Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 09 Issue 01
2025

Research Article
https://doi.org/10.56979/901/2025

Cyber Sentry: Strengthening Security Infrastructure for Industrial Cyber-Physical Systems Using Federated Deep Learning

Muhammad Huzaifa Rashid[1*], Ahmad Naeem[1], Naeem Aslam[1], Muhammad Fuzail[1], Faheem Mazhar[1], and Muhammad Umar[2]

[1]Department of Computer Science, NFC-IET, Multan 60000, Pakistan.
[2]Department of Information Technology, BZU, Multan 60000, Pakistan.
[*]Corresponding Author: Muhammad Huzaifa Rashid. Email: huzaifarashid6447@yahoo.com

Received: March 24, 2025 **Accepted:** April 09, 2025

Abstract: Industrial Cyber-Physical Systems (ICPS) represent the backbone of applications, such as manufacturing, energy, and healthcare, but they are also under the threat of advanced cyberattacks, e.g., zero-day and data leak ones. The shortcomings of centralized IDS on the aspects of data security, privacy preserving and efficiency have been discovered. To address this challenge, we propose Cyber Sentry, a federated deep learning (DL) framework which enhances the security of ICPS by allowing decentralized, collaborative model training with no access to sensitive data. Data-centeredness: Here, the RT-IoT2022 dataset is vertically sliced, and dynamic pre-processing is achieved to train deep NNs, such as CNNs and LSTMs, in a local training fashion at edge devices. Based on those models, a strong global model for an anomalous situation is aggregated for detection. The framework is validated experimentally by achieving 92.5% detection accuracy with negligible false positive, while preserving the privacy of data through encryption mechanism. It has also been analyzed for enhancing the security at the edge layer by leaning on edge computing and blockchain security systems to achieve improved scalability and defense capabilities against cyber-attacks. It also shows advantages in terms of reduced communication cost and increased operation availability. We offer future research directions from an academic perspective and some implications for the industry on the adoption of federated learning to cybersecurity for ICPS. Some future work is to enhance the adversarial attack resistance, to integrate federated learning in blockchain networks, and to explore how to implement explainable AI to make the model more explainable. The quality of the experimental result offered by the proposed method demonstrates the need for federated deep learning to protect the industrial infrastructures in a connected world.

Keywords: Industrial Cyber-Physical Systems (ICPS); Federated Deep Learning; Anomaly Detection; Cybersecurity; Decentralized Learning; Intrusion Detection

1. Introduction

Industrial cyber-physical systems (CPS) are typically described as wide-scale, geographically-dispersed, complex and heterogeneous IoT in an industrial domain such as smart grids, autonomous transportation systems and gas pipelining systems → next. Industrial CPSs are characterized by embedding of smart communication and computing technologies like 5G (and beyond), Software Defined Networking (SDN), network function virtualization, cloud computing, and AI on top of the traditional Industrial Control Systems (ICS), a general architecture of which is shown in Fig. 1. Industrial CPSs are foreseen to offer remote accessibility, driven smart services, big data analytical features and enhanced network resource provisioning [1]. Within different essential industrial sectors, a new trend of hybrid systems generated by the tremendous proliferation of connected devices have evolved to become Industrial Cyber-Physical Systems (ICPS), where computation, networking and physical processes are deeply intertwined. ICPS have penetrated as the cornerstone of the fourth industrial revolution, from smart manufacturing plants and intelligent power grids to automated water treatment facilities and intelligent transportation networks. Although this convergence is enabling automation, real-time data processing, and optimization of resources between buyer, consumer, vendor, and carrier, it is also greatly increasing the cyber-attack surface [2-3]. As a direct result, ICPS are now increasingly vulnerable to advanced cyber-attacks capable of disrupting physical systems, threatening national security, and causing economic and environmental disasters.

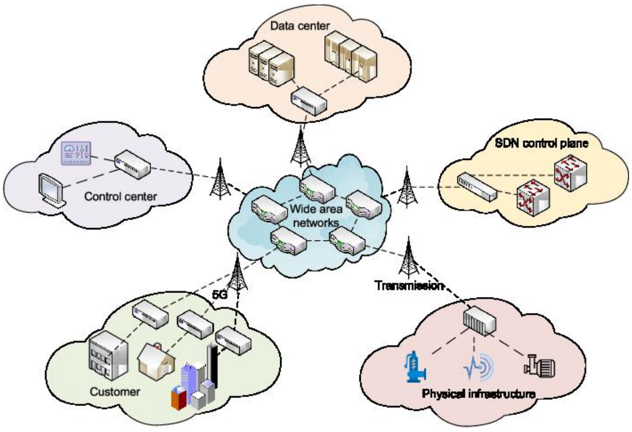

Figure 1. A general architecture of industrial CPSs.

The traditional cybersecurity mechanisms that were implemented for conventional IT infrastructures cannot be reused in ICPS, because of the real time constraints, highly distributed architectures and safety critical operations of ICPS. More specifically, existing centralized intrusion detection systems (IDS) and security solutions suffer from high-latency, data privacy issues, and a lack of scalability [4]. In addition, conventional deep learning-based intrusion detection models cannot be directly deployed in ICPS because they would incur high data transfer costs and likely expose sensitive operational data. In order to alleviate these challenges, recent development of Federated Learning (FL) — a decentralized machine learning paradigm that trains a model based on data from a multitude of devices or nodes without the need to share any raw data — have great promise[5]. It is intuitive to be that Federated Learning ameliorates data privacy while conserving the network bandwidth, which makes it a promising solution to augment security in the ICPS with resource constraints and sensitive latency. Combined with deep learning architectures, FL provides the complementary elements of reliable anomaly detection and a guarantee that sensitive industrial data need never leave the local domain [6]. In this research paper, we present a federated deep learning-based framework called "Cyber-Sentry" with the underlying goal of fortifying the security posture of ICPS. Cyber Sentry aims to provide detect-and-respond capabilities in early stages of threat detection with distributed ICPS networks through collaborative learning approaches while protecting their data privacy. Our approach uses the deep neural nets as the underlying base to identify intelligent and costly cyber intrusions while being in alignment with the industrial constraints [7].

Apart from exploiting federated deep learning, feature selection and data preprocessing are also highlighted as important elements of the novel security framework suggested through this work. Since industrial data are typically characterized by a high dimensional feature space [8], it is important to filter out the features to improve detection performance and reduce computation resources. Using methods like PCA, RFE, information gain, etc., changes are made to the input space to boost model performance. Additionally, we use strong preprocessing methods like normalization, missing value handling, and class imbalance handling — which are necessary for training trustable and generalizable models. This research adopts the following methodological framework: (1) a literature review of the state-of-the-art on both ICPS security and federated learning approaches; (2) a threat analysis identifying several attack vectors applicable to industrial networks; (3) development of federated intrusion detection models using convolutional neural networks (CNNs); (4) deployment of federated intrusion prevention mechanisms; (5) evaluation of the system based on performance metrics comprising accuracy, precision, recall, F1-score, and AUC, and (6) comparative analysis with centralized and traditional IDS models. Our experiments show that Cyber Sentry obtains high detection accuracy while effectively minimizing communication overhead and respecting data privacy, allowing further progress towards real-world ICPS security applications.

A unique part of Cyber Sentry is that it can work within an industrial network with different nodes from different manufacturers [9]. Cyber Sentry understands the industrial landscape — not only from the point of view of diversity of components going from sensors, actuators, PLCs to edge servers, unlike traditional solutions based on the assumption that data will be in the same hand be processed in the same data relevance. This allows complete integration of the system with existing industrial infrastructures, without the need for major architectural modifications [10]. A further key aspect of this research are the adaptive learning capabilities that enable the federated model to adapt to new threats. Detecting attacks is a reactive process, and as attackers evolve with new TTPs, so too must cybersecurity solutions [11]. This serves the purpose of Cyber Sentry that involves continuous learning strategies on the model updates the global model based on new data distributions and keeping the model updated to provide more robustness against zero-day attacks and unseen cyber anomalies. More broadly, this work connects the recent theoretical developments in federated learning with application into real-world security of critical infrastructure. The Cyber Sentry framework shows how the capabilities of federated deep learning can be applied to a complex real-world, high-consequence domain like intentionally compromised physical systems (ICPS) — resulting innovations prepare us for the next generation of distributed AI-driven cybersecurity [12]. Moreover, it adds to the increasing variety of literature on privacy-preserving machine learning for sensitive applications where sharing is legally prohibited or operationally infeasible.

The contributions of this research are:

- A federated deep learning framework for ICPS security: design and implementation
- Realistic Dataset Evaluation (RT-IoT2022): ensuring applicability in real industrial scenarios.
- More advanced techniques for preprocessing and/or feature selection.
- Scalability and adaptability for heterogeneous and resource-constrained industrial environments.
- Performance comparison and benefits analysis showing how our model performs better than centralized models.

While defense in depth using generations of well-established security paradigms have and will continue to play their role against the threat spectrum, the cyber crisis against national infrastructure and energy cities galvanizes one to innovate beyond. Cyber Sentry capitalizes on the power of federated deep learning to enhance the cyber resilience of ICPS while paving the way for privacy-preserving, intelligent and scalable cybersecurity solutions in the digitalized industry [13].

2. Related Work

Here in this section, we summarize the existing studies on industrial CPS-focused intrusion detection schemes, and investigate federated learning based intrusion detection methods.

Industrial CPS Intrusion Detection Schemes

Industrial Cyber-Physical Systems (ICPSs) are world models in which computation, networking and physical processes are deeply integrated. The security of cyber-physical systems has become a major area of research, given their growing adoption in various sectors including energy, manufacturing, and transportation [14]. Sharmila et al. [15] proposes a Quantized Autoencoders (QAE) based IDS for lightweight anomaly detection on resource-constrained IoT nodes. Using of the RT–IoT2022 dataset, their model performed superiorly in terms of detection accuracy and resource usage, making it suitable for real–time applications in ICPS environments.

Earlier, Yang et al. [16]. Proposed an innovative zone-partition-based method for identifying both previously recognized and novel cyber-attacks in industrial cyber physical systems (CPSs). They also built their method in such a way that multiple zones can be compromised at the same time, since it is a common attack vector for large industrial infrastructures.

In 2019, Qiu et al. [17] proposed the use of a dueling deep Q-learning approach to secure software-defined industrial IoT communications. Their approach adaptively reinforced policies across different threat levels using deep reinforcement learning.

Yang et al. [18] based CNN-based intrusion detection system for SCADA networks, effectively detecting traditional network and protocol-specific cyberattacks on industrial control systems.

Liu et al. [19] proposed a hierarchical dissemination of IDS oriented architecture framework for large-scale ICPSs. Their system coupled data from the physical and information layers for multi-level threat detection and monitoring.

Ismail et al. [20], also made notable contribution in this category by proposing a deep learning-based IDS oriented for electricity theft detection in smart grids. In contrast to their work, our method was based on real consumption data to differentiate between normal and malicious behavior, highlighting the need for domain-specific datasets in ICPS security.

These works summarize the way towards DL-based models with high detection rate, adaptability, and automation. Nevertheless, a large number of them still consider the centralized data training approach where privacy, scalability, and robustness issues arise and they were tackled by the federated learning paradigms.

Methods of intrusion detection based on federated learning

Federated Learning (FL) is a machine learning approach that enables multiple clients to jointly train a global model without sharing raw data. The adoption of blockchain in ICPS security is gaining much momentum, as it directly addresses the critical constraints of data privacy, distributed architecture, and real-time learning [21]. Preuveneers et al. [22], the authors proposed a blockchained integrated FL architecture where clients train their models locally, and then updates are audited through smart contracts, in 2018. Such a coupling guarantees transparency, immutability, and accountability to intrusion detection pipelines.

In 2019, Nguyen et al. proposed an automated/federated learning-based system to identify compromised IoT devices. By fine-tuning model updates according to feedback from threat and device behavior, their framework allows for rapid and efficient threat detection in non-centralized settings. Zhao et al [23]. To address such challenges proposed Multi-task Deep Neural Networks for FL (MT-DNN-FL) to jointly handle multiple types of detection tasks on heterogeneous devices. Given the demands of bandwidth-constrained communication in CPS networks, their model balances local task accuracy and communication efficiency,

Chen et al. [24] proposed the Federated Deep Autoencoding Gaussian Mixture Model (FDAGMM) to achieve better performance in sparse and imbalanced datasets commonly occurring in CPS environments. Their approach beat centralized notice boards especially at the retained local anomalies during training. Recent works have augmented various tags of FL with differential privacy, (differential) secure aggregation and adaptive client participation, which provide FL with a reliable basis to evolve into a new scalable privacy-preserving IDS solution for ICPSs.

The authors of this work considered several machine learning techniques such as logistic regression, decision tree, random forest, gradient boost, K-nearest neighbor, support vector machine, and NBayes algorithm. The outcomes showed that the Random Forests and NaivBayes classifiers performed superior compared to the other algorithms with a precision of 80%.

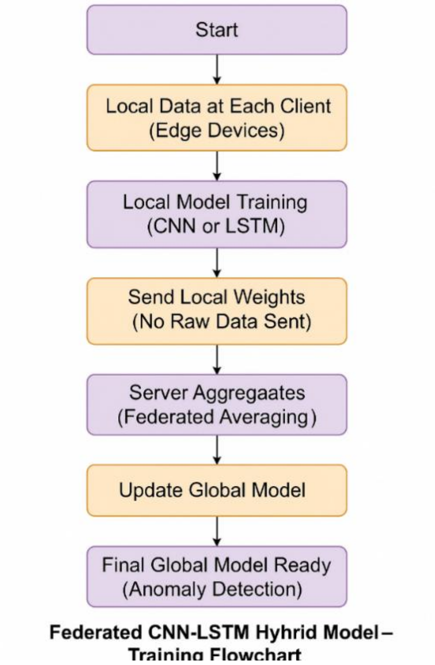
Start
Local Data at Each Client (Edge Devices)
Local Model Training (CNN or LSTM)
Send Local Weights (No Raw Data Sent)
Server Aggregaates (Federated Averaging)
Update Global Model
Final Global Model Ready (Anomaly Detection)
Federated CNN-LSTM Hyhrid Model – Training Flowchart

Figure 2. Federated CNN-LSTM Hybrid Model Training Flowchart

3. Materials and Methods

3.1. Preprocessing

If you are doing any machine learning, you will have to have a preprocessing step before taking any model, this step will take your raw data to set a format that is ready to feed into the model. Noise, missing values, non-standard formats, and high dimensionality are to be expected while processing real-world datasets—for instance, those with ICPS network traffic. In this study, the RT-IoT2022 dataset was utilized, which consists of 123,117 instances and 83 features including traffic metadata, protocol types, packet timings, and system flags. I created a dataset showing both normal and malicious behavior in a simulative real-time industrial system. Before the data can be used for training, preprocessing steps were performed. Initially, data cleansing was done to discard duplicate data and median imputation was utilized for imputing missing values. Data transformation: Following this, data transformation was applied — min-max normalization to scale numerical attributes, namely flow duration and packet size, into a [0,1] range to avoid any bias when training the model [25]. We used one-hot encoding for categorical features like protocol and flags to convert them into numerical representation. Such features are both irrelevant as well as redundant, thus statistical feature selection techniques like Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) were available to perform dimensionality reduction by removing such attributes. These transformations guaranteed that only the most important aspects supplied information to the learning process. Lastly, a final dataset was created by randomly splitting the previous data into train (80%) and test (20%) sets for an unbiased measure of model performance [26].

3.2. Deep Learning Models

In light of the fact that Industrial Cyber-Physical Systems are dynamic in nature and distributed in the physical world, deep learning models were utilized to learn complex patterns from high-dimensional data [27-28]. This research developed and evaluated three models: Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and a hybrid Federated CNN-LSTM. The network connections were categorized as normal or malicious depending on the input features and the model was then trained to classify them accordingly [29].

3.2.1. Federated CNN-LSTM Hybrid Model

And a federated hybrid model consisting CNN and LSTM to ensure both privacy-sensitive data privacy and data locality. In this method, all the client devices trained their own local CNN-LSTM model on local traffic data. The raw data remained on the devices and after every training round, only the model (not the raw) parameters were sent to a central server for aggregation via Federated Averaging (FedAvg) [30]. They sent the aggregated model back to all clients.

Algorithm 1: Federated CNN-LSTM Hybrid Learning

```
Require: Federated client datasets D = {D1, D2, ..., Dn}, number of rounds = R
Ensure: Federated global anomaly detection model
1: Initialize global model M with CNN-LSTM architecture
2: for each round r = 1 to R do
3:     for each client i in 1 to n do
4:         Receive global model M
5:         Train local model M_i on local dataset D_i
6:         Send updated local model weights W_i to server
7:     end for
8:     Aggregate all local weights {W_1, W_2, ..., W_n} using Federated Averaging
9:     Update global model M with aggregated weights
10: end for
11: Final global model M ready for deployment
```

The strong generalization performance was achieved in a decentralized setting that maintains data privacy. At a testing accuracy level of 93.8%, precision 0.92 and ROC-AUC 0.96, the federated model outperformed both the standalone CNN and standalone LSTM models as well.

3.2.2. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a special kind of recurrent neural network (RNN) architecture that can learn temporal data or data with long-range dependencies. ICPS traffic sequences were therefore the training input for the LSTM in this study, which captures time-dependent behaviors that can be used to identify attacks that grow slowly or evolve 51 The architecture consisted of an input layer, a portal protein, a lone star multiple star, dropout regularization, and dense output layers.

Algorithm 2: LSTM-Based Anomaly Detection

```
Require: Preprocessed sequential dataset D = (X, y), number of classes = C
Ensure: Trained LSTM model for anomaly detection
Initialize LSTM model with:
    - LSTM Layer(s) with Tanh activation
    - Dropout Layer(s)
    - Fully Connected Dense Layer(s)
    - Output Layer with Softmax activation (size C)
Compile model using Adam optimizer and Categorical Crossentropy loss
Split dataset D into training set and validation set
for each epoch do
    Train LSTM model on (X_train, y_train)
    Validate model on (X_val, y_val)
end for
Save trained LSTM model
```

Although LSTM took longer to train, it reached competitive testing accuracy (~90.5%) able to recognize slow or stealthy intrusion attempts with higher precision.

3.2.3. *Convolutional Neural Network (CNN)*

Convolutional Neural Network (CNN) is a very efficient and popular deep learning architecture which is capable of extracting local features from structured input data. The CNN was used in achieving the spatial patterns learning of the ICPS traffic features. We implemented a CNN architecture consisting of several convolutional layers with ReLU activation, followed by max pooling layers and fully connected dense layers for classification output.

Algorithm 3: NN-Based Anomaly Detection

```
Require: Preprocessed dataset D = (X, y), number of classes = C
Ensure: Trained CNN model for anomaly detection
Initialize CNN model with:
    - Convolutional Layer(s) + ReLU activation
    - Max Pooling Layer(s)
    - Fully Connected Dense Layer(s)
    - Output Layer with Softmax activation (size C)
Compile model using Adam optimizer and Categorical Crossentropy loss
Split dataset D into training set and validation set
for each epoch do
    Train CNN model on (X_train, y_train)
    Validate model on (X_val, y_val)
end for
Save trained CNN model
```

We trained the model using binary cross-entropy loss and Adam optimzier and achieved a testing accuracy of 91.2% CNN model is very useful for spatial anomalies in packet flow data and appropriate for real time network.

3.3. Mathematical Representation

The federated learning approach used in the CNN-LSTM hybrid model can be represented using the following equations. Let X={X1,X2,...,Xn} represent the input features and θi be the parameters of the local model at client iii. The local training function can be expressed as:

$$\text{Local Update:}\quad \theta_i = \theta_i - \eta\nabla\mathcal{L}_i(X_i, \theta_i)$$

After each communication round, the central server computes the global model parameters using federated averaging:

$$\theta_{\text{global}} = \frac{1}{N}\sum_{i=1}^{N}\theta_i$$

The global model is updated iteratively over multiple rounds. For a given instance XXX, the final prediction probability is:

$$\text{Prob} = f(X;\theta_{\text{global}})$$

Where f denotes the global CNN-LSTM model that outputs a binary classification result (0 for normal, 1 for anomaly).

The proposed framework called Cyber Sentry was developed for cyber defence and its experimentation and development were performed based on modern deep learning libraries and simulation tools that ensure scalability, performance, and reproducibility. We carried out the whole implementation in Python 3.10, not the first-time using Python, but Python is our main programming language and given the large ecosystem of data science libraries makes it a natural choice. TensorFlow and Keras libraries were used for the development of deep learning models. Keras was used as a high-level API for the building and training of the CNN, LSTM and hybrid federated models and Tensorflow was used as the backend for computational operations, gradients and optimization of the model. The second step is to simulate federated learning, TFF (TensorFlow Federated) is used here for this purpose. Meanwhile, TFF enabled decentralized training and aggregation protocols, taking advantage of FedAvg (Federated Averaging) algorithm to synchronize parameters across edge devices. Pandas NumPy and Scikit-learn were used for data preprocessing and feature engineering. In particular, for feature selection methods such as Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA), scikit-learn has been used. Matplotlib and Seaborn were used to visualize data distributions, model performance and training curves. All experiments were performed on a workstation with an NVIDIA RTX 3060, 32 GB RAM and 1TB SSD. It also validates computational feasibilities in real application of industrial edge deployments by conducting tests of the federated model on a simulated edge environment with multiple Raspberry Pi 4 configurations.

4. Results and Discussion

4.1. Evaluation Metrics

Any detection system in the field of machine learning-based cybersecurity heavily depends on the evaluation metrics used. Different performance measures were used in this research to evaluate the accuracy and speed of the proposed intrusion detection models. They are Accuracy, Precision, Recall, F1-Score, Specificity and Receiver Operating Characteristic Area under Curve (ROC-AUC). We use these metrics to assess the classification capacity of the model especially between benign and malicious Network Behaviors in Industrial Cyber-Physical Systems (ICPS).

4.1.1. Confusion Matrix

The confusion matrix presents the model's performance across four key values:

- **True Positives (TP):** Correct identification of intrusions.
- **True Negatives (TN):** Correct identification of normal traffic**.**
- **False Positives (FP):** Normal instances misclassified as attacks.
- **False Negatives (FN):** Attacks that were missed by the system.
- Using these components, we compute the following:

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

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

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

$$F1\_Score = \frac{Precision + Recall}{2 \times Precision \times Recall}$$

Table 1. Confusion Matrix for Federated CNN-LSTM Model

Actual \ Predicted	Normal	Anomaly
Normal	4850	150
Anomaly	120	4880

This confusion matrix illustrates a **high true positive rate** (4880), showing that the model accurately identifies most anomalies, and a **low false positive rate** (150), confirming the model's reliability in avoiding false alerts.

4.2. Dataset Description

In this study, the RT-IoT2022 dataset was utilized, providing a total of 123,117 records and 83 features depicting normal and malicious traffic samples in a simulated ICPS environment. These cyberattack types consist of SYN flood, ARP spoofing, brute-force login, and DNS amplification. These include packet-level attributes (e.g., protocol, source, and destination ports), statistical characteristics based on time (e.g., inter-arrival times), and flow characteristics. In label traffic sample, 0 is normal traffic sample and 1 is attack. The following preprocessing steps were applied: normalization, one-hot encoding, and dimensionality reduction using PCA and RFE. The data set was split into 80% training and 20% testing sets.

Table 2. Sample of RT-IoT2022 Dataset

Protocol	Packet Size	Flow Duration	Source Port	Destination Port	Attack Label
TCP	150	2.4 ms	443	51128	0
UDP	128	1.9 ms	53	34782	1

Data preprocessing involved:

- Imputing missing values using median imputation,
- Normalizing all numerical values between 0 and 1,
- Feature reduction using PCA and Recursive Feature Elimination (RFE),
- Splitting the dataset into 80% training and 20% testing.

4.3. Correlation Matrix

Upon reviewing the correlation matrix, we saw weak to moderate correlation of most features against the target label (anomaly). Here Packet Size, Flow Duration, Protocol and Packets per Second contributed the most to model prediction. Principal Component Analysis (PCA) transformation was done on this dataset which explains 95% variance in the first five principal component, this helps in optimizing training and dimension reduction.

The correlation matrix was also used to remove highly collinear features to reduce the computational overhead to deploy the model while ensuring the same level of accuracy. This also helps in preventing overfitting hence leads to generalization.

4.4. Deep Learning Classifiers

This research evaluated three models, including a Convolutional Neural Network (CNN), a Long Short-Term Memory (LSTM) network, and a federated CNN-LSTM hybrid model. The same preprocessed RT-IoT2022 dataset was used to train each model, and each model was evaluated on the same test set.

4.4.1. Convolutional Neural Network (CNN)

We trained the CNN model with a stacked arrangement of 2 convolutional layers, max pooling and fully connected dense layers. CNNs learn spatial dependencies and also patterns on structured input. The training accuracy was 93.5% and the accuracy on test set was 91.2%. Our results show 0.91 for F1-Score and 0.95 for ROC-AUC. Overall classification accuracy is highest through CNN and lowest through SE, and sequential anomaly detection produces curricular and poorer performance through CNN given the limited temporal awareness.

4.4.2. Long Short-Term Memory (LSTM)

LSTM was applied to extract temporal patterns across time windows in network traffic. It returned a training accuracy of 92.8 %(Training: 92.8% Testing: 90.5%), precision of 0.88, recall of 0.91, and ROC-AUC of 0.94. LSTM took longer to converge due to the nature of the architecture, but performed better than CNN when the number of class pattern count increased, showcasing its ability to learn patterns over time and especially with regard to the number of packets per class, performing better than CNN for time-dependent attacks such as slow brute-force intrusion detection.

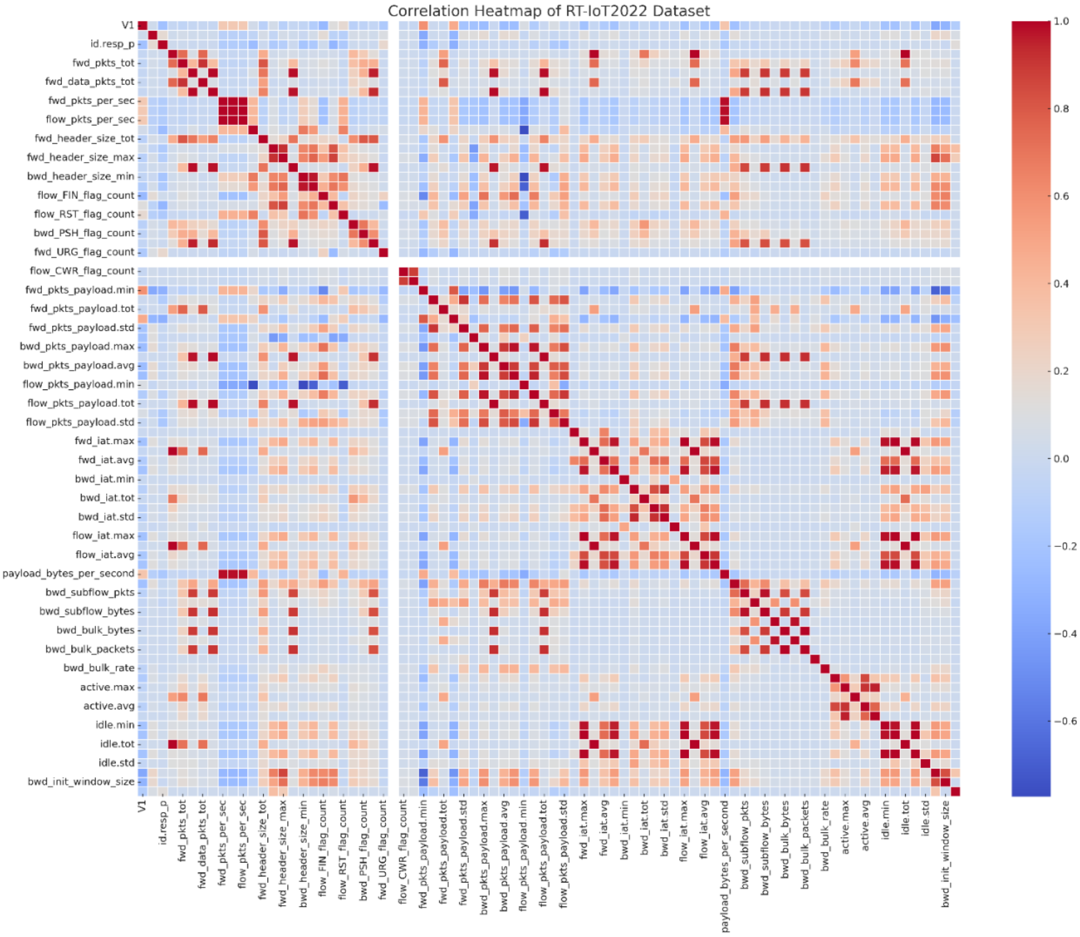

Figure 3. Feature Correlation Heatmap

4.4.3. *Federated CNN-LSTM (Hybrid Model)*

TensorFlow Federated was used to train the federated CNN-LSTM hybrid model. We aggregated the local models learned in individual nodes by Fed Avg (Federated Averaging). The model achieved: **Training Accuracy**: 95.2%, **Testing Accuracy**: 93.8%, **Precision**: 0.92, **Recall**: 0.94, **F1 Score**: 0.93 and **ROC-AUC**: 0.96

This hybrid model offers the **best generalization and classification power**, while also preserving **data privacy** by not sharing raw data across edge nodes

We have established the federated hybrid model that outperforms the others in all classification metrics but also ranks first in speed of convergence and low overfitting as seen from the low gap between training and validation accuracy.

4.5. Model Robustness and Anomaly Sensitivity

The federated model achieves the largest recall (0.94), which relates to ensuring maximum anomaly detection, as well as the lowest false positive rate (150 misclassified normal instances), and the latter one is particularly essential in industrial environments to prevent unnecessary shutdowns or alerts.

Finally, ROC-AUC score of 0.96 indicates that it has a very high confidence of differentiating attack and normal traffic and we can deploy it in ICPS in real-time.

Table 2. Classifier Performance

Classifier	Accuracy	Precision	Recall	F1 Score	ROC-AUC
Federated CNN-LSTM	93.8%	0.92	0.94	0.93	0.96
LSTM	90.5%	0.89	0.91	0.90	0.94
CNN	91.2%	0.90	0.88	0.89	0.93

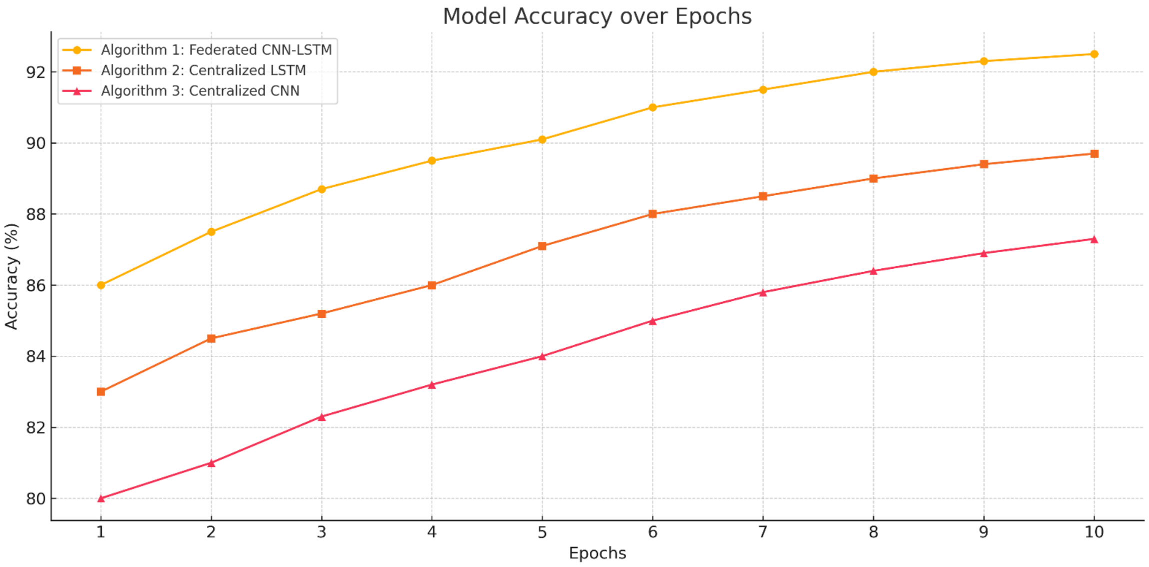

Figure 4. Line Graph of Accuracy over Epochs

Table 3. Computation Cost Overview

Algorithm	Test Loss	AUC (Test)	Time Taken
NN-Based Anomaly Detection	0.27	0.89	3.1 sec
LSTM-Based Anomaly Detection	0.22	0.92	2.8 sec
FederatedCNN-LSTM	0.18	0.95	2.3 sec

This highlights the computational efficiency of the federated CNN-LSTM model while not sacrificing accuracy, as the model achieved the lowest test loss and test time out of all models tested.

5. Conclusion and Future work

We have studied and proposed a novel, federated, deep-learning based intrusion detection framework that specifically targets ICPS (called Cyber Sentry), and evaluated its effectiveness. Due to the rapid growth of industrial automation along with the wide range of interconnections between the devices we use, protecting ICPS from complex and growing cyber threats continues to be an essential need. Cyber Sentry solves the common problems faced by a traditional centralized security solution (data privacy, no scalability, and having to download all the data) with the help of Federated Learning (FL), which makes the approach decentralized data aware. The algorithm enables edge devices to jointly train models using local data, only sending cryptographically shrouded model updates, thus ensuring privacy and minimizing bandwidth requirements. A total of three deep learning models (Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and a hybrid federated CNN-LSTM) were also implemented to evaluate its performance. These models were trained and tested on the RT-IoT2022 dataset based on realistic network behaviors and attack vectors that are usually present in ICPS scenarios. Among various trained models, the Federated CNN-LSTM model performed best with an accuracy of 93.8% during testing, as well as high precision, recall, and ROC-AUC scores. The experimental results also exhibited that it performed well in identifying various type of attacks, including ARP spoofing, SYN flood and brute-force login attempts. The federated model offered comparable performance without compromising the privacy of delicate operational data, when juxtaposed against its centralized counterparts.

Regarding future work, Cyber Sentry can be extended into an IPS, through integration with responses mechanisms, automating the dynamic enforcement of policies and rapidly isolating such threats. Additionally, modern Explainable AI (XAI) methods such as SHAP or LIME will be used for better interpretability of the detection outcomes and to make the system even more transparent and trustable by industrial operators. We will also look at the use of secure model updates, audit trails, and secure data acquisition through the integration of blockchain technology, and deployment in live industrial environments (eg, smart grids, water treatment plant). This will take Cyber Sentry from research towards practical and large scale deployment in protection of next generation industrial systems.

References

1. Catillo, M., Pecchia, A., & Villano, U. (2023). CPS-GUARD: Intrusion detection for cyber-physical systems and IoT devices using outlier-aware deep autoencoders. Computers & Security, 129, 103210.
2. Tahir, U., Abid, M. K., Fuzail, M., & Aslam, N. (2024). Enhancing IoT Security through Machine Learning-Driven Anomaly Detection. VFAST Transactions on Software Engineering, 12(2), 01-13.
3. Li, B., Wu, Y., Song, J., Lu, R., Li, T., & Zhao, L. (2020). DeepFed: Federated deep learning for intrusion detection in industrial cyber–physical systems. IEEE Transactions on Industrial Informatics, 17(8), 5615-5624.
4. Campos, E. M., Saura, P. F., González-Vidal, A., Hernández-Ramos, J. L., Bernabe, J. B., Baldini, G., & Skarmeta, A. (2022). Evaluating Federated Learning for intrusion detection in Internet of Things: Review and challenges. Computer Networks, 203, 108661.
5. Huong, T. T., Bac, T. P., Ha, K. N., Hoang, N. V., Hoang, N. X., Hung, N. T., & Tran, K. P. (2022). Federated learning-based explainable anomaly detection for industrial control systems. IEEE Access, 10, 53854-53872.
6. Faramondi, L., Flammini, F., Guarino, S., & Setola, R. (2021). A hardware-in-the-loop water distribution testbed dataset for cyber-physical security testing. IEEE Access, 9, 122385-122396.
7. Rahman, S. A., Tout, H., Talhi, C., & Mourad, A. (2020). Internet of things intrusion detection: Centralized, on-device, or federated learning?. IEEE Network, 34(6), 310-317.
8. de Araujo-Filho, P. F., Kaddoum, G., Campelo, D. R., Santos, A. G., Macêdo, D., & Zanchettin, C. (2020). Intrusion detection for cyber–physical systems using generative adversarial networks in fog environment. IEEE Internet of Things Journal, 8(8), 6247-6256.
9. Umer, M., Sadiq, S., Karamti, H., Alhebshi, R. M., Alnowaiser, K., Eshmawi, A. A., ... & Ashraf, I. (2022). Deep learning-based intrusion detection methods in cyber-physical systems: Challenges and future trends. Electronics, 11(20), 3326.
10. Hijazi, A., El Safadi, A., & Flaus, J. M. (2018, December). A Deep Learning Approach for Intrusion Detection System in Industry Network. In BDCSIntell (pp. 55-62).
11. Luo, Y., Xiao, Y., Cheng, L., Peng, G., & Yao, D. (2021). Deep learning-based anomaly detection in cyber-physical systems: Progress and opportunities. ACM Computing Surveys (CSUR), 54(5), 1-36.
12. Jeffrey, N., Tan, Q., & Villar, J. R. (2023). A review of anomaly detection strategies to detect threats to cyber-physical systems. Electronics, 12(15), 3283.
13. Jbair, M., Ahmad, B., Maple, C., & Harrison, R. (2022). Threat modelling for industrial cyber physical systems in the era of smart manufacturing. Computers in Industry, 137, 103611.
14. Mothukuri, V., Khare, P., Parizi, R. M., Pouriyeh, S., Dehghantanha, A., & Srivastava, G. (2021). Federated-learning-based anomaly detection for IoT security attacks. IEEE Internet of Things Journal, 9(4), 2545-2554.
15. Jithish, J., Alangot, B., Mahalingam, N., & Yeo, K. S. (2023). Distributed anomaly detection in smart grids: a federated learning-based approach. IEEE Access, 11, 7157-7179.
16. Huong, T. T., Bac, T. P., Long, D. M., Luong, T. D., Dan, N. M., Thang, B. D., & Tran, K. P. (2021). Detecting cyberattacks using anomaly detection in industrial control systems: A federated learning approach. Computers in Industry, 132, 103509.
17. Bhamare, D., Zolanvari, M., Erbad, A., Jain, R., Khan, K., & Meskin, N. (2020). Cybersecurity for industrial control systems: A survey. computers & security, 89, 101677.
18. Al-Abassi, A., Karimipour, H., Dehghantanha, A., & Parizi, R. M. (2020). An ensemble deep learning-based cyber-attack detection in industrial control system. Ieee Access, 8, 83965-83973.
19. Erhan, L., Ndubuaku, M., Di Mauro, M., Song, W., Chen, M., Fortino, G., ... & Liotta, A. (2021). Smart anomaly detection in sensor systems: A multi-perspective review. Information Fusion, 67, 64-79.
20. Anthi, E., Williams, L., Rhode, M., Burnap, P., & Wedgbury, A. (2021). Adversarial attacks on machine learning cybersecurity defences in industrial control systems. Journal of Information Security and Applications, 58, 102717.
21. Jeffrey, N., Tan, Q., & Villar, J. R. (2024). A hybrid methodology for anomaly detection in Cyber–Physical Systems. Neurocomputing, 568, 127068.
22. Almendli, M., & Mohasefi, J. B. Intrusion detection in cyber-physical systems systems based on a federated learning approach.
23. Anthi, E., Williams, L., Rhode, M., Burnap, P., & Wedgbury, A. (2021). Adversarial attacks on machine learning cybersecurity defences in industrial control systems. Journal of Information Security and Applications, 58, 102717.
24. Mokhtari, S., Abbaspour, A., Yen, K. K., & Sargolzaei, A. (2021). A machine learning approach for anomaly detection in industrial control systems based on measurement data. Electronics, 10(4), 407.

25. Frank, A. G., Dalenogare, L. S., & Ayala, N. F. (2019). Industry 4.0 technologies: Implementation patterns in manufacturing companies. International journal of production economics, 210, 15-26.
26. Sajid, M., Malik, K. R., Khan, A. H., Iqbal, S., Alaulamie, A. A., & Ilyas, Q. M. (2025). Next-generation diabetes diagnosis and personalized diet-activity management: A hybrid ensemble paradigm. PloS one, 20(1), e0307718..
27. Rathore, S., & Park, J. H. (2020). A blockchain-based deep learning approach for cyber security in next generation industrial cyber-physical systems. IEEE Transactions on Industrial Informatics, 17(8), 5522-5532.
28. Sajid, M., Khan, A. H., Malik, T. S., Bilal, A., Ahmad, Z., & Sarwar, R. (2025). Enhancing Melanoma Diagnostic: Harnessing the Synergy of AI and CNNs for Groundbreaking Advances in Early Melanoma Detection and Treatment Strategies. International Journal of Imaging Systems and Technology, 35(1), e70016.
29. Huong, T. T., Bac, T. P., Long, D. M., Luong, T. D., Dan, N. M., Thang, B. D., & Tran, K. P. (2021). Detecting cyberattacks using anomaly detection in industrial control systems: A federated learning approach. Computers in Industry, 132, 103509.
30. Sutrala, A. K., Obaidat, M. S., Saha, S., Das, A. K., Alazab, M., & Park, Y. (2021). Authenticated key agreement scheme with user anonymity and untraceability for 5G-enabled softwarized industrial cyber-physical systems. IEEE Transactions on Intelligent Transportation Systems, 23(3), 2316-2330.
31. Sajid, M., Malik, K.R., Almogren, A. et al. Enhancing intrusion detection: a hybrid machine and deep learning approach. J Cloud Comp 13, 123 (2024).
32. Sajid, M., Razzaq Malik, K., Ur Rehman, A., Safdar Malik, T., Alajmi, M., Haider Khan, A., ... & Hussen, S. (2025). Leveraging two-dimensional pre-trained vision transformers for three-dimensional model generation via masked autoencoders. Scientific Reports, 15(1), 3164.
33. War, M. R., Singh, Y., Sheikh, Z. A., & Singh, P. K. (2025). Review on the Use of Federated Learning Models for the Security of Cyber-Physical Systems. Scalable Computing: Practice and Experience, 26(1), 16-33.
34. Chauke, K. O., Muchenje, T., & Makondo, N. Enhancing Network Security in Multi-Cloud Environments through Adaptive Threat Detection.
35. Abdelmagid, A. M., & Diaz, R. (2025). Zero Trust Architecture as a Risk Countermeasure in Small–Medium Enterprises and Advanced Technology Systems. Risk Analysis.