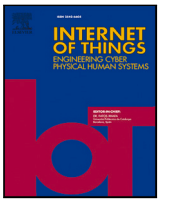
Internet of Things 27 (2024) 101252

Contents lists available at ScienceDirect

Internet of Things

journal homepage: www.elsevier.com/locate/iot



Enhancing cybersecurity in Edge IIoT networks: An asynchronous federated learning approach with a deep hybrid detection model

Syed Muhammad Salman Bukhari [a], Muhammad Hamza Zafar [b], Mohamad Abou Houran [c], Zakria Qadir [d], Syed Kumayl Raza Moosavi [b], Filippo Sanfilippo [b,e,*]

[a] Department of Electrical Engineering, Capital University of Science and Technology, Islamabad, 44000, Pakistan
[b] Department of Engineering Sciences, University of Agder, Grimstad, 4879, Norway
[c] School of Electrical Engineering Xi'an Jiaotong University, Xi'an, 710049, China
[d] School of Engineering, Design and Built Environment, Western Sydney University, Sydney, 2750, Australia
[e] Department of Software Engineering, Kaunas University of Technology, Kaunas, 44029, Lithuania

ARTICLE INFO

Keywords:
Cybersecurity
Federated learning
Industrial Internet of Things (IIoT)
Network intrusion detection
Data privacy
Convolutional Neural Network (CNN)
Gated Recurrent Unit (GRU)
Long Short-Term Memory (LSTM) networks
Asynchronous learning

ABSTRACT

In the rapidly evolving field of the Industrial Internet of Things (IIoT), advancements in wireless technology have resulted in significant cybersecurity vulnerabilities. These weaknesses pose serious risks such as damage to manufacturing systems, theft of intellectual property, and substantial financial losses. This study introduces an advanced deep hybrid learning model in an asynchronous federated learning setup, aimed at improving the detection of cyberattacks and ensuring robust data privacy. The combination of Convolutional Neural Networks (CNN), Gated Recurrent Units (GRU), and Long Short-Term Memory (LSTM) networks provides an effective solution for quickly identifying anomalies in IIoT sensor traffic. Our model operates asynchronously, ensuring data remains localised to improve security while avoiding the need for complete node synchronisation. Demonstrating outstanding effectiveness, the model achieves an accuracy of 1.00%, precision of 1.00%, recall of 1.00%, and an F1 score of 1.00% across a variety of IIoT environments. These results highlight the model's exceptional adaptability and its capability to rapidly respond to emergent threats, marking a significant step forward in the protection of IIoT infrastructures and the rigorous maintenance of data privacy.

1. Introduction

The Internet of Things (IoT) represents a significant paradigm shift in digital connectivity and intelligent system integration. With its network of interconnected devices spanning numerous industries, IoT has experienced unprecedented growth due to its versatility, scalability, and the integration of advanced technological features. At the forefront of this technological revolution is the Industrial Internet of Things (IIoT), which is pivotal in transforming industrial production processes and optimising resource utilisation. This transformation is integral to advancing Industry 5.0, which advocates for a seamless amalgamation of digital and physical industrial systems [1,2]. Despite these advancements, the rapid expansion of the IoT ecosystem presents formidable challenges. The IoT market, having expanded to over 8 billion devices and expected to reach approximately 41 billion by 2027, is projected to grow from a valuation of over $380 billion in 2021 to nearly $1.8 trillion by 2028 [3,4]. This growth is particularly notable in sectors such as automotive, smart homes, and healthcare, where it has been accompanied by an increase in security vulnerabilities. The frequency

* Corresponding author at: Department of Engineering Sciences, University of Agder, Grimstad, 4879, Norway.
E-mail address: filippo.sanfilippo@uia.no (F. Sanfilippo).

https://doi.org/10.1016/j.iot.2024.101252
Received 1 January 2024; Received in revised form 20 May 2024; Accepted 11 June 2024
Available online 18 June 2024

and sophistication of cyberattacks have escalated, with incidents nearly doubling within a year (2020–2021), exemplified by high-profile cases like the Colonial Pipeline attack. These incidents highlight the urgent need for robust security measures [3,5]. The traditional security solutions often prove inadequate for the complex demands of IIoT systems, which require not only reliability and scalability but also high energy efficiency [5]. Addressing these challenges necessitates the development of advanced security solutions tailored to the unique requirements of IIoT environments.

This study introduces a novel hybrid learning model encapsulated within an asynchronous Federated Learning (FL) framework, designed to address the escalating security challenges in the IoT and IIoT sectors. Our model leverages the synergy of Convolutional Neural Networks (CNN), Gated Recurrent Units (GRU), and Long Short-Term Memory (LSTM) networks to meet the intricate security needs of these environments. Focusing on the detection and mitigation of cyber threats, the model prioritises critical aspects of industrial cybersecurity, such as data privacy and operational efficiency. This approach marks a significant advancement in industrial network security, providing robust defence mechanisms against the evolving landscape of cyber threats that endanger critical infrastructures. To bolster the model's practical applicability and improve its adaptability, this research integrates the Edge-IIoTset dataset — a comprehensive and publicly available resource — enabling the deployment of our solution across various real-world scenarios in different IIoT sectors. This integration highlights the potential of our model as a flexible and potent solution to the pressing cybersecurity challenges within the dynamic realms of IoT and IIoT.

1.1. Motivation and contributions

The motivation behind this study stems from two primary considerations. Firstly, the rapid proliferation of IIoT devices in various industrial sectors has escalated the risks of cyberattacks, potentially leading to significant operational disruptions. Traditional security solutions often fall short of addressing the dynamic and heterogeneous nature of IIoT environments effectively. Secondly, the recent advances in machine learning, particularly in deep learning, open new opportunities to improve cybersecurity measures but require careful adaptation to the decentralised and data-sensitive contexts of IIoT. This study makes several significant contributions to the field of IIoT cybersecurity:

1. **Hybrid Deep Learning Model:** We propose a hybrid deep learning model that integrates Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRU) architectures. This model leverages the strengths of each approach to improve detection accuracy and response times to cyber threats within IIoT environments.
2. **Asynchronous and Collaborative Federated Learning Framework:** The implementation of an asynchronous federated learning framework boosts efficient model training without compromising data privacy. This framework supports incremental and collaborative learning from decentralised data sources, crucial for dynamic and scalable IIoT environments.
3. **Validation with Real-World Data:** We validate our model using a comprehensive dataset from actual IIoT operations, ensuring that the model is robust and effective across various scenarios and attack vectors.

These contributions aim to provide a scalable, efficient, and privacy-preserving solution to cybersecurity challenges in the IIoT sector, leveraging both asynchronous and collaborative learning modalities in federated learning.

1.2. Paper organisation

This article is organised into six main sections to provide a thorough analysis of IIoT cybersecurity. Section 1 introduces the transformative impact of the Internet of Things (IoT) and the Industrial Internet of Things (IIoT) on industrial innovation, discussing the significant security challenges that arise with these advancements. Section 2 reviews existing literature, examining prior research in IoT and IIoT cybersecurity, identifying methodologies used, and highlighting research gaps our study addresses. In Section 3, we detail our proposed approach, integrating CNN, LSTM, and GRU within an asynchronous FL framework to improve cybersecurity defences. Section 4 elaborates on the dataset used for validating our model, explaining its creation, relevance, and categorisation of included cyberattacks. Section 5, Results and Discussion, evaluates the performance of our model against several metrics, demonstrating its effectiveness and superiority over conventional models. Finally, Section 6 concludes the paper by summarising our key contributions to IIoT cybersecurity and outlining future research directions to continue advancing this crucial field.

2. Related work

The field of IoT and IIoT cybersecurity has seen the emergence of many datasets crucial for the progress of Machine Learning (ML) based Intrusion Detection Systems (IDS). This segment explores various significant datasets that have played a key role in the development of IoT/IIoT-based IDS, evaluating their performance with different ML models. Vaccari et al.'s MQTTset [6] is particularly noteworthy, centring on the Message Queuing Telemetry Transport (MQTT) protocol's vulnerabilities in IoT devices. This dataset underwent validation employing a range of techniques including Neural Network, Decision Tree (DT), Random Forest (RF), Naive Bayes (NB), Multilayer Perceptron (MP), and Gradient Boost (GB). Its primary constraint, however, is the exclusive focus on MQTT traffic, which omits IIoT-specific data like the Modbus protocol, thereby constraining its utility in IIoT security scenarios. Similarly, Meidan et al.'s dataset [7], designed for network-based anomaly detection using deep autoencoders, excels in identifying irregular traffic from compromised IoT devices. It was rigorously tested with Local Outlier Factor (LOF), One-Class SVM, Isolation Forest (IF), and optimised deep autoencoders, demonstrating substantial efficacy for most devices. Nonetheless, its limitation lies in the lack of IIoT traffic representation and a narrow focus on botnet attacks, reducing its applicability for a wider range of IoT security

needs. Koroniotis et al. put forth dataset [8], aimed at botnet detection within IoT networks, integrating both real and simulated IoT traffic. The dataset's evaluation using Recurrent Neural Network (RNN), Support Vector Machine (SVM), and Long-Short Term Memory (LSTM) revealed that SVM exhibited the highest accuracy. However, the absence of IIoT-specific data in this dataset limits its relevance for IIoT security applications. Moustafa et al.'s dataset [2] includes a wide range of data from various IoT services and network traffic. Despite its broad scope, the absence of specific IIoT traffic and limited intrusion analysis restricts its usefulness for IIoT security evaluations. Al-Hawawreh et al.'s dataset [9] was evaluated using advanced techniques such as Deep Neural Networks (DNN) and Gated Recurrent Units (GRU). Although it offers detailed insights, its focus on centralised learning highlights the need for federated learning to improve privacy and efficiency in IoT and IIoT systems.

Zolanvari et al.'s dataset [1], specifically designed for IIoT cybersecurity, was tested using various models including Logistic Regression (LR) and Support Vector Machines (SVM). While it showed high accuracy, especially with the Random Forest (RF) model, its exclusive focus on IIoT limits its applicability to broader IoT scenarios. Recent studies have also contributed significantly. Mohy et al. [10] examined realistic traffic patterns in IoT environments, focusing on botnet attack patterns, and achieved high scores for accuracy, MCC, and AUC. However, their study lacks scenarios specific to IIoT. Yazdinejad et al. [11] developed a parallel ensemble model for detecting anomalies in IIoT systems, achieving precise results though it falls short in representing a diverse range of real-world IIoT deployments. Wang et al. [12] introduced a semi-supervised model that uses limited labelled data and integrates edge intelligence to reduce latency and protect privacy. However, their reliance on pre-trained models may not effectively address all new types of malicious behaviour. In 2024, Maddali et al. [13] implemented a comprehensive malware detection framework using DCGAN and ConvNeXt, achieving excellent accuracy on both the MaleVis and Malimg datasets, though this method demands substantial computational resources which may not always be available. Rajak [14] utilised a 5G-based DL-SkLSTM system to classify various cyberattacks on the Edge-IIoTset, showing great potential but raising concerns about its suitability for environments with limited resources. Hassini [15] proposed a straightforward CNN1D model that excelled in detecting sophisticated threats in I-IoT, achieving nearly perfect accuracy and demonstrating its effectiveness with k-fold cross-validation. Nevertheless, its training on a specific dataset could limit its generalisability to other I-IoT scenarios not covered in the training data. Emerging technologies within the Industrial Internet of Things (IIoT) sector continue to address critical concerns over operational efficiency and security. In recent work, [16] have developed a Quality of Service and Privacy-Aware Routing protocol (QoSPR) for 5G-IIoT, which optimises routing through federated reinforcement learning to improve latency and load balancing while ensuring data privacy. Complementarily, [17] propose a federated learning-based anomaly detection strategy for IIoT, which reduces privacy risks by localising deep reinforcement learning algorithms, thereby achieving high throughput and low latency in privacy preservation across diverse IIoT scenarios. Additionally, [18] introduces a blockchain-based secure data aggregation strategy (BSDA) that incorporates security labels into block headers to manage task assignment efficiently among mobile data collectors, significantly enhancing data security and system performance in IoT environments. These studies collectively advance the security frameworks, demonstrating significant improvements in both the privacy and operational efficiency of IIoT and IoT systems. Table 1 below summarises these datasets, their evaluation methodologies, and the limitations impacting their relevance in the broader context of IoT/IIoT cybersecurity.

Our research introduces a novel method to address the security challenges of IoT and IIoT by uniquely combining the selected Edge-IIoTset dataset with innovative deep learning models. This approach distinctively integrates Convolutional Neural Networks (CNN) [19], Gated Recurrent Units (GRU) [20], and Long Short-Term Memory (LSTM) [21] within a federated learning framework [22], creating a deep hybrid learning system. This system harnesses the complementary strengths of these technologies to achieve exceptional accuracy in intrusion detection across IoT and IIoT landscapes. Our architecture employs 1D-CNN as the backbone, leveraging its convolutional and pooling layers to efficiently extract features from sequential data while integrating activation functions like ReLU to model complex data patterns. The GRU component manages temporal dependencies, essential for recognising sophisticated cyberattack patterns, and the LSTM networks are utilised for their long-term memory capabilities, crucial in identifying ongoing threats. Additionally, we implement dropout layers and normalisation techniques to prevent overfitting, ensuring robust performance against unseen attack vectors. Our asynchronous federated learning model distributes the dataset across multiple clients, allowing for independent model learning and enhancement of privacy and computational efficiency without necessitating simultaneous updates. This integration of convolutional and recurrent architectures not only improves feature extraction and sequence modelling but also provides the system with the capability to accurately detect and classify a diverse array of attack types. The overall system architecture sets a new standard in efficiency, versatility, and privacy protection in industrial network cybersecurity, aiming to revolutionise IoT/IIoT security applications by bridging technological advancements with practical, real-world challenges. The integration of these models within an asynchronous FL framework represents a significant advancement in the field of IIoT cybersecurity. This proposed approach promises to deliver high accuracy and efficiency, ensuring robust cyber defence mechanisms that are scalable and adaptable to the ever-evolving threat landscape. Fig. 1 presents the basic abstract of the proposed model flow.

3. Proposed model

This study introduces a hybrid learning model that combines deep learning techniques to enhance cybersecurity in IoT and IIoT environments. The model employs 1D Convolutional Neural Networks (1D-CNNs), which are designed to effectively recognise complex data patterns through convolutional and pooling layers using activation functions like Rectified Linear Units (ReLU). To prevent overfitting and ensure the model's efficacy on new datasets, techniques such as dropout layers and normalisation are integrated. The model's design also includes Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) networks, which are critical for managing data sequences and recognising complex cyberattack patterns. These components are implemented within an asynchronous federated learning framework, enhancing operational efficiency and adaptability while prioritising data privacy. This setup allows for rapid model updates and training, essential for timely response to cybersecurity threats [23–25].

Table 1
Comparison of previous studies.

Reference	Year	Description	Lacks
[7]	2018	Comprises traffic from 9 IoT devices and two botnets (BASHLITE and Mirai) for 10 attack types.	Limited to a narrow IoT threat model; does not include IIoT traffic.
[6]	2020	Concentrates on MQTT protocol traffic and various attack streams associated with IoT devices.	Primarily contains MQTT traffic, lacking broader applicability for IIoT security.
[9]	2021	Integrates device-agnostic data, suitable for ML/DL-based IDS in both IoT and IIoT systems.	Focuses solely on centralised learning methods, not encompassing federated learning approaches.
[1]	2021	Combines data from IIoT and industrial devices to simulate an industrial environment for testing a machine learning-based Intrusion Detection System.	Excludes IoT-related traffic and data, limiting its suitability for broader IoT security evaluations.
[10]	2023	Focuses on IoT traffic and botnet attack patterns for IDS evaluations. High scores show effectiveness.	Does not provide explicit IIoT-specific scenarios.
[11]	2023	Focuses on threat hunting in IIoT systems with a parallel ensemble model. Classifies anomalies efficiently with high accuracy using Multi-class AdaBoost and majority voting.	Lacks a broad variety of real-world IIoT deployment scenarios.
[12]	2023	Applies semi-supervised model to classify IIoT malicious traffic, enhancing performance with limited data. Uses edge intelligence to reduce latency and enhance privacy.	Relies on pre-trained models which may not capture all novel malicious behaviours effectively.
[13]	2024	Deploys a multi-stage malware framework with DCGAN and ConvNeXt, achieving high accuracy and generalisability on MaleVis and Malimg datasets.	Requires extensive computational resources which may not be available on all edge devices.
[14]	2024	Employs a 5G-enhanced DL-SkLSTM system for accurate cyber-attack classification on the Edge-IIoTset, surpassing conventional DL techniques.	While effective, the complexity and resource requirements of 5G-based systems may limit deployment in resource-constrained environments.
[15]	2024	Developed a CNN1D model with 99.96% accuracy for detecting 14 I-IoT threats, validated by k-fold cross-validation for robust real-world application.	While highly effective, the model's training on a specific dataset might limit its adaptability to other I-IoT environments not represented in the training data.

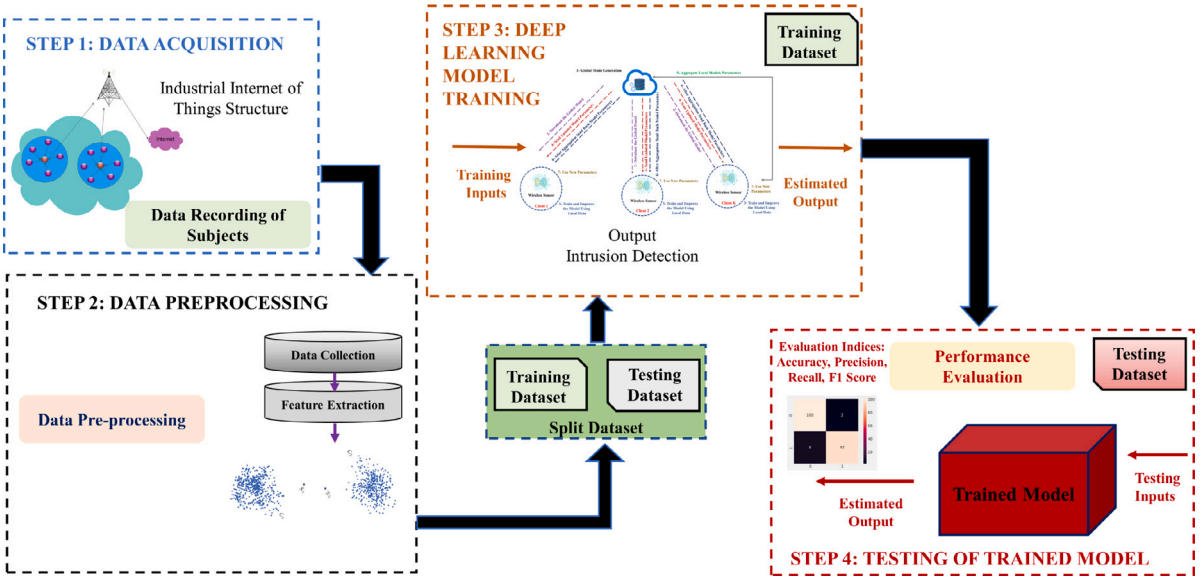


Fig. 1. Proposed asynchronous FL-based intrusion detection model for Edge IIoT.

3.1. 1D convolutional neural networks (1D-CNNs)

1D-CNNs are crucial components of our model, specifically tailored for processing sequential data efficiently. These networks leverage convolutional and pooling layers to extract and learn essential features from data. A key advantage of using 1D-CNNs is

their ability to employ the Rectified Linear Unit (ReLU) activation function [19], which introduces nonlinearity into the model. This nonlinearity is vital for distinguishing complex patterns within the data, which is central to effective cybersecurity measures in IoT and IIoT systems. The architecture of 1D-CNNs is particularly beneficial in real-time cybersecurity applications. By operating on 1D arrays, these networks require significantly less computational power compared to their 2D counterparts, making them ideal for use in environments with limited resources. This efficiency does not compromise their performance, as the streamlined data processing allows for quick and effective pattern recognition and anomaly detection, addressing the critical need for rapid response in security-sensitive contexts. 1D-CNNs are pivotal for extracting and interpreting complex patterns essential for cybersecurity measures in IoT and IIoT environments. The core operation within these networks is the convolution operation, which can be described by the following equation:

$$O_l = f(W_l * X + b_l) \tag{1}$$

Here, O_l is the output from the lth convolutional layer, W_l represents the kernel or filter weights applied to the input X, and b_l is the bias. The activation function f, typically a Rectified Linear Unit (ReLU), introduces non-linearity, enhancing the network's ability to learn complex and nonlinear relationships in the data. This convolution process is instrumental in feature extraction, crucial for identifying nuanced cybersecurity threats. By applying W_l across the input data, the network effectively captures essential features that are pivotal for detecting anomalies and potential security breaches. Moreover, the network updates its weights to optimise performance, based on the gradient of the loss function, as represented by:

$$W_{new} = W_{old} - \eta \cdot \nabla W \tag{2}$$

In this equation, W_{new} and W_{old} denote the updated and previous weights, respectively, while η is the learning rate, and ∇W is the weight gradient. This weight adjustment is fundamental in refining the model's accuracy and responsiveness over time, particularly in the dynamic contexts of IoT and IIoT.

The efficiency of 1D-CNNs in processing 1D data makes them well-suited for real-time applications, where fast data analysis is crucial. This efficiency, combined with their ability to deeply understand temporal and sequential data, positions 1D-CNNs as a powerful tool in our cybersecurity arsenal, particularly useful in environments with large data volumes and demanding performance needs.

3.2. Gated Recurrent Units (GRUs)

Gated Recurrent Units (GRUs) represent a significant advancement in the evolution of Recurrent Neural Networks (RNNs), engineered specifically for efficient processing of sequential data [20]. The Gated Recurrent Unit (GRU) is an advanced architecture for processing sequential data. As illustrated in Fig. 2, the GRU comprises two gates: the update gate and the reset gate. These gates determine how much of the past information needs to be passed along to the future. This mechanism allows GRUs to capture dependencies from large intervals of time effectively, making them suitable for tasks such as time series prediction or language modelling. GRUs adeptly address the vanishing gradient problem commonly encountered in standard RNNs, utilising unique gating mechanisms. These mechanisms are instrumental in controlling the information flow within the unit, ensuring the maintenance of data relevance over various time steps.

The architecture of GRUs encompasses two fundamental components, known as gates:

1. **Reset Gate**: This gate determines the extent to which past information should be forgotten, thus enabling the GRU to discard irrelevant data and concentrate on recent inputs.
2. **Update Gate**: It balances between retaining important information from the previous state and incorporating new inputs.

Gated Recurrent Units (GRUs) optimise the processing of sequential data through specialised mechanisms detailed by mathematical equations, significantly enhancing their efficiency in learning and maintaining long-term data dependencies. The Reset Gate, crucial for deciding how much past information to discard, operates through the equation:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \tag{3}$$

Here, r_t is the gate's output, W_r is the weight matrix, h_{t-1} the previous hidden state, x_t the current input, b_r the bias, and σ the sigmoid activation function.

Similarly, the Update Gate, which determines the amount of information to carry forward, is defined as:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \tag{4}$$

These gates enable the GRU to effectively balance between maintaining relevant historical information and incorporating new data inputs. This capability is vital for applications that require analysis over long periods, such as language modelling or time-series forecasting. The candidate's hidden state, influenced by the Reset Gate, further refines the model's accuracy:

$$\widetilde{h}_t = tanh(W \cdot [r_t * h_{t-1}, x_t] + b) \tag{5}$$

Finally, the GRU updates its hidden state h_t at each timestep by blending the previous state and the candidate state, facilitated by both gates:

$$h_t = z_t * h_{t-1} + (1 - z_t) * \widetilde{h}_t \tag{6}$$

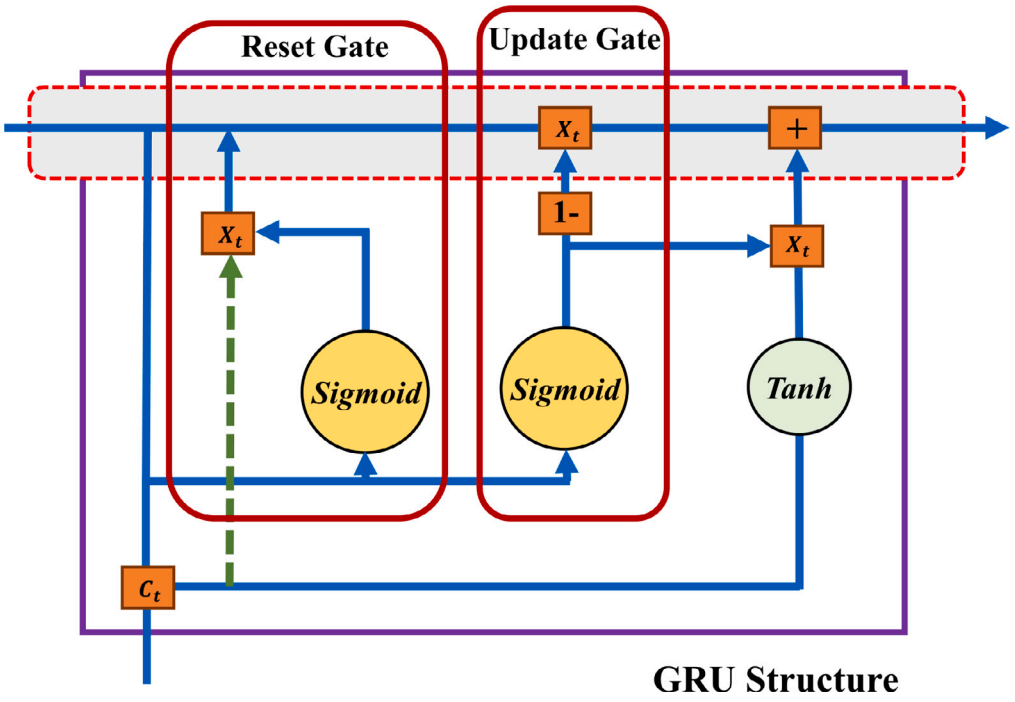


Fig. 2. Gated Recurrent Unit (GRU) architecture with update and reset gates.

GRUs demonstrate remarkable efficacy in various sequence modelling tasks by capturing dependencies across varying time scales efficiently. Their design is particularly advantageous in scenarios where performance and computational efficiency are critical, making them an essential tool for dynamic and resource-constrained environments.

3.3. Long Short-Term Memory (LSTM) networks

LSTM networks are a specialised type of RNNs designed to handle long sequences of data. They are adept at capturing both long-term and short-term dependencies in data, a critical feature for many sequential modelling tasks. LSTMs include memory cells that carry information across long sequences, which helps to overcome the vanishing gradient problem common in traditional RNNs [21].

LSTM unit, depicted in Fig. 3, is a type of RNN architecture that is designed to remember information for long periods. An LSTM unit includes three gates: the forget gate, input gate, and output gate, which regulate the flow of information into and out of the cell state. This structure allows the network to effectively capture long-term dependencies and handle issues like vanishing gradients, which can occur with standard recurrent neural networks. This capability allows LSTMs to maintain information over longer sequences and adapt to a wide range of sequential data tasks. LSTM units are a type of RNN specifically designed to handle long sequences and dependencies in data. The LSTM architecture consists of various components, each governed by specific equations:

The input gate controls the extent to which a new value flows into the cell state.

$$i_n = \sigma(W_{xi} \cdot x_t + W_{hi} \cdot h_{t-1} + b_i) \tag{7}$$

Here, i_n represents the input gate's activation, σ is the sigmoid activation function, W_{xi} and W_{hi} are the weight matrices, x_t is the input at the current time step, h_{t-1} is the previous hidden state, and b_i is the bias. The forget gate decides what information is discarded from the cell state. It is calculated as:

$$f_n = \sigma(W_{xf} \cdot x_t + W_{hf} \cdot h_{t-1} + b_f) \tag{8}$$

In this equation, f_n is the forget gate's activation, with W_{xf}, W_{hf}, x_t, h_{t-1}, and b_f analogous to those in the input gate equation. The output gate determines the part of the cell state to be outputted into the next layer. It is given by:

$$o_n = \sigma(W_{xo} \cdot x_t + W_{ho} \cdot h_{t-1} + b_o) \tag{9}$$

Where o_n is the output gate's activation, and the other variables are as previously defined. The cell state represents the memory component of the LSTM. It is updated as follows:

$$C_n = f_n * C_{t-1} + i_n * \widetilde{C}_t \tag{10}$$

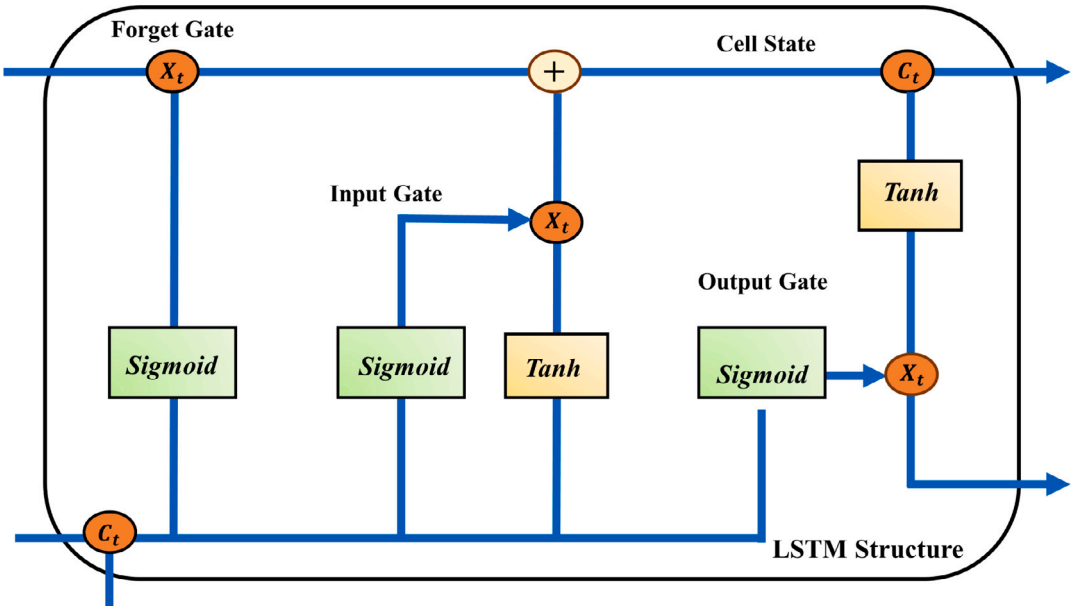


Fig. 3. Long Short-Term Memory (LSTM) unit with forget, input, and output gates.

C_n is the new cell state, C_{t-1} is the previous cell state, and $\widetilde{C}_t$ is the candidate cell state. The forget gate's output f_n and the input gate's output i_n modulate the cell state. The candidate cell state $\widetilde{C}_t$ is calculated by:

$$\widetilde{C}_t = \tanh(W_{xc} \cdot x_t + W_{hc} \cdot h_{t-1} + b_c) \tag{11}$$

This equation introduces a new candidate value, which is considered for addition to the cell state. Finally, the hidden state h_n is the actual output of the LSTM cell at each time step, defined as:

$$h_n = o_n * \tanh(C_n) \tag{12}$$

LSTM networks are a specific type of RNN that are particularly good at handling long-term connections, which is why they are useful in cybersecurity. Their ability to deal with the vanishing gradient problem makes them very effective at processing long sequences of data, which is essential in modern security settings such as IoT applications. LSTMs can be used to analyse real-time IoT data for identifying and reacting to cybersecurity threats in a dynamic manner, aiding in continuous threat surveillance and predictive security actions.

3.4. Asynchronous federated learning

Asynchronous Federated Learning (FL) represents a significant advancement in distributed machine learning. It is an approach where multiple clients, each with their own dataset, contribute to the training of a global model [26].

Fig. 4 illustrates the process flow of asynchronous Federated Learning (FL). In this model, a global model is first generated and then downloaded by the clients (nodes). Each client independently trains the model using its local data and sends the updated model parameters back to the server. The server aggregates these local model parameters without waiting for all clients to send their updates, allowing for a more flexible and potentially faster learning process compared to synchronous FL. The uniqueness of this method lies in the asynchronous nature of updates, which is particularly beneficial in environments with diverse computational capacities and network conditions [27,28]. Each client in the FL network is responsible for computing a local update based on its dataset D_k. This computation is aimed at optimising the local model by minimising a loss function L. The local update is mathematically formulated as:

In asynchronous Federated Learning (FL), the process of optimising the model parameters involves several critical mathematical equations. One key equation is:

$$\Delta w_k = \underset{\Delta w}{\operatorname{argmin}} \sum_{i \in D_k} L(x_i, y_i, w + \Delta w) \tag{13}$$

In this equation, Δw_k represents the optimal weight adjustment for the client k. This adjustment is determined by minimising the loss function L across all data points (x_i, y_i) in the client's dataset D_k. Here, x_i and y_i denote the input and output pairs, respectively, and w signifies the current parameters of the model. The objective is to find a change in the model's weights (Δw) that minimises the loss on the client's data. A crucial aspect of asynchronous FL is aggregating these individual updates into the global model. This

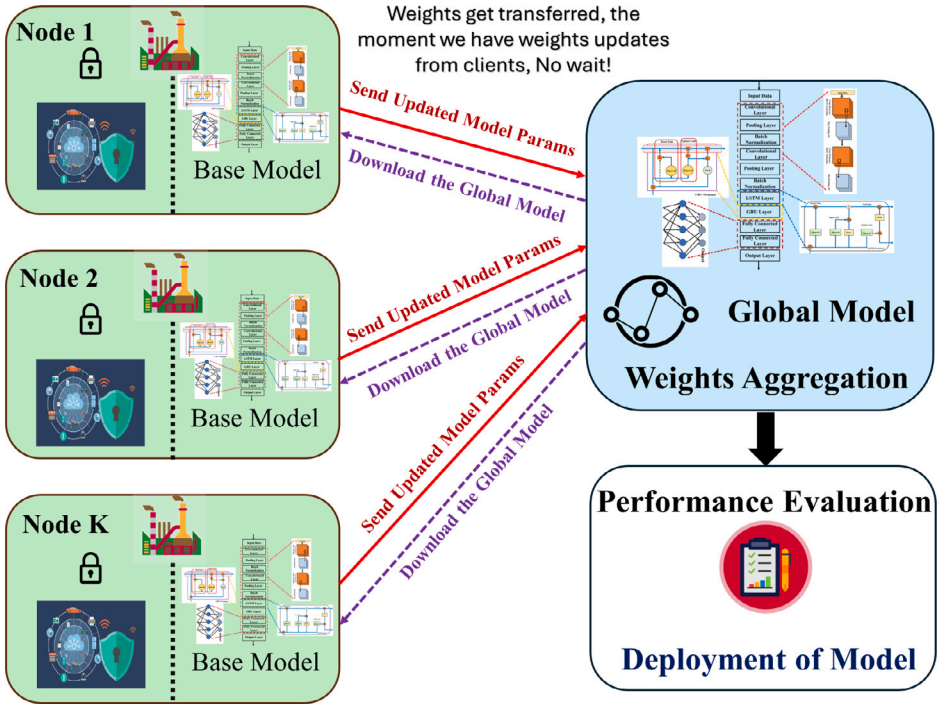


Fig. 4. Asynchronous Federated Learning (FL) process flow.

aggregation process is characterised by its asynchronous nature, allowing updates to be incorporated into the global model as they are sent by clients, without requiring synchronisation. The global model update is mathematically expressed as:

$$w(t+1) = w(t) + \eta \sum_{k} \alpha_k \Delta w_k \tag{14}$$

In this equation, $w(t)$ and $w(t+1)$ represent the global model parameters at times t and $t+1$, respectively. The term η is the learning rate, α_k the relative importance or weight assigned to the update from client k, and Δw_k the local update computed by client k. This asynchronous update mechanism enables the global model to evolve continuously, integrating diverse insights from various clients, thereby enhancing its performance and accuracy.

To further ensure the stability of the learning process in asynchronous FL, an additional equation is often employed to manage the learning rate:

$$\eta(t) = \frac{\eta_0}{1 + \delta t} \tag{15}$$

Here, $\eta(t)$ is the learning rate at time t, with η_0 being the initial learning rate and δ a decay factor. This equation adjusts the learning rate over time, reducing it as the model trains, which helps in stabilising the learning process and ensuring convergence. Asynchronous FL improves machine learning's capabilities in distributed networks by allowing for independent and non-synchronous updates from diverse client systems [29]. This approach is particularly advantageous in environments where real-time data processing is crucial, and client systems vary widely in computational power and network connectivity.

In asynchronous FL, each client autonomously updates the shared global model using their local data, eliminating the need for synchronisation with other clients. This independence is particularly effective in IoT and mobile computing environments, where data is continuously generated, allowing for real-time data integration and maintaining the model's accuracy and relevance over time [30]. Asynchronous FL is well-suited for dynamic settings, such as healthcare and mobile computing, which experience variable data rates and network conditions. It supports real-time data utilisation, crucial for immediate decision-making without the delays associated with batch processing. However, the asynchronous nature of updates can introduce inconsistencies in the global model due to simultaneous updates from multiple clients. To mitigate this, techniques like elastic averaging and adaptive learning rates are employed to synchronise updates towards a global average and manage data variability, respectively, ensuring stable training across diverse clients [31]. The conflict resolution mechanisms are essential in asynchronous FL to handle conflicts arising from concurrent modifications of the same model parameters, thereby maintaining the integrity and accuracy of the global model. These mechanisms are particularly beneficial in Edge IoT environments, where devices operate under varied capacities and network conditions, allowing each client to independently update the model, which facilitates quicker convergence and optimal resource utilisation [30,31]. The process is depicted in Fig. 5, which illustrates a flowchart within a dotted outline, representing the components of a larger system designed for intrusion detection via federated learning. This decentralised approach enables multiple participants to contribute to a machine-learning model without sharing their data, thus preserving privacy and reducing communication overhead.

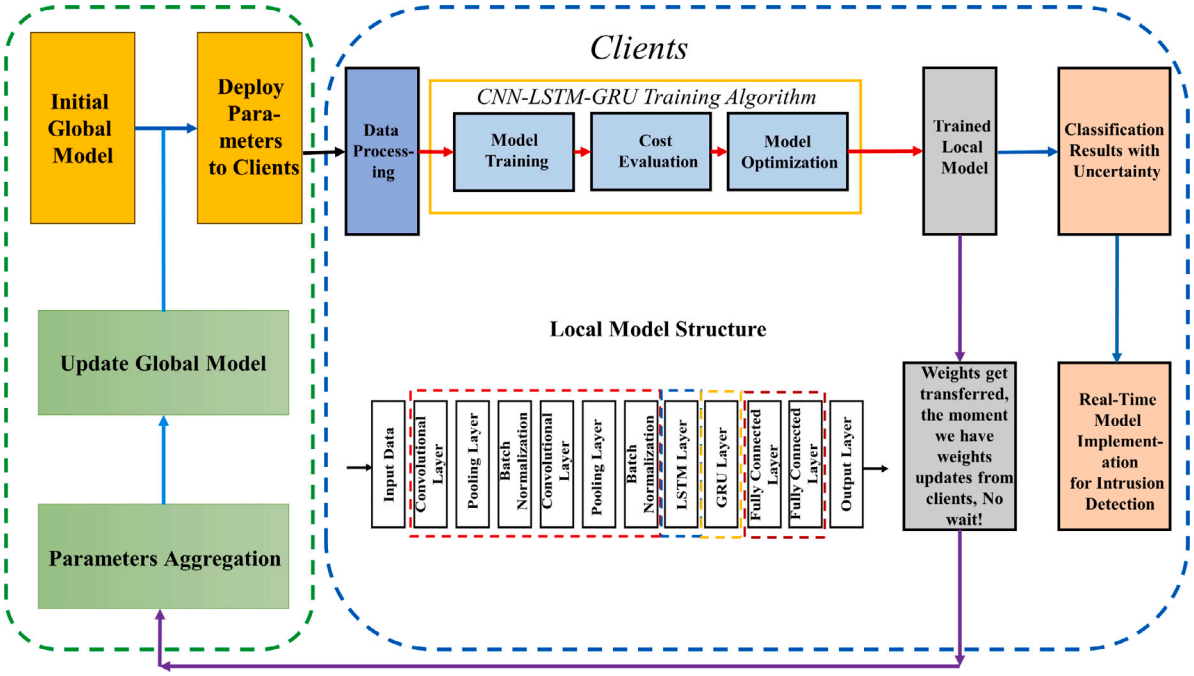


Fig. 5. Detailed architecture of asynchronous FL-based IDS system.

3.5. Integration of CNN, LSTM, and GRU with asynchronous FL

The integration of deep learning models such as CNNs, LSTMs, and GRUs represents a novel approach to addressing the complexities of cybersecurity in IoT environments. Each model brings unique strengths that are synergistically combined to increase the detection and analysis of security threats in vast and varied data streams.

As depicted in Fig. 6, our proposed model architecture utilises CNN layers for their robust feature extraction capabilities, which are critical for identifying nuanced patterns in data indicative of potential security breaches. This is followed by the application of the LSTM unit, which excels in processing sequential data, thus capturing temporal dependencies necessary for understanding the context within which anomalies occur. Complementing this, GRUs are employed to efficiently manage the flow of information, ensuring that only relevant data influences the decision-making process. This integrated approach not only improves the model's ability to discern complex sequences and anomalies but also optimises the learning process across a distributed network. The flexibility afforded by asynchronous FL allows each client within the network to contribute independently to the global model, catering to the diverse and dynamic nature of IoT data. This model architecture is particularly tailored to the demands of cybersecurity in distributed systems, where rapid and accurate threat detection is paramount. The process of integrating CNN-LSTM-GRU models within an asynchronous FL framework is outlined through a pseudo algorithm, detailing the workflow from initialisation to prediction. As shown in Algorithm 1, the integration of CNN, LSTM, and GRU models with asynchronous federated learning can significantly improve the performance of intrusion detection systems.

The integration of CNNs, LSTMs, and GRUs within an asynchronous FL framework is structured to maximise the efficiency and effectiveness of cybersecurity measures in IoT environments. Below is a detailed workflow that outlines the key steps involved in this integration, ensuring a robust and dynamic learning process across distributed networks.

1. **Server Initialisation**: The process begins with the server initialising the global model, which serves as the foundational framework for all client models within the federated network.
2. **Client-Side Operations**: Each client in the FL network initialises its local CNN-LSTM-GRU model using the global model's weights. This step guarantees consistency across the network and fosters collaborative learning from the outset, leveraging the unique strengths of each model component for superior feature extraction and sequence modelling.
3. **Asynchronous Learning Process**: At the core of the model's functionality is the asynchronous update mechanism, where clients independently train their local models on their respective datasets and compute updates. These updates are then sent back to the server without waiting for synchronisation with other clients.

    - **Local Training**: Each client utilises its dataset to train the model locally, tailoring the learning process to specific data characteristics.
    - **Model Update**: Post training, each client computes an update and sends this back to the server, contributing to the global model's improvement.

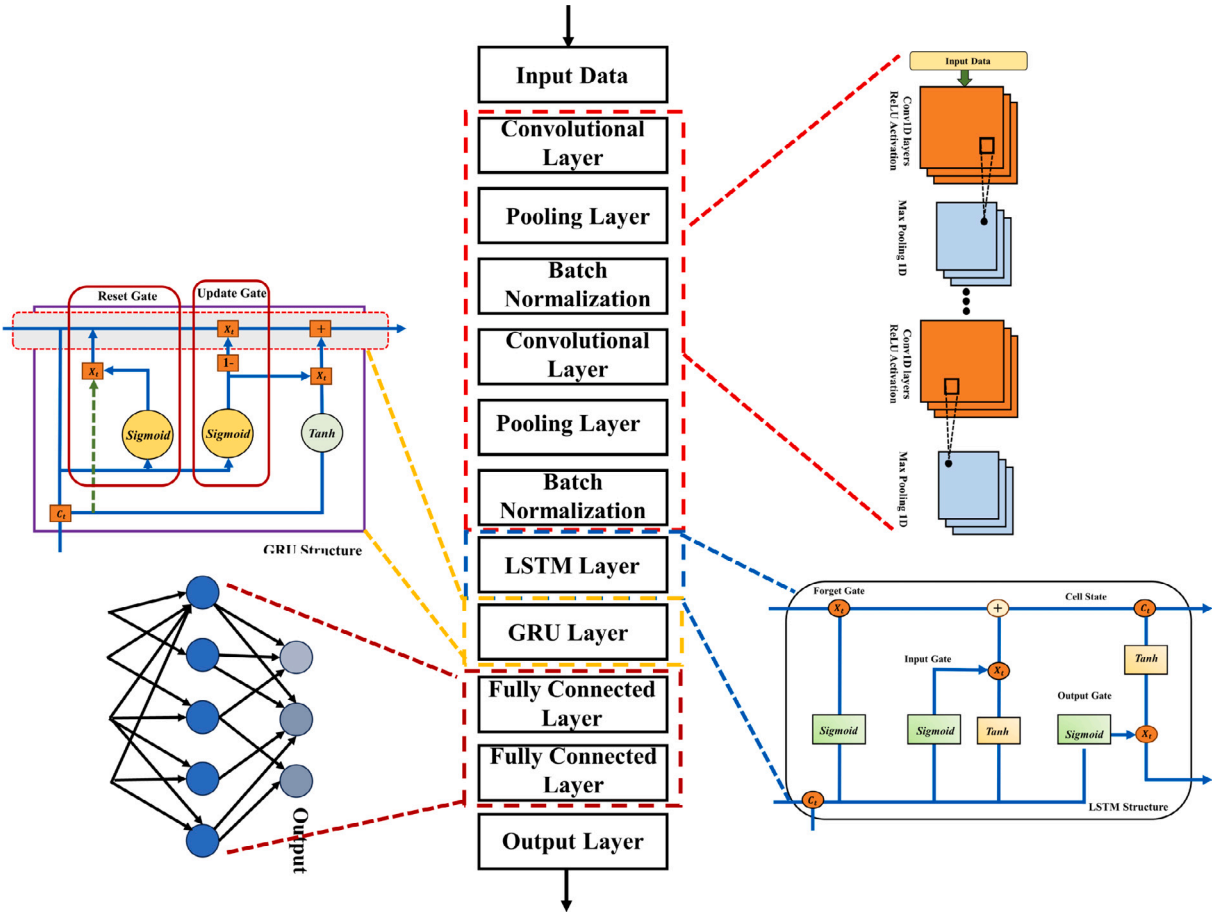


Fig. 6. Combined CNN, GRU, and LSTM model architecture.

4. **Server-Side Aggregation**: The server receives updates asynchronously and aggregates them to improve the global model. This step is crucial for integrating diverse insights from various clients, thus enriching the model's learning and adaptability.
5. **Model Distribution**: After aggregation, the server redistributes the updated global model back to each client. This ensures that all clients are equipped with the latest model version, maintaining uniformity and synchronisation across the network.
6. **Convergence and Prediction**: The cycle of updates and distributions continues until the model achieves convergence. Upon convergence, clients are equipped to perform predictions using the refined model, either locally or centrally, enabling real-time response to cybersecurity threats.

This structured approach not only leverages the individual strengths of CNNs, LSTMs, and GRUs but also harnesses the collective power of these models in a federated learning context. The asynchronous nature of the updates improves learning efficiency and allows the model to rapidly adapt to new and evolving data patterns, all while maintaining client data privacy and minimising communication overhead. Such a dynamic system is particularly suited for applications requiring high responsiveness and adaptability across distributed and diverse computing environments.

3.6. Importance of hyperparameter tuning in deep learning models

Hyperparameters are critical parameters that are set before the training process begins and significantly influence the behaviour and performance of deep learning models. Unlike model parameters that are learned during training, hyperparameters must be defined in advance and are crucial for guiding the training process. In our deep learning model designed for cybersecurity in the IIoT setting, several hyperparameters need careful tuning to ensure optimal performance:

- **Number of Units in Layers**: The number of neurons in Conv1D and LSTM layers affects the model's capacity to learn complex patterns. For instance, our model uses 400 units in the first LSTM layer to capture detailed sequence dependencies.
- **Kernel Size**: In Conv1D layers, the kernel size impacts how the model perceives input data, influencing its ability to recognise patterns of varying scales.

Algorithm 1 Integration of CNN-LSTM-GRU with Asynchronous FL

Server Initialisation:
Initialise global model weights $w(0)$
Client-Side Operations:
for each client k **do**
    Initialise local CNN-LSTM-GRU model with weights $w_k(0) \leftarrow w(0)$
Asynchronous Federated Learning:
while not converged **do**
    for each client k asynchronously **do**
        Fetch current global model weights $w(t)$
        Train local model on local dataset D_k
        Compute local update Δw_k using loss function L
        Send update Δw_k to server
    Server-Side Aggregation:
    for each received update Δw_k **do**
        Update global model: $w(t+1) \leftarrow w(t) + \eta \alpha_k \Delta w_k$
    Distribute Updated Global Model:
    for each client k **do**
        Send updated global model $w(t+1)$ to client k
Prediction:
for each client k **do**
    Perform local or central predictions using the final model

Table 2
Key hyperparameters in the CNN-LSTM-GRU model.

Hyperparameter	Value
Conv1D filters	400
Conv1D kernel size	2
LSTM units (1st, 2nd, 3rd layers)	400, 300, 200
GRU units	150
Dropout rate	0.1–0.3
Activation function	relu, softmax
Optimizer	adam
Loss function	sparse categorical cross-entropy

- **Dropout Rate**: This is essential for preventing overfitting by randomly dropping units during training phases, thus ensuring the model generalises well to new, unseen data.
- **Activation Functions**: Functions like ReLU (used in Conv1D and Dense layers) help introduce non-linearity, enabling the model to learn more complex functions.
- **Optimizer and Loss Function**: The choice of optimiser and loss function directs how the model adjusts its weights in response to the error it observes, impacting convergence speed and stability.

The exact adjustment of these hyperparameters is crucial for our model, ensuring it avoids both underfitting and overfitting while maintaining computational efficiency. Particularly in IIoT cybersecurity, well-tuned hyperparameters improve the model's capacity to accurately identify and categorise cyberattacks, ensuring precision and resilience against diverse attack vectors in IIoT settings. This strategy safeguards data privacy and optimises the use of distributed computational resources, vital for prompt threat detection and control in decentralised networks (see Table 2).

4. Dataset description

The Edge-IIoTset Dataset, utilised in our study, was originally developed and detailed by Ferrag et al. [3], who meticulously curated a comprehensive dataset reflective of realistic IIoT environments. This dataset, significant in scale, captures a wide array of interactions and cyberattacks over a period from November 21, 2021, to January 10, 2022. It was initially comprised of 1176 features, distilled down to 61 high-correlation features to improve manageability and relevance for effective cyberattack detection. Our research leverages this established dataset to demonstrate the efficacy of our novel security analysis model. We particularly focus on the application of the dataset in a federated learning framework, where it serves as the foundation for assessing the robustness of our proposed cybersecurity solutions. This involves utilising the dataset's diverse scenarios of normal operations and cyberattacks, including DoS/DDoS attacks, Information Gathering, Man-in-the-Middle attacks, Injection Attacks, and Malware Attacks, to validate our model's performance in detecting and mitigating potential threats. The composition of the dataset includes an extensive setup

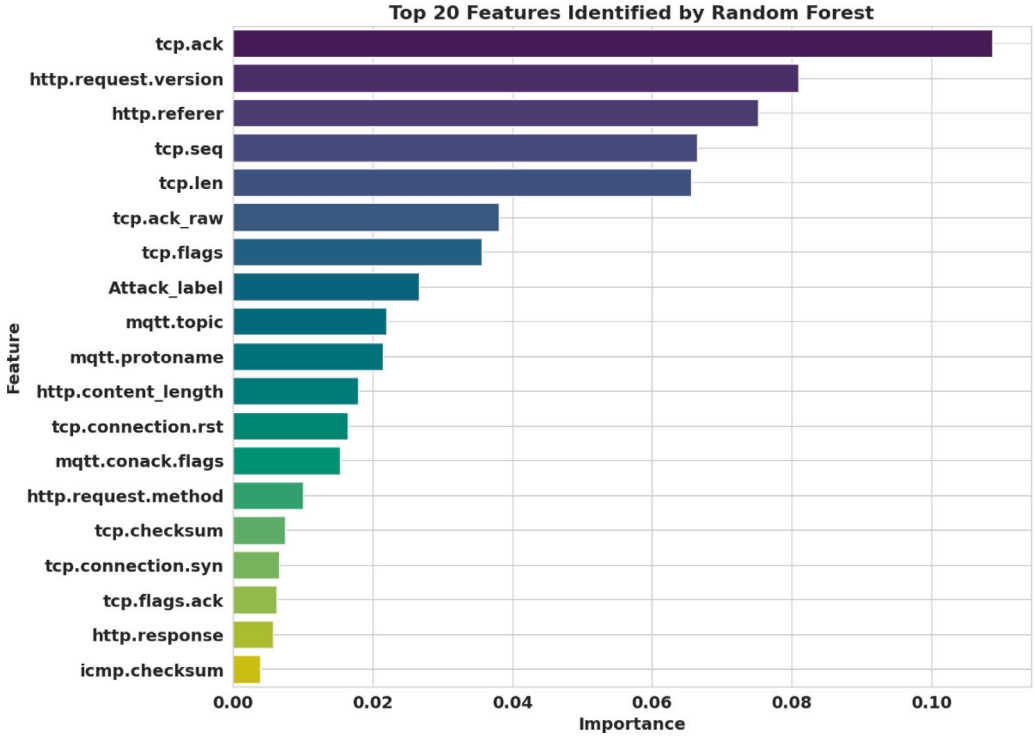


Fig. 7. Top 20 most important features identified by the random forest model.

of network layers—ranging from cloud computing to Software-Defined Networking (SDN)—which are critical for simulating an authentic IIoT environment. The detailed phase of data collection and feature extraction was adapted in our study to test the resilience and accuracy of our intrusion detection algorithms under varied conditions. This comprehensive portrayal of IIoT security dynamics supports the testing and validation of our models, particularly highlighting the effectiveness of federated learning in enhancing data privacy and model robustness across decentralised networks.

Fig. 7 presents the top 20 features identified by a Random Forest classifier, ranked according to their importance in predicting the Attack types in our dataset. The Random Forest algorithm was employed due to its efficacy in handling high-dimensional data and its robustness against overfitting, which is particularly beneficial in analysing complex datasets with a large number of features. This approach not only helped in determining the most influential features but also in enhancing the overall interpretability of our model's predictions. The selected features significantly improve the predictive accuracy of the models tested, including our proposed algorithm. The visualisation serves as a crucial tool for understanding the feature dynamics within the dataset, offering insights into which variables most significantly impact the classification process. By highlighting these key features, we can better align our data collection and preprocessing efforts to focus on the most impactful variables, thereby optimising the performance of our proposed security solutions in real-world scenarios. This methodological choice highlights the complementary relationship between traditional machine learning techniques and advanced models, illustrating how foundational algorithms like Random Forest can validate and improve the deployment of new predictive strategies in cybersecurity.

4.1. Attacks in edge-IIoTset dataset

The integrity and diversity of the Edge-IIoTset Dataset, encompassing both legitimate and malicious entries, are pivotal for establishing a normal behavioural profile for IIoT systems and for identifying novel attack patterns. As illustrated in Fig. 8, the Kernel Density Estimates (KDE) provide insights into the distribution of six key features in our dataset, selected specifically for visualisation due to the extensive number of features available. These plots are particularly valuable for understanding the underlying patterns and variations within each feature. KDEs offer a smooth representation of data distribution, which is essential for identifying skewness, modality (i.e., whether the data has multiple peaks), and other distribution characteristics. These visualisations help us grasp the complex behaviour of network traffic under both normal and attack scenarios, emphasising features that show significant variances. This selective visualisation approach not only makes the analysis more manageable but also allows us to focus on features that are most indicative of the security state within IIoT systems. For instance, the KDE for the feature packet length might reveal a bi-modal distribution, which can indicate typical and atypical packet sizes that characterise normal operations and cyberattack scenarios, respectively. By highlighting these distinctions, we augment our model's ability to discriminate between benign and malicious behaviours, thereby improving the reliability of our intrusion detection system. The study categorises the multitude of

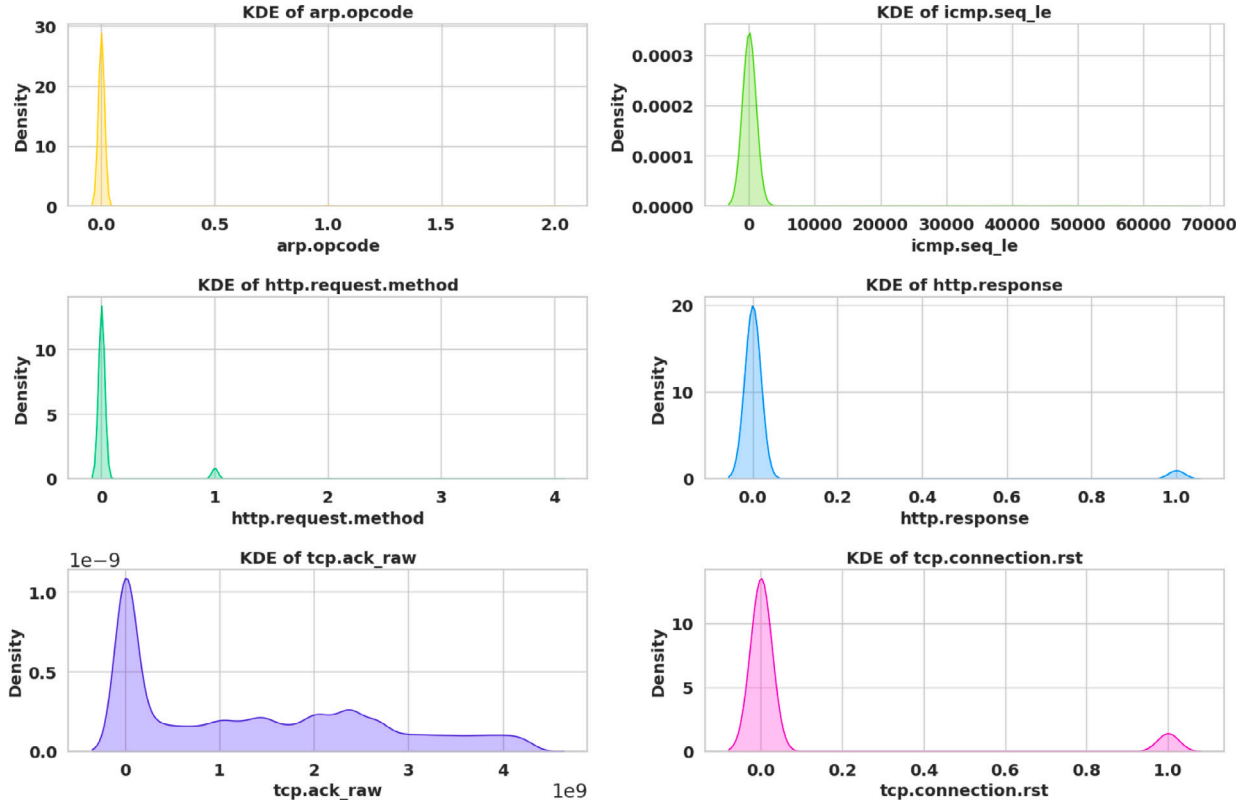


Fig. 8. Kernel density estimates of selected features.

cyberattack methodologies into clearly defined groups, each comprising various techniques aimed at compromising the integrity and availability of IIoT systems [32,33].

4.1.1. DoS/DDoS attacks

DoS and DDoS attacks, critical components of this dataset, are designed to disrupt the normal function of services by denying access to legitimate users. These attacks include TCP SYN Flood, where attackers exploit the TCP handshake protocol to overload the server with SYN requests, and UDP Flood attacks that involve sending numerous UDP packets to random ports on a host, saturating the network with ICMP 'Destination Unreachable' responses. HTTP Flood attacks overwhelm the target with HTTP requests, incapacitating web servers, while ICMP Flood attacks aim to deplete the network's resources by flooding it with ICMP packets.

4.1.2. Information gathering

The dataset also addresses the initial phase of most cyberattacks, which involves information gathering. This phase includes Port Scanning to identify accessible ports on a system, OS Fingerprinting to determine the operating system of a host, and Vulnerability Scanning, which is an automated process aimed at identifying security weaknesses within the network or infrastructure.

4.1.3. Man-in-the-middle attacks

Included in the dataset are Man-in-the-Middle (MitM) attacks, which intercept and alter communications between two parties who believe they are directly communicating with each other. This category features DNS Spoofing, which corrupts the DNS resolution process to redirect traffic to an attacker-controlled system, and ARP Spoofing, where falsified ARP messages link an attacker's MAC address with the IP address of a legitimate computer or server.

4.1.4. Injection attacks

Injection attacks, aiming at the integrity and confidentiality of systems, come in various forms within our dataset. These include Cross-site Scripting (XSS), which allows attackers to inject malicious scripts into benign websites; SQL Injection, which manipulates backend databases through malicious SQL queries; and Uploading Attacks, where attackers upload malicious files to a server to execute unauthorised activities.

Table 3
Mapping of specific attack types to general attack classes.

General attack class	Specific attack types
Normal	Normal
DDoS	DDoS_UDP, DDoS_ICMP, DDoS_TCP, DDoS_HTTP
Injection attacks	SQL_injection, Uploading, XSS
Malware attacks	Password, Backdoor, Ransomware, MITM
Information gathering	Vulnerability_scanner, Port_Scanning, Fingerprinting

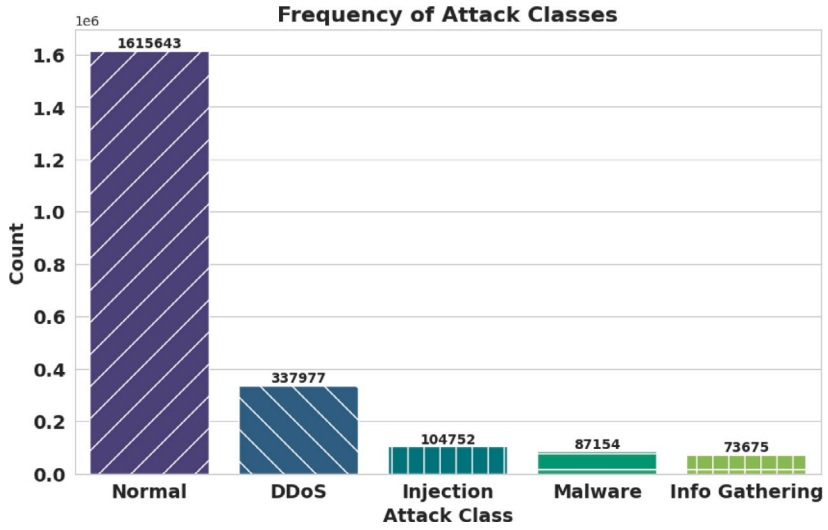


Fig. 9. Frequency distribution of attack classes.

4.1.5. Malware attacks

This dataset encapsulates a range of malware exploits, including Backdoor Attacks that allow remote system control, Password Cracking using brute force or dictionary methods, and Ransomware, which blocks system or data access until a ransom is paid. These varied attack methodologies are critical for establishing a robust behavioural profile for IIoT systems and aid in the development and testing of novel intrusion detection systems.

In this study, we systematise various attack types into broader classes to streamline the analysis of cybersecurity threats within the IIoT framework. Table 3 outlines this systematic categorisation, consolidating diverse attacks into five principal classes: 'Normal', 'DDoS', 'Injection attacks', 'Malware attacks', and 'Information gathering'. Each class aggregates specific types of attacks that exhibit common behaviours or goals, enhancing our approach to identifying and mitigating cybersecurity threats. For example, the 'DDoS' class encompasses methods like 'DDoS_UDP', 'DDoS_ICMP', 'DDoS_TCP', and 'DDoS_HTTP', each representing a variant of Distributed Denial of Service attacks. Fig. 9 illustrates the frequency distribution of these attack classes within our dataset. The bar chart provides a visual representation of the prevalence of each attack class, highlighting how often 'Normal' activities occur in comparison to attack scenarios. This visualisation is instrumental for gauging the significance of each attack type, crucial for the development of targeted cybersecurity strategies and refined models for anomaly detection in IIoT environments.

5. Results and discussion

To effectively address the challenges of data privacy and efficiency in cyberattack detection across distributed environments, our proposed model utilised an asynchronous FL framework. The dataset was distributed among ten different clients, mimicking a realistic IIoT network scenario where data is inherently decentralised. Each client independently trained a local model on their specific data subset, utilising a combination of CNN, LSTM, and GRU architectures to improve feature extraction and sequence modelling capabilities. The core innovation of our approach involved asynchronously aggregating the learned weights from these local models to construct a robust global model. This global model leveraged the aggregated intelligence to classify attack types accurately. The asynchronous nature of our FL framework facilitated flexible and efficient model training, as it allowed for updates without the need for all clients to synchronise simultaneously. This strategy not only maintained the privacy of the data on individual clients but also capitalised on their collective intelligence, significantly improving the overall predictive accuracy of our system.

Table 4
Classification report by attack class.

Attack class	Precision	Recall	F1-Score	Support
Normal	1.00	1.00	1.00	1 615 643
DDoS	1.00	1.00	1.00	337 977
Injection Attacks	1.00	1.00	1.00	104 752
Malware Attacks	1.00	1.00	1.00	87 154
Information Gathering	1.00	1.00	1.00	73 675
Overall Accuracy		1.00 (443700 instances)		
Macro Avg	1.00	1.00	1.00	443 700
Weighted Avg	1.00	1.00	1.00	443 700

5.1. Metrics

The performance evaluation of our proposed model in the IIoT cybersecurity context utilises several key metrics, each offering distinct insights into different aspects of the model's effectiveness. The mathematical formulations of these metrics are presented below.

Accuracy is calculated as the ratio of correctly predicted observations to the total observations:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \tag{16}$$

High accuracy, as shown in Eq. (16), indicates the model's general capability in accurately classifying both attack and normal activities.

Precision is crucial in minimising false positives, defined as the ratio of correctly predicted positive observations to the total predicted positive observations:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{17}$$

In cybersecurity, precision, detailed in Eq. (17), is vital for ensuring that actual attacks are correctly identified, reducing the risk of false alarms.

Recall or Sensitivity measures the model's ability to identify all actual positive cases:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{18}$$

High recall, as described in Eq. (18), is critical to ensure the detection of the majority of attacks, despite the risk of false positives.

F1-Score provides a balance between Precision and Recall, calculated as the harmonic mean of Precision and Recall:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{19}$$

This metric, captured in Eq. (19), is particularly useful in imbalanced datasets as it encapsulates both false positives and false negatives into a single metric.

ROC-AUC represents the area under the ROC curve, which plots the True Positive Rate (Recall) against the False Positive Rate. The AUC quantifies the overall performance of the model:

$$\text{AUC} = \int_0^1 \text{TPR}(x)\, dx \tag{20}$$

A higher AUC, detailed in Eq. (20), indicates a better-performing model capable of effectively distinguishing between classes.

5.2. Comprehensive performance evaluation

Our deep learning model demonstrates exceptional performance in cybersecurity detection within the IIoT environment, as evidenced by the detailed results in our classification report (Table 4). This report clearly shows that the model achieves perfect precision, recall, and F1-score of 1.00 across all categories, including 'Normal', 'DDoS', 'Injection Attacks', 'Malware Attacks', and 'Information Gathering'. These results indicate flawless identification and classification capabilities for each attack type, demonstrating the model's robustness and adaptability across varied cyber threat scenarios.

The consistent performance across different attack scenarios highlights the model's strength and flexibility. The 'Support' numbers within the report reflect the dataset's composition, with 'Normal' activities being predominant. Nevertheless, the model maintains high accuracy even with less frequent attack types. The overall accuracy of 1.00, calculated from a substantial dataset of 443,700 instances, confirms the model's reliability. Both the macro and weighted averages also stand at 1.00, underscoring the model's consistent performance across all classes, irrespective of their occurrence frequency in the dataset. Such precision is vital in the IIoT cybersecurity context, where the accurate detection of a wide range of attack types is essential for effective threat prevention.

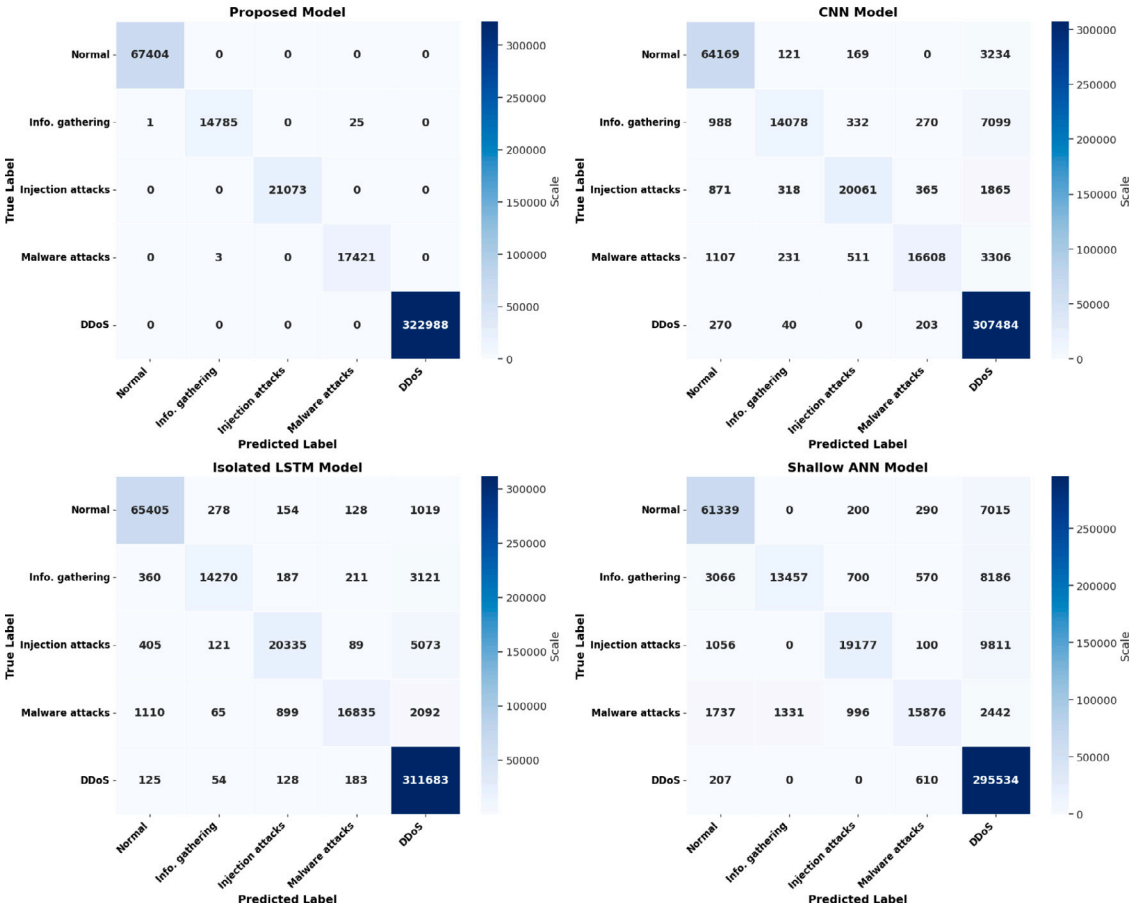


Fig. 10. Confusion matrix of the competing techniques for the Edge-IIoT dataset.

5.2.1. Analysis of the confusion matrix

The confusion matrix shown in Fig. 10 provides a visual representation of the model's accuracy in classifying different attack classes within the IIoT dataset. The prevalence of true positive predictions in the diagonal cells indicates the model's high accuracy. It showcases the model's ability to detect various attack types such as DDoS, normal activities, injection attacks, malware attacks, and information gathering. The model exhibits precision and reliability in distinguishing normal activities from different cyber threats, ensuring strong cybersecurity for IIoT systems. These findings validate the model's effectiveness in accurately identifying and categorising various cyberattack types, highlighting its performance and resilience in the IIoT cybersecurity domain.

5.2.2. Bar chart of performance metrics

Further illustrating the model's outstanding performance, the bar chart in Fig. 11 displays the precision, recall, and F1-score for each class within the IIoT dataset. Remarkably, all metrics for the classes 'Normal', 'Information Gathering', 'Injection Attacks', 'Malware Attacks', and 'DDoS' achieve the maximum score of 1.0. This exceptional level of uniformity across all classes not only reflects the model's acute precision in predicting true positives but also its effectiveness in minimising both false positives and false negatives. Such consistently high scores across varied metrics emphasise the model's efficacy as a dependable tool for comprehensive cybersecurity in IIoT environments, ensuring robust and accurate threat detection and classification.

5.2.3. Receiver Operating Characteristic (ROC) curve analysis

The ROC curve is a fundamental tool for evaluating the performance of a classification model. As depicted in the figures, different models exhibit varying levels of discriminative ability between the different classes within the IIoT dataset. Fig. 12 shows the ROC curves for the CNN model, which demonstrates good classification performance with high AUC values across all classes. The CNN model effectively identifies most classes but shows a slight variation in performance for different attack types. Fig. 13 presents the ROC curves for the Isolated LSTM model. This model shows a slight improvement over the CNN model, with higher AUC values, indicating better performance in distinguishing between normal and attack scenarios. Fig. 14 displays the ROC curves for the proposed model, which integrates CNN, GRU, and LSTM architectures. The proposed model achieves nearly perfect classification with AUC values of 1.00 for all classes, demonstrating exceptional discriminative ability and confirming its robustness in accurately

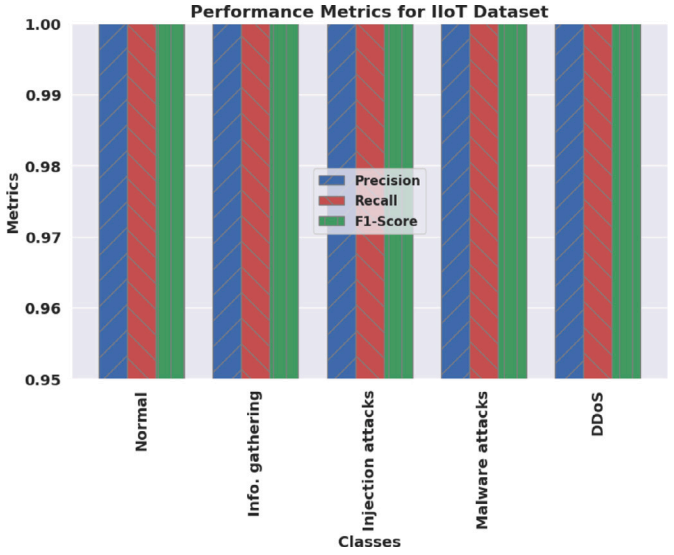


Fig. 11. Performance metrics bar chart for IIoT dataset.

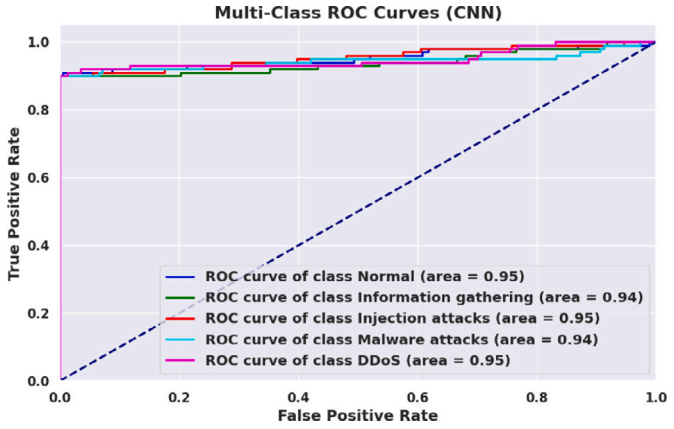


Fig. 12. Multi-class ROC curve for the CNN Model.

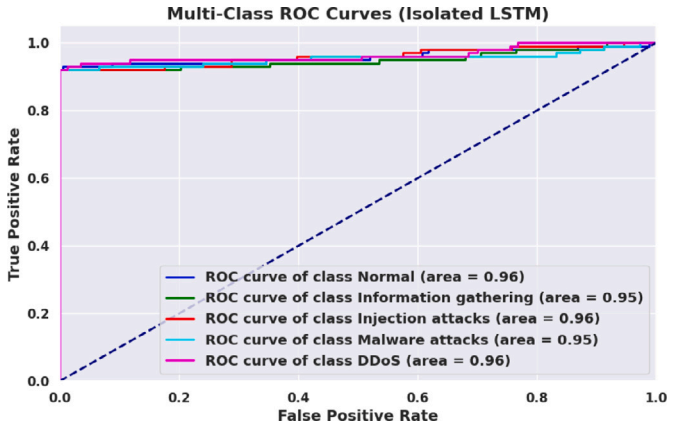


Fig. 13. Multi-class ROC curve for the isolated LSTM model.

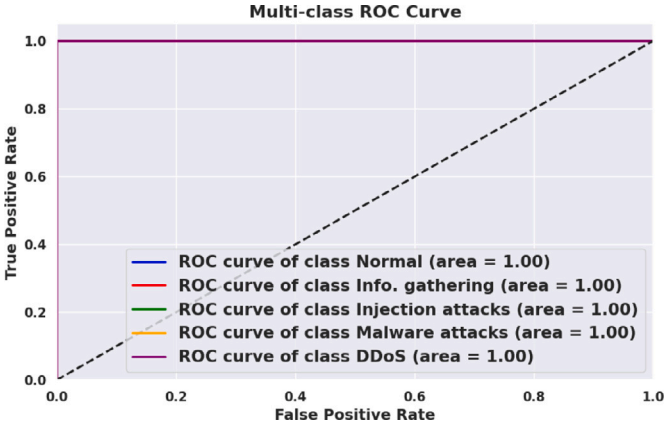


Fig. 14. Multi-class ROC curve for the proposed model.

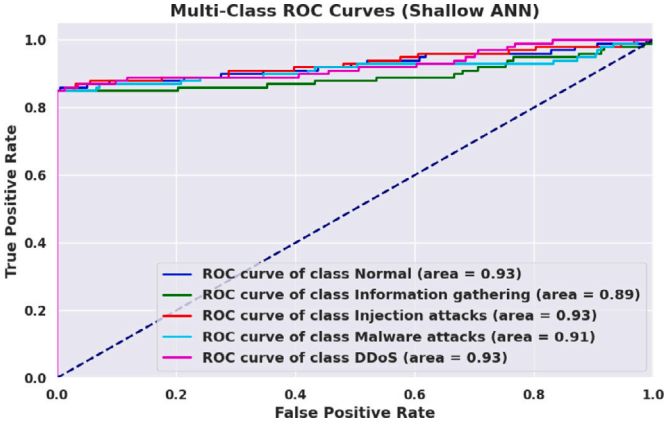


Fig. 15. Multi-class ROC curve for the shallow ANN model.

identifying and classifying various cyberattack types. Fig. 15 illustrates the ROC curves for the Shallow ANN model. While this model performs adequately, achieving reasonable AUC values, it does not match the performance of the more complex models, particularly the proposed hybrid model. Overall, the ROC analysis confirms that the proposed model outperforms other models by a significant margin, with near-perfect classification accuracy across all classes. This superior performance underscores the effectiveness of the proposed hybrid model in enhancing cybersecurity within IIoT systems, ensuring reliable threat detection and maintaining operational stability.

5.2.4. Round-wise average metrics across clients

The iterative nature of our proposed model's training process involves multiple rounds of learning, where each round contributes to the refinement and enhancement of the model's performance. Fig. 16 illustrates the average metrics achieved across clients over these rounds — accuracy, precision, and recall. As the rounds progress, a consistent improvement in all metrics is observed, indicating the model's increasing proficiency in correctly identifying and classifying cyberattack types. Notably, accuracy shows a steady upward trend, reflecting the model's improved generalisation capabilities with each successive round. Precision and recall also exhibit significant improvement, crucial for detecting cyber threats in IIoT systems. This upward trend in metrics emphasises the benefits of the FL approach, where aggregated insights from distributed clients cumulatively improve the model's predictive power. By round 20, the model achieves near-optimal performance, with all metrics approaching the maximum value of 1.00. This trend highlights the model's effectiveness in learning from distributed data sources, showcasing its potential for real-world deployment in safeguarding IIoT environments against a diverse array of cyber threats.

The rigorous evaluation of the asynchronous FL model across the IIoT dataset has demonstrated exceptional performance, underpinned by comprehensive quantitative metrics. The model achieved perfect scores in precision, recall, and F1-score across all attack classes, signifying its superior capability in cyberattack detection within IIoT environments. The integration of CNN, LSTM, and GRU models has proven effective in capturing complex patterns and dependencies in the data, contributing to the model's robust predictive accuracy. These results are further validated by a detailed statistical analysis, including ROC curve assessments that indicate an Area Under the Curve (AUC) of 1.00 for all classes, confirming the model's flawless discrimination between normal

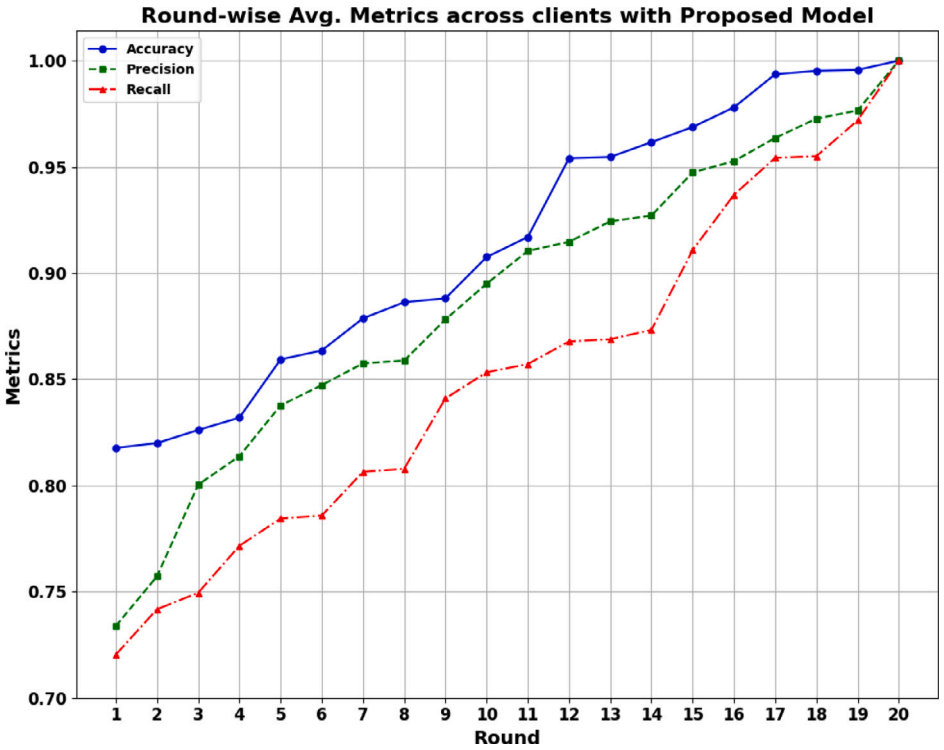


Fig. 16. Round-wise average metrics scores across clients with the proposed model.

Table 5
Performance metrics across different hyperparameter configurations.

Configuration	Filters	Units	Dropout	Accuracy	Precision	Recall	F1-Score	Support
Config 1	200	100, 50	0.2	95%	95%	94%	94.5%	443 700
Config 2	300	200, 100	0.3	97%	97%	97%	97%	443 700
Config 3	350	250, 150	0.1	98%	98%	98%	98%	443 700
Optimal	400	400, 300, 200	0.1–0.3	100%	100%	100%	100%	443 700

operations and various cyberattacks. The iterative improvement of accuracy, precision, and recall over multiple rounds of learning, as shown in our round-wise evaluation, illustrates the model's adaptability and learning efficiency, which are critical for real-time and dynamic IIoT scenarios. This performance not only meets but exceeds the benchmarks set by existing cybersecurity solutions in the IIoT domain, establishing a new standard for future developments in the field.

5.3. Evaluation of model performance across different hyperparameter settings

This section outlines the performance of our deep learning model across various hyperparameter configurations. The final row showcases the optimal setting that achieves perfect scores in all evaluated metrics.

As illustrated in Table 5, varying the number of filters and LSTM/GRU units, along with adjusting the dropout rates, has significant effects on the model's performance. The optimal configuration, which utilises 400 filters in the Conv1D layer, LSTM units set at 400, 300, and 200 for the successive layers, and dropout rates between 0.1 and 0.3, achieve the highest possible scores across all metrics. This configuration is particularly effective in our cybersecurity application within the IIoT environment, ensuring the model is highly accurate, robust against overfitting, and capable of handling the diverse and complex data patterns characteristic of cyberattack detection.

5.4. Comparative analysis

This subsection presents a comparative analysis of various deep learning techniques applied in IoT environments, with a focus on their effectiveness in addressing the challenges of cyberattack detection in IIoT systems.

This Table 6 provides a succinct overview of how various deep learning models perform across different IoT and IIoT datasets, highlighting the effectiveness of each method in cyberattack detection. Our model's integration within an asynchronous Federated

Table 6
Comparative analysis of different techniques on used datasets.

Ref.	Year	Technique	Summary	Results	Privacy	Centralised
[34]	2022	EECA-LSTM	Employs PCA for feature selection and LSTM for classification.	Accuracy: 0.996, Recall: 0.996, Precision: 0.996, F1-Score: 0.995	✓	×
[35]	2022	DenseNet & Inception Time	Utilises DenseNet and Inception Time for multi-class classification of cyber-attacks across ToN-IoT, Edge-IIoT, and UNSW2015 datasets.	Accuracy: 100% on Windows 10 with Inception Time, 99.9% with DenseNet; 94.94% on Edge-IIoT; 98.4% on UNSW-NB15; 98.6% with sliding window enhancement.	✓	✓
[36]	2022	LSTM-KPCA	End-to-end network attack detection with LSTM.	Accuracy: 0.98, Precision: 0.91, Recall: 0.84, F1-Score: 0.87	×	×
[37]	2023	Stacked unsupervised FL	Utilises unsupervised FL and deep autoencoder for NIDS.	Accuracy: 0.98, Recall: 0.88, Precision: 0.91, F1-Score: 0.90	✓	×
[10]	2023	ML with PCC and IF	Utilises machine learning with Pearson's Correlation Coefficient for feature selection and Isolation Forest for outlier removal to enhance IIoT security.	ACC: 99.99% on Bot-IoT, 99.12% on WUSTl_IIoT_2021; MCC: 92.17% on Bot-IoT, 93.96% on WUSTl_IIoT_2021; AUC: 92.48% on Bot-IoT, 99.3% on WUSTl_IIoT_2021	✓	✓
[38]	2024	FL-MA	Combines ML and the Firefly Algorithm for WSN security.	Accuracy: 0.992, Recall: 0.981, Precision: 0.995, F1-Score: 0.962	×	✓
Our Model	2024	Asynchronous FL-CNN-GRU-LSTM	Employs CNN, GRU, and LSTM in an asynchronous FL setting for IIoT security.	Accuracy: 1.00, Recall: 1.00, Precision: 1.00, F1-Score: 1.00	✓	✓

Table 7
Performance comparison of different models.

Model description	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Shallow ANN	91.7%	91.2%	91.5%	91.4%	0.915
Isolated LSTM Model	96.5%	96.3%	96.4%	96.4%	0.964
CNN Model	95.2%	95.0%	95.1%	95.1%	0.952
Proposed CNN-GRU-LSTM Model	100%	100%	100%	100%	1.00

Learning framework showcases superior performance metrics, reflecting its robust capacity to handle complex IIoT security challenges without requiring simultaneous updates from all network nodes. This approach not only preserves data privacy but also leverages the collective intelligence of distributed data points to improve the overall predictive accuracy of our system.

We also performed a comparative performance analysis to showcase the superiority of our deep hybrid learning model, which merges CNN, GRU, and LSTM networks in an asynchronous federated learning framework. This comparison not only proves the feasibility of our approach but also emphasises its advantages over simpler neural network models commonly used in similar situations. The models examined include the Shallow Artificial Neural Network (ANN), a basic model with fewer hidden layers serving as a baseline; the standalone LSTM Model, concentrating solely on LSTM layers to capture temporal dependencies; the CNN Model, employing only convolutional layers for effective spatial feature extraction; and the Proposed CNN-GRU-LSTM Model, our advanced hybrid model that combines the strengths of CNN, GRU, and LSTM architectures. The models were assessed using the same dataset and under comparable federated learning conditions to ensure a fair comparison. The key metrics used for evaluation included accuracy, precision, recall, F1-score, and ROC-AUC.

The results in Table 7 show that all models perform well, but our proposed CNN-GRU-LSTM model achieves perfect scores on every metric. This demonstrates its strength and effectiveness in handling complex IIoT cybersecurity tasks. The Shallow ANN, while adequate, shows the lowest performance, indicating that complexity and depth in neural architectures enhance model capability significantly. The isolated LSTM and CNN models, specialising in temporal and spatial features, respectively, performed better than the shallow ANN but did not reach the effectiveness of the hybrid approach. The results of this comparison clearly show that our proposed model is superior, taking advantage of the best features of CNN, GRU, and LSTM architectures to improve performance and detect cyberattacks more efficiently in IIoT settings.

5.5. Discussion

Our proposed deep hybrid learning model, which combines CNN, GRU, and LSTM architectures, has been tested in the real world and compared to other models. It is better at finding and categorising cyber risks in the IIoT environment. This model demonstrates its potential for real-world applications by achieving perfect scores across all evaluation metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. Furthermore, it outperforms simpler baseline models such as Shallow ANN, Isolated LSTM, and standalone CNN models. The comparisons provide evidence for the usefulness of our approach in detecting cyberattacks in linked device settings, where the interconnectivity of devices requires strong security measures. The results of our study show that the model has a sophisticated capacity to properly distinguish between distinct cyberattacks. This ensures that legitimate operations are correctly identified, hence preserving operational continuity and protecting against future security breaches. An asynchronous FL architecture improves this capacity by facilitating quick model flexibility, which is essential in dynamic situations that prioritise data protection and operational efficiency. This novel strategy not only protects the confidentiality of data across dispersed networks but also harnesses the combined knowledge and expertise of all participants, significantly enhancing the overall accuracy of the model's predictions. Although our model has shown outstanding performance in controlled experiments, its implementation in real-world IIoT systems may face difficulties owing to the wide range of network topologies and unforeseen attacker routes. Regular monitoring and frequent modifications to the model are essential to maintain long-term efficacy. The current focus of research is to improve the model's ability to adapt to changing threats and investigate adaptive learning methods. These methods aim to enable the model to quickly respond to new threats, making it a strong defence against advanced cyber threats. Furthermore, the enduring superiority of our model in different assault situations highlights its adaptability and wide range of uses. Nevertheless, this also raises concerns over its capacity to handle bigger and more diverse datasets in terms of scalability and speed. Future developments may prioritise resolving these scaling concerns to evaluate the model's potential to withstand a wider range of intricate assaults, therefore strengthening IIoT infrastructures against the ever-changing strategies of cyber attackers. The exceptional performance of our model confirms its crucial role in improving the cybersecurity framework of IIoT systems. It also emphasises the need for continuous improvements to ensure its relevance in a fast-evolving digital environment. The effective application of this approach offers substantial enhancements in the proactive and reactive skills of cybersecurity teams, serving as a strong defence against the evolving landscape of digital threats.

6. Conclusion

This study has focused on the crucial issue of identifying cyberattacks in the Industrial Internet of Things (IIoT) by using an advanced deep learning model integrated into an asynchronous Federated Learning (FL) framework. Our model has outstanding accuracy, precision, recall, and F1 scores, making it very beneficial for establishing strong security measures in IIoT contexts. Asynchronous federated learning (FL) is used to maintain data privacy and improve system performance. This is achieved by processing data locally at the edge and only distributing necessary model changes. This approach allows for the utilisation of the combined knowledge of a dispersed network. The combination of Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) has greatly improved the model's capacity to identify various types of cyber threats accurately, while minimising the occurrence of false positives. This improvement has resulted in reduced operational interruptions and ensured the continuity of operations. Although the paradigm has notable advantages, it encounters difficulties in terms of scalability, ongoing learning, and adjusting to emerging and changing threats. Subsequent investigations will prioritise the task of surmounting these obstacles by augmenting the model's adaptability and computational efficacy. Our research provides valuable knowledge for the advancement of cybersecurity solutions for IIoT systems. We strongly support the ongoing development and integration of Federated Learning approaches to address the complex security requirements of networked settings. Through the process of improving and perfecting these methods, our goal is to strengthen the security measures of IIoT infrastructures against the ever-changing environment of digital dangers. This will guarantee their capacity to withstand and remain dependable in the presence of progressively advanced cyberattacks.

CRediT authorship contribution statement

Syed Muhammad Salman Bukhari: Methodology, Data curation, Conceptualization. **Muhammad Hamza Zafar:** Writing – original draft, Methodology, Investigation, Formal analysis. **Mohamad Abou Houran:** Writing – review & editing, Writing – original draft, Resources, Formal analysis. **Zakria Qadir:** Writing – original draft, Validation, Resources, Formal analysis. **Syed Kumayl Raza Moosavi:** Investigation, Formal analysis. **Filippo Sanfilippo:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition.

Declaration of competing interest

All authors claim that there is not any conflict of interest regarding the above submission. The work of this submission has not been published previously. It is not under consideration for publication elsewhere. Its publication is approved by all authors and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder.

Data availability

Openly available datasets are used.

Acknowledgments

This research is supported by the Artificial Intelligence, Biomechatronics, and Collaborative Robotics research group, Top Research Centre Mechatronics (TRCM), University of Agder (UiA), Norway.

References

[1] M. Zolanvari, M.A. Teixeira, L. Gupta, K.M. Khan, R. Jain, Machine learning-based network vulnerability analysis of industrial Internet of Things, IEEE Internet Things J. 6 (4) (2019) 6822–6834.
[2] N. Moustafa, M. Keshky, E. Debiez, H. Janicke, Federated TON_IoT Windows datasets for evaluating AI-based security applications, in: 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom, IEEE, 2020, pp. 848–855.
[3] M.A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, H. Janicke, Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning, IEEE Access 10 (2022) 40281–40306.
[4] S. Prajapati, A. Singh, Cyber-attacks on internet of things (IoT) devices, attack vectors, and remedies: a position paper, IoT Cloud Comput. Soc. Good (2022) 277–295.
[5] J. Sengupta, S. Ruj, S.D. Bit, A comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoT, J. Netw. Comput. Appl. 149 (2020) 102481.
[6] I. Vaccari, G. Chiola, M. Aiello, M. Mongelli, E. Cambiaso, MQTTset, a new dataset for machine learning techniques on MQTT, Sensors 20 (22) (2020) 6578.
[7] Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, A. Shabtai, D. Breitenbacher, Y. Elovici, N-BaIoT—network-based detection of IoT botnet attacks using deep autoencoders, IEEE Pervasive Comput. 17 (3) (2018) 12–22.
[8] N. Koroniotis, N. Moustafa, E. Sitnikova, B. Turnbull, Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset, Future Gener. Comput. Syst. 100 (2019) 779–796.
[9] M.A. Ferrag, O. Friha, L. Maglaras, H. Janicke, L. Shu, Federated deep learning for cyber security in the Internet of Things: Concepts, applications, and experimental analysis, IEEE Access 9 (2021) 138509–138542.
[10] M. Mohy-eddine, A. Guezzaz, S. Benkirane, M. Azrour, An effective intrusion detection approach based on ensemble learning for IIoT edge computing, J. Comput. Virol. Hacking Tech. 19 (4) (2023) 469–481.
[11] A. Yazdinejad, B. Zolfaghari, A. Dehghantanha, H. Karimipour, G. Srivastava, R.M. Parizi, Accurate threat hunting in industrial Internet of Things edge devices, Digit. Commun. Netw. 9 (5) (2023) 1123–1130.
[12] M. Wang, B. Zhang, X. Zang, K. Wang, X. Ma, Malicious traffic classification via edge intelligence in iIoT, Mathematics 11 (18) (2023) 3951.
[13] D. Maddali, Convnext-Eesnn: An effective deep learning based malware detection in edge based IIoT, J. Intell. Fuzzy Systems (Preprint) 1–17.
[14] A. Rajak, R. Tripathi, DL-SkLSTM approach for cyber security threats detection in 5G enabled IIoT, Int. J. Inf. Technol. 16 (1) (2024) 13–20.
[15] K. Hassini, S. Khalis, O. Habibi, M. Chemmakha, M. Lazaar, An end-to-end learning approach for enhancing intrusion detection in Industrial-Internet of Things, Knowl.-Based Syst. (2024) 111785.
[16] X. Wang, J. Hu, H. Lin, S. Garg, G. Kaddoum, M.J. Piran, M.S. Hossain, QoS and privacy-aware routing for 5G-enabled industrial Internet of Things: A federated reinforcement learning approach, IEEE Trans. Ind. Inform. 18 (6) (2021) 4189–4197.
[17] X. Wang, S. Garg, H. Lin, J. Hu, G. Kaddoum, M.J. Piran, M.S. Hossain, Toward accurate anomaly detection in industrial Internet of Things using hierarchical federated learning, IEEE Internet Things J. 9 (10) (2021) 7110–7119.
[18] X. Wang, S. Garg, H. Lin, G. Kaddoum, J. Hu, M.S. Hossain, A secure data aggregation strategy in edge computing and blockchain-empowered Internet of Things, IEEE Internet Things J. 9 (16) (2020) 14237–14246.
[19] M. Azizjon, A. Jumabek, W. Kim, 1D CNN based network intrusion detection with normalization on imbalanced data, in: 2020 International Conference on Artificial Intelligence in Information and Communication, ICAIIC, IEEE, 2020, pp. 218–224.
[20] R. Dey, F.M. Salem, Gate-variants of gated recurrent unit (GRU) neural networks, in: 2017 IEEE 60th International Midwest Symposium on Circuits and Systems, MWSCAS, IEEE, 2017, pp. 1597–1600.
[21] Y. Yu, X. Si, C. Hu, J. Zhang, A review of recurrent neural networks: LSTM cells and network architectures, Neural Comput. 31 (7) (2019) 1235–1270.
[22] J. Konečný, H.B. McMahan, F.X. Yu, P. Richtárik, A.T. Suresh, D. Bacon, Federated learning: Strategies for improving communication efficiency, 2016, arXiv preprint arXiv:1610.05492.
[23] A.H. Mirza, S. Cosan, Computer network intrusion detection using sequential LSTM neural networks autoencoders, in: 2018 26th Signal Processing and Communications Applications Conference, SIU, IEEE, 2018, pp. 1–4.
[24] C. Xu, J. Shen, X. Du, F. Zhang, An intrusion detection system using a deep neural network with gated recurrent units, IEEE Access 6 (2018) 48697–48707.
[25] R. Vinayakumar, K. Soman, P. Poornachandran, Applying convolutional neural network for network intrusion detection, in: 2017 International Conference on Advances in Computing, Communications and Informatics, ICACCI, IEEE, 2017, pp. 1222–1228.
[26] J. Nguyen, K. Malik, H. Zhan, A. Yousefpour, M. Rabbat, M. Malek, D. Huba, Federated learning with buffered asynchronous aggregation, in: International Conference on Artificial Intelligence and Statistics, PMLR, 2022, pp. 3581–3607.
[27] L. Li, Y. Fan, M. Tse, K.-Y. Lin, A review of applications in federated learning, Comput. Ind. Eng. 149 (2020) 106854.
[28] W.Y.B. Lim, N.C. Luong, D.T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, D. Niyato, C. Miao, Federated learning in mobile edge networks: A comprehensive survey, IEEE Commun. Surv. Tutor. 22 (3) (2020) 2031–2063.
[29] X. Lu, Y. Liao, P. Lio, P. Hui, Privacy-preserving asynchronous federated learning mechanism for edge network computing, IEEE Access 8 (2020) 48970–48981.
[30] J. Liu, H. Xu, Y. Xu, Z. Ma, Z. Wang, C. Qian, H. Huang, Communication-efficient asynchronous federated learning in resource-constrained edge computing, Comput. Netw. 199 (2021) 108429.
[31] Z. Chen, W. Liao, K. Hua, C. Lu, W. Yu, Towards asynchronous federated learning for heterogeneous edge-powered Internet of Things, Digit. Commun. Netw. 7 (3) (2021) 317–326.
[32] M. Shahin, F.F. Chen, A. Hosseinzadeh, H. Bouzary, R. Rashidifar, A deep hybrid learning model for detection of cyber attacks in industrial IoT devices, Int. J. Adv. Manuf. Technol. 123 (5–6) (2022) 1973–1983.
[33] T. Gueye, Y. Wang, M. Rehman, R.T. Mushtaq, S. Zahoor, A novel method to detect cyber-attacks in IoT/IIoT devices on the modbus protocol using deep learning, Cluster Comput. (2023) 1–27.

[34] L. Zhiqiang, G. Mohiuddin, Z. Jiangbin, M. Asim, W. Sifei, Intrusion detection in wireless sensor network using enhanced empirical based component analysis, Future Gener. Comput. Syst. 135 (2022) 181–193.
[35] I. Tareq, B.M. Elbagoury, S. El-Regaily, E.-S.M. El-Horbaty, Analysis of TON-IoT, UNW-NB15, and Edge-IIot datasets using dl in cybersecurity for IoT, Appl. Sci. 12 (19) (2022) 9572.
[36] V. Ravi, R. Chaganti, M. Alazab, Recurrent deep learning-based feature fusion ensemble meta-classifier approach for intelligent network intrusion detection system, Comput. Electr. Eng. 102 (2022) 108156.
[37] G. de Carvalho Bertoli, L.A.P. Junior, O. Saotome, A.L. dos Santos, Generalizing intrusion detection for heterogeneous networks: A stacked-unsupervised federated learning approach, Comput. Secur. 127 (2023) 103106.
[38] M. Karthikeyan, D. Manimegalai, K. RajaGopal, Firefly algorithm based WSN-IoT security enhancement with machine learning for intrusion detection, Sci. Rep. 14 (1) (2024) 231.