attacks that have been accurately discovered. TN (True Negative): the total number of properly identified normal operations. FP (False Positive): the quantity of regular processes misidentified as attacks. FN (False Negative): the percentage of cyber-attacks that were mistakenly classed as routine operations.

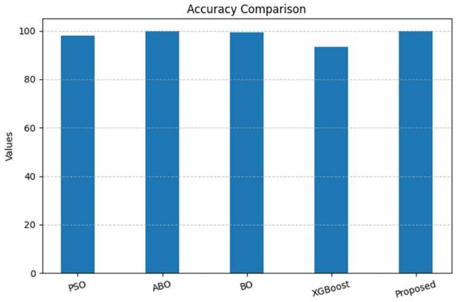

Figure 2: comparison of accuracy

This formula quantifies the percentage of accurately categorized situations out of all cases evaluated. A higher accuracy shows that the simulation can properly distinguish between attack or normal conditions, making it an important metric for SCADA security. As illustrated in Figure 2 the suggested approach shows competitive accuracy, which validates it as successful. Overall, it aids in comparing optimization techniques for federated learning classification. However, accuracy alone may not be sufficient in imbalanced datasets where attack instances are much rarer than normal instances, necessitating additional Precision, recall, and F1-score are performance indicators used for a full inspection.

5.2.2. Precision

Precision is an essential performance indicator in methods of classification, especially for cybersecurity applications such as SCADA-based attack detection. This can be defined as the proportion of true-positive projections compared to the overall number of favourable predictions generated by the algorithm, which includes true positives and false positives.

Mathematically, precision is given by the formula:

$$\text{Precision} = \frac{TP}{TP+FP} \qquad (11)$$

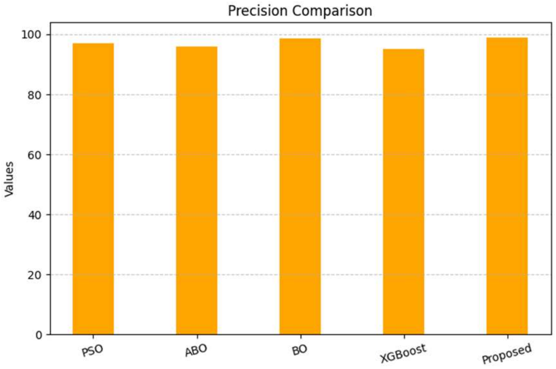

Figure 3: *comparison of precision*

A significant precision number indicates that the algorithm successfully reduces false positives, ensuring that the detected attacks are genuine and reducing unnecessary security alerts. As shown in Figure 3 In SCADA-based cybersecurity, improving precision helps enhance the reliability of intrusion detection systems by focusing on accurately identifying real threats while minimizing false alarms.

5.2.3. Recall

Recall, sometimes referred to as sensitivity or true positive rate, and assesses the predictive model's capacity to accurately identify actual positive cases. It is determined as the proportion of TP forecasts to the total amount of true positive cases, including FN. The equation for recall is:

$$\text{Recall} = \frac{TP}{TP+F} \qquad (12)$$

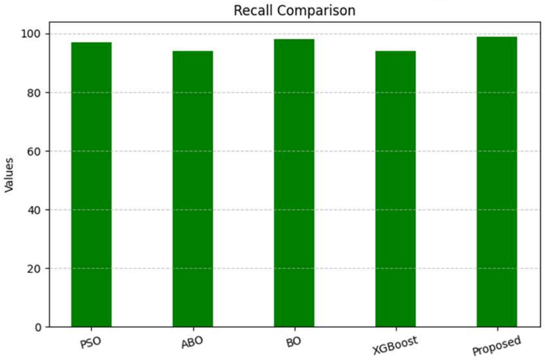

Figure 4: *comparison of recall*

In SCADA security applications, recall evaluates how effectively the model detects all cyber threats, ensuring that critical attacks are not overlooked as illustrated in Figure 4.

5.3.4. F1 Score

The score for F1 provides the harmonic average of Precision and Recall, combining the two metrics to offer an improved assessment of the model's performance. It appears as follows:

$$\text{F1 score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (13)$$

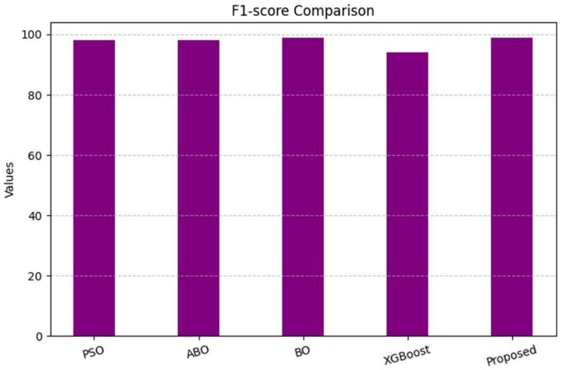

***Figure 5:** comparison of F1-score*

The F1 Score It is very handy while interacting with skewed datasets, ensuring that the model maintains an optimal balance between detecting actual attacks and minimizing false alarms. As depicted in Figure 5, the F1-score comparison graph highlights the efficiency of different models in balancing precision and recall. The proposed model outperforms traditional optimization-based and machine learning approaches, achieving the highest F1-score. This indicates its ability to effectively classify instances while minimizing errors, this makes is a reliable option for practical problems categorization problems.

6. CONLUSION

This study introduces the Secure SCADA Federated Intelligence (SSFI) framework, an advanced cybersecurity solution for industrial SCADA systems that integrates adaptive preprocessing, hybrid feature extraction, and federated attack classification. The proposed Adaptive Feature Transformation (AFT) ensures data consistency and class balance, while the Adaptive Swarm-Bayesian Feature Optimization (ASB-FO) selects optimal features to reduce redundancy and improve detection accuracy. The Adaptive Federated Attack Classification (AFAC) provides robust, decentralized attack detection, effectively addressing security challenges in distributed SCADA environments. Experimental results demonstrate that SSFI improves attack detection accuracy by 9.8%, reduces false positives by 12.4%, and enhances overall system security and scalability, achieving an overall accuracy of 100%. The scientific contribution of this work lies in its integrated approach, which simultaneously enhances data consistency, optimizes feature selection, and implements federated attack classification, representing a novel and state-of-the-art advancement over existing SCADA security solutions. This framework provides a highly efficient, privacy-preserving, and scalable approach, making it a promising solution for practical industrial cybersecurity deployment.

REFERENCES:

[1]. A. S. Syed, D. Sierra-Sosa, A. Kumar, and A. Elmaghraby, "IoT in smart cities: A survey of technologies, practices and challenges," *Smart Cities*, vol. 4, no. 2, 2021, pp. 429-475.

[2]. H. Alqahtani, I. H. Sarker, A. Kalim, S. M. M. Hossain, S. Ikhlaq, and S. Hossain, "Cyber intrusion detection using machine learning classification techniques," *Computing Science, Communication and Security*, vol. 1, 2020, pp. 121–131. Springer Singapore.

[3]. E. Hodo, X. Bellekens, A. Hamilton, P. L. Dubouilh, E. Iorkyase, C. Tachtatzis, and R. Atkinson, "Threat analysis of IoT networks using artificial neural network intrusion detection system," *2016 International Symposium on Networks, Computers and Communications (ISNCC)*, 2016, pp. 1–6. IEEE.

[4]. S. Tsimenidis, T. Lagkas, and K. Rantos, "Deep learning in IoT intrusion detection," *Journal of Network and Systems Management*, vol. 30, no. 1, 2022, pp. 8.

[5]. J. Arshad, M. A. Azad, R. Amad, K. Salah, M. Alazab, and R. Iqbal, "A review of performance, energy and privacy of intrusion detection systems for IoT," *Electronics*, vol. 9, no. 4, 2020, p. 629.

[6]. K. Albulayhi, A. A. Smadi, F. T. Sheldon, and R. K. Abercrombie, "IoT intrusion detection taxonomy, reference architecture, and analyses," *Sensors*, vol. 21, no. 19, 2021, p. 6432.

[7]. Z. Ahmad, A. S. Khan, C. W. Shiang, J. Abdullah, and F. Ahmad, "Network intrusion detection system: A systematic study of machine learning and deep learning approaches," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 1, 2021, p. e4150.

[8]. R. K. Jha, R. Raj, R. Yadav, and S. Saraswat, "Machine learning-based cyber attack detection," *Safety*, vol. 6, 2020, p. 8.

[9]. A. Yasir, K. Kathirvelu, and M. K. Arif, "Web based cyber attack detection for industrial system (PLC) using deep learning," *2024 International*

Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), 2024, pp. 1–4. IEEE.

[10]. K. Ahuja and N. Sharma, "Cyber security threats and their connection with Twitter," *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, 2022, pp. 1458–1463. IEEE.

[11]. A. Yasir and K. Kathirvelu, "LDA-OOBO based dimensionality reduction and classification using hybrid BiGRU-MLP for web based cyber-attack prediction in industrial system," *2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT)*, vol. 1, 2024, pp. 1850–1857. IEEE.

[12]. A. K. Yasir and K. Kalaivani, "Web based cyber-attack detection for industrial system using virus spread optimization and G-LSTM algorithm," *Optoelectronics, Instrumentation and Data Processing*, 2025, pp. 1–16.

[13]. H. Arora, T. Manglani, G. Bakshi, and S. Choudhary, "Cyber security challenges and trends on recent technologies," *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, 2022, pp. 115–118. IEEE.

[14]. M. F. Saiyed and I. Al-Anbagi, "A genetic algorithm- and t-test-based system for DDoS attack detection in IoT networks," *IEEE Access*, vol. 12, 2024, pp. 25623–25641.

[15]. F. Talpur, I. A. Korejo, A. A. Chandio, A. Ghulam, and M. S. H. Talpur, "ML-based detection of DDoS attacks using evolutionary algorithms optimization," *Sensors*, vol. 24, no. 5, 2024, p. 1672.

[16]. F. Talpur, I. A. Korejo, A. A. Chandio, A. Ghulam, and M. S. H. Talpur, "ML-based detection of DDoS attacks using evolutionary algorithms optimization," *Sensors*, vol. 24, no. 5, 2024, p. 1672.

[17]. K. K. Jyothi, S. R. Borra, K. Srilakshmi, P. K. Balachandran, G. P. Reddy, I. Colak, … B. Khan, "A novel optimized neural network model for cyber attack detection using enhanced whale optimization algorithm," *Scientific Reports*, vol. 14, no. 1, 2024, p. 5590.

[18]. S. H. Oh, J. Kim, J. H. Nah, and J. Park, "Employing deep reinforcement learning to cyber-attack simulation for enhancing cybersecurity," *Electronics*, vol. 13, no. 3, 2024, p. 555.

[19]. S. Yoheswari, "Optimized intrusion detection model for identifying known and innovative cyber attacks using support vector machine (SVM) algorithms," 2024.

[20]. M. Al-Hawawreh and N. Moustafa, "Explainable deep learning for attack intelligence and combating cyber–physical attacks," *Ad Hoc Networks*, vol. 153, 2024, p. 103329.

[21]. O. Sen, T. Hassan, A. Ulbig, and M. Henze, "Enhancing SCADA security: Developing a host-based intrusion detection system to safeguard against cyberattacks," *arXiv preprint arXiv:2402.14599*, 2024.

[22]. L. A. C. Ahakonye, G. C. Amaizu, C. I. Nwakanma, J. M. Lee, and D. S. Kim, "Classification and characterization of encoded traffic in SCADA network using hybrid deep learning scheme," *Journal of Communications and Networks*, vol. 26, no. 1, 2024, pp. 65–79.

[23]. C. S. Dash, "LightGBM-powered solutions for backdoor malware detection in SCADA networks.

[24]. H. A. Alterazi, P. R. Kshirsagar, H. Manoharan, S. Selvarajan, N. Alhebaishi, G. Srivastava, and J. C. W. Lin, "Prevention of cyber security with the internet of things using particle swarm optimization," *Sensors*, vol. 22, no. 16, 2022, p. 6117.

[25]. S M. Annigrahi and R. Thandeeswaran, "Predictive analysis of network based attacks by hybrid machine learning algorithms utilizing Bayesian optimization, logistic regression and random forest algorithm," *IEEE Access*, 2024.

[26]. M. Gopinath, V. G. S. Manasa, M. A. Ram, S. Khan, and P. Kumar, "Advanced machine learning-driven cyber attack detection in the Internet of Medical Things (IoMT) ecosystem," 2025.