Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: https://www.science-gate.com/IJAAS.html

Towards a heterogeneity-aware federated intelligence framework for proactive cyber-physical defense in the Internet of Energy

Laila Nassef*

Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 21 April 2026
Received in revised form
18 September 2026
Accepted 25 September 2026

Keywords:
Internet of Energy
Federated split learning
Intrusion detection
Cyber-physical security
Proactive defense

ABSTRACT

The Internet of Energy is a critical national infrastructure that must be protected against security attacks to prevent significant economic losses. Its communication infrastructure integrates electrical power grids with data analytics tools to automate energy generation, transmission, distribution, and consumer-related operations. However, it remains vulnerable to cyber-physical attacks and sensitive data leakage. Recent federated split learning-based intrusion detection systems help preserve data privacy but face significant challenges in large-scale heterogeneous environments. The literature highlights the need to address data, computational, and communication heterogeneity, which can reduce model performance. Therefore, this prospective paper proposes a hierarchical federated split learning framework that distributes processing across local, regional, and global levels while considering cross-layer dependencies. Graph neural networks are integrated to model network topology and provide structured feature embeddings for multi-agent deep reinforcement learning. This integration aims to understand complex infrastructure dependencies, optimize conflicting parameters, coordinate actions, and automate responses. The proposed framework aims to shift cyber-physical defense from centralized, reactive detection towards collaborative, proactive, and predictive systems. It also aims to enhance situational awareness, risk assessment, continuous adaptation, threat mitigation, and response optimization to support the early detection of complex and unknown cyber-physical threats, including coordinated attacks across critical infrastructure, thereby promoting a proactive, cognitive, and sustainable defense ecosystem.

© 2026 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Internet of Energy (IoE) is one of the critical cyber-physical ecosystems that integrates Internet of Things (IoT), communication networks, and Artificial Intelligence (AI) approaches to enable monitoring, optimization, and automated control of the smart grids. Generally, the smart grids include three main domains: customers, traditional & renewable energy generation sources, and Transmission & Distribution (T&D) infrastructures (Farhan et al., 2023). The customer domain deals with customers' consumption data, where smart meters collect power consumption data and communicate over an independent communication infrastructure called the Advanced Meter Infrastructure (AMI), which also collects data on renewable Distributed Energy Resources (DES). The other two domains continuously collect operational data from generation sources and T&D domains using the traditional Supervisory Control and Data Acquisition (SCADA) and the more advanced Wide Area Monitoring, Protection and Control (WAMPAC) systems. These two data acquisition systems continuously collect data from different types of sensors attached to all critical parts of traditional power generators, renewable energy plants, transformers, power lines, storage units, charging points, stations, and substations. SCADA collects operational data and mainly relies on Intelligent Electronic Devices (IEDs) to integrate monitoring and control functions and Programmable Logic Controllers (PLCs). SCADA collects conventional measurements such as voltage, current, and power flow to perform supervisory monitoring and control through periodic measurements and control

* Corresponding Author.
Email Address: lmohamed@kau.edu.sa
https://doi.org/10.21833/ijaas.2026.09.014
Corresponding author's ORCID profile:
https://orcid.org/0000-0001-9707-1259
2313-626X/© 2026 The Authors. Published by IASE.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

commands and communicates through the Remote Terminal Unit (RTUs) to enforce local automation and protection. WAMPAC relies heavily on more advanced sensors such as Phasor Measurement Units (PMUs) to provide high-speed synchronized measurements to support wide area monitoring, protection, and stability control. PMU data include voltage phase angles and rate of change of frequency to enable high-resolution and synchronized measurements required to enhance stability and protection conditions across T&D networks. AMI collects customers' meter data and DER data (Algarni et al., 2025).

The integration of different types of data streams from independent infrastructures is complex due to the use of different types of communication protocols. SCADA uses specialized vendor-specific communication protocols such as DNP3, IEC 60870-5-104, and Modbus to support communication with industrial control devices and substations, while WAMPAC systems use other communication protocols such as IEC 61850-90-5 and IEEE 1588. The synchronized measurements use IEEE C37.118 to communicate PMU data to local Phasor Data Concentrators (PDCs) for local processing. Data streams are forwarded from local to regional PDCs up to the utility management center based on standard communication protocols (Minh et al., 2022).

Table 1 summarizes the main differences between the two main data acquisition systems of generation and T&D domains.

Table 1: Differences between SCADA and WAMPAC data acquisition systems

Feature category	SCADA	WAMPAC
System name	Supervisory control and data acquisition	Wide area monitoring, protection, and control
Purpose	Real-time monitoring and control of grid components	Wide-area monitoring and dynamic stability analysis
Monitoring scope	Local and regional grid monitoring	Wide-area grid monitoring across multiple substations
Data source	PLC, RTUs, IEDs, sensors	PMUs
Measurement type	Conventional measurements	Time-synchronized phasor measurements
Data synchronization	Not time-synchronized	GPS time-synchronized measurements
Measurement accuracy	Moderate accuracy	High-precision synchronized measurements
Data sampling rate	Low (2–10 seconds typical)	High (30–120 samples per second)
Key operational features	Voltage magnitude, power flow, breaker status, transformer tap position	Voltage magnitude, voltage angle, frequency, rate of change of frequency
Control capabilities	Direct supervisory control of devices	Mainly monitoring and advanced protection coordination
Grid stability monitoring	Limited capability	Advanced dynamic stability monitoring
Communication protocols	IEC 60870-5-104, DNP3, Modbus	IEEE C37.118, IEC 61850-90-5, IEEE 1588
Data latency	Seconds	Milliseconds
Applications	Load monitoring, switching operations, equipment control	Oscillation detection, transient stability monitoring, blackout prevention
Physical anomaly detection	Operational anomaly detection	Operational anomaly detection

The vendor-specific proprietary protocols were originally designed for isolated utility networks and often lack built-in authentication, encryption, and integrity verification. Consequently, vendor-specific legacy protocols alone cannot provide adequate protection against modern cyber threats or maintain cybersecurity isolation. Although data acquisition systems are usually connected to control centers through secure gateways to ensure cybersecurity isolation, they cannot defend against recent attacks. The communication access points support basic security features such as encryption and authentication to protect against unauthorized access, yet they cannot defend against emerging malicious activities. Over the past decade, the energy sector has experienced a series of sophisticated security attacks targeting the physical and cyber infrastructures to compromise operations and manipulate control actions, potentially cause physical damage, and disrupt critical infrastructure, causing huge economic losses. Recently, the national platforms and energy-related services of multiple nations have been targets of attacks to steal sensitive information or cause severe destruction, with many of these attacks attributed to state-sponsored threat actors. Cyber campaigns have increased to involve network intrusion attempts, ransomware deployment, phishing operations, and hacktivist activities. These illustrate a clear escalation from isolated cyber incidents to coordinated attacks against critical national infrastructure and the growing security risks. This emphasizes the challenges facing the modern critical infrastructures of nations from organized attacks, which usually reflect the escalation of cyber warfare into political conflicts. A summary of recent documented cyberattacks against the energy sector is provided in (Abraham et al., 2025). Table 2 presents some cyber-physical attacks, indicates their impact, and features.

Currently, any disruptions to the data impact grid operation, block control actions, and directly affect grid stability (Tariq et al., 2024). The major sources of disruptions can be divided into two categories based on their target and impact. The first category includes physical attacks that target operational data and control commands, in which adversaries directly manipulate sensor measurements, state estimation inputs, or remotely manipulate control systems to disconnect substations and cause large-scale outages.

Physical attacks on parameters such as voltage magnitude, voltage angle, frequency, power flow, frequency deviation, have direct impact on state of devices causing severe consequences including physical anomalies such as sudden load spike, frequency oscillation which indicates equipment faults or abnormal operational conditions such as voltage instability, frequency deviation, line overload, and transformer overheating. The manipulation of control settings causes incorrect

control actions to directly influence physical infrastructure. Nevertheless, the interconnections between physical systems and communications networks increase the susceptibility to cyber-induced manipulation of data with physical consequences. Physical anomalies are generally detected at local and regional levels using Anomaly Detection Systems (ADSs), which detect abnormal deviations such as manipulated measurements, unusual command sequences, or unexpected system states by continuously comparing real-time operational measurements with expected patterns. The second category includes cyberattacks against communication infrastructure itself, where attackers intercept the data transmission channels without necessarily modifying physical measurements, even though they may lead to physical impacts if they disrupt critical control actions.

Table 2: Data sources of cyber-physical attacks, impact, and extracted features

Data sources	Cyber-physical attacks	Impacts	Extracted features
PMU/SCADA	False data injection attack	Manipulation of sensor measurements	Voltage magnitude deviation, phase angle difference, abnormal power flow
PMU/SCADA	Data integrity attack	Tampering with transmitted grid measurements	Voltage/frequency inconsistency
SCADA	Command injection attack	Unauthorized control commands sent to devices	Abnormal control commands, unexpected breaker status
PMU	Grid stability attack	Attacks affecting physical grid stability	Frequency deviation, voltage instability, power imbalance
Authentication and security logs	Malware/ransomware	Malicious software affecting grid control systems	Abnormal system logs, unauthorized processes
	Unauthorized access	Unauthorized login to grid control systems	Authentication failures, abnormal login patterns
Network monitoring logs	Denial-of-service	Flooding network communication channels	Packet rate, communication latency, packet loss
	Man-in-the-middle attack	Interception or modification of communications	Communication delay, authentication failures
	Replay attack	Reuse of captured legitimate packets	Timestamp mismatch, repeated packet sequences

Communication networks usually monitor network traffic, communication patterns, and protocol behavior to analyze protocol data such as headers and communication sessions to identify suspicious patterns of unusual packet loss, delays, or unauthorized access. Intrusion Detection Systems (IDSs) are generally based on collecting monitoring logs, application event logs, and security device logs to identify malicious activities. Examples of such attacks include false data injection, denial of service, man-in-the-middle, command injection, replay, data integrity, ransomware, and unauthorized access (Mallidi and Ramisetty, 2025). Currently, both ADS and IDS continuously receive their specific logs and independently provide evidence of physical anomalies or cyberattacks. Combining both systems into one integrated system enables the detection of interdependencies between operational behavior and communication patterns of malicious activities to recognize cyber-physical activities that are not directly observed by the isolated ADS and IDS systems.

The huge volume of data collected from independent infrastructures requires highly complex learning models, which have enforced a centralized approach for management, control, and cyberattack detection. Although the centralized learning models have achieved high performance, they suffer from high communication latencies and connectivity problems typically found in harsh operational environments and high-density deployment scenarios. Additionally, the centralized server could become a focal point for cyberattacks with catastrophic consequences that threaten the nation's security. The transfer of sensitive raw data raises a concern about the privacy of critical data with geolocation features that could reveal information about the locations of sensitive power plants, where attackers may use system vulnerabilities to plan direct physical destruction with serious threats to nearby populations or damage to the economy. The current cyber-physical attack incidents demonstrated the importance of protecting the critical infrastructures of nations from adversaries and emphasized the importance of analyzing the increasing capabilities of adversaries to interfere with sensitive raw data.

The increasing complexity of centralized learning models has contributed to the urgent need to distribute learning models (Nassef, 2026). Lightweight detection systems have been developed; however, they are isolated and unsuitable for large-scale and geographically distributed infrastructures. Recent advancements in machine learning have contributed to the development of Federated Learning (FL) that utilizes a central server to initialize a global Deep Neural Network (DNN) model and distribute the complete model among multiple clients to independently train the learning model using their own local data (Papadopoulos et al., 2024). After local training, clients only share the intermediate representations, known as smashed activations, with the central server to ensure that raw data remains exclusively local at the client side. The central server then aggregates the clients' models and trains the server-side model to update the global learning model and send back gradients to clients to start the next round of training. Several aggregation approaches are found; however, Federated Averaging (FedAvg), derived from Stochastic Gradient Descent (SGD), is commonly used to average the updated parameters received from all participating clients. Despite the benefits of FL, the frequent exchange of complete global models increases the communication overhead and limits scalability (Hukkeri et al., 2025).

As the learning models become deeper and wider, the increasing complexity of learning models

poses significant barriers for resource-constrained clients to host DNN models such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), and BiLSTM–Attention models (Almarshdi et al., 2024). Therefore, Split Learning (SL) has emerged to enable clients to host part of the DNN. The model is cut at a predefined split layer into two parts: a client-side model and a server-side model. Clients download the initial shallow layers of the model and train using their own raw data. During training, clients perform processing and feature extraction, forward propagation through their local layers up to the cut layer, and send the resulting smashed activations data upward to the server to train the remaining server-side model, perform forward propagation, and perform backpropagation up to the cut layer. The resulting gradients at the cut layer are then transmitted back to the clients to complete the backpropagation through their local layers. Clients may send local alerts and summaries to the server to avoid sending raw data. SL significantly reduces the computational burden on clients as deeper and more complex layers of the learning model are processed on the server. Typically, pure SL does not have aggregation across the fully independent clients; instead, the server-side model is updated centrally based on the incoming smashed activations and gradients from the sequential training of multiple clients. However, this increases the training time and reduces scalability (Hu et al., 2025). In addition, SL suffers from the struggle effects which force faster clients to wait for a slower client to complete training which contributes to more performance degradation. Furthermore, SL cannot be applied to every learning model due to the difficulty of splitting model layers and the need to synchronize model updates among clients.

Generally, the performance of both FL and SL is less than the centralized learning approaches due to variation in the quality, diversity, and quantity of local data sizes. Therefore, Federated Split Learning (FSL) has emerged as a promising solution to combine the benefits of FL parallel client training and SL model partitioning at a specific cut layer into client-side and server-side models. Each client in FSL maintains only the client-side shallow layers of the model, while the main server hosts the deeper server-side layers of the model (Ao et al., 2025; Duan et al., 2022). Training is performed collaboratively through the forward and backward propagation mechanism similar to SL; however, unlike pure SL, FSL utilizes a federated aggregation to periodically combine updates from multiple clients and reconstruct the updated global model. These features make FSL particularly well suited for hierarchical, critical data, heterogeneous, and large-scale infrastructures (Nanayakkara et al., 2024). Table 3 differentiates features of federated learning, split learning, and federated split learning.

Table 3: Main features of federated learning, split learning, and federated split learning

Feature	Federated learning (FL)	Split learning (SL)	Federated split learning (FSL)
Training architecture	Fully distributed model training	Model split training	Hybrid FL and SL model training
Model location	Entire model at each client	Model partitioned among clients	Partial model with client-side and server-side submodels
Privacy	Raw data remains exclusive to clients	Raw data remains with clients	Raw data remains at clients
Communicated data	Exchange of model gradients	Exchange of intermediate activations	Exchange of gradient activations and model updates
Communication overhead	High (communicate the entire model and the frequent gradient activations)	Low due to partial computation of shallow layers	Medium communication overheads-enhanced with optimization
Computation at clients	High computational requirements	Reduced computational requirements	Computations depend on the position of cut layer-enhanced with optimization
Scalability support	Limited scalability	Limited scalability	High scalability with hierarchical coordination
Data heterogeneity	Limited support	Limited support	Strong support for heterogeneity-aware systems using optimization
Resource efficiency	Requires powerful local clients	Suitable for resource-constrained local clients	Support resource-aware training - enhance with optimization
Training latency	High due to full model updates and frequent activation gradients	High latency due to sequential workflow and straggler effect	Optimized through integrated learning
Typical applications	Mobiles	Healthcare	Large-scale infrastructures; IoE
Multimodal cyber-physical data	Computationally expensive	Limited	Efficient for distributed hierarchical architecture
Optimization complexity	Limited	Limited	High - requires joint optimization

Despite the substantial publications on centralized IDSs, few studies have addressed FSL-based IDSs, and most of the works have assumed homogeneous environments and are still facing significant challenges in dealing with heterogeneous environments (Lin et al., 2024). In practice, large-scale infrastructures suffer from three types of heterogeneities. First, data heterogeneity results from different data acquisition systems with different data sizes, feature distributions, class distributions, and attack types that vary significantly across different regions and time. The data heterogeneity is commonly known as non-Independent and non-Identically Distributed (non-IID), which violates the common assumptions of homogeneous data typically assumed in centralized learning models (Jimenez et al., 2024). Second, computational heterogeneity resulting from the different computational capabilities of participating systems makes them unable to host complex learning models. These differences create the straggler problem, where resource-constrained clients take longer training times, slow down the overall training process, and cause synchronization problems during model aggregation (Zhu et al., 2024). Third, communication heterogeneity which

results from variations in network technologies, protocols, and harsh operational environments. Furthermore, the coexistence of different proprietary protocols with standard protocols generate different types of traffic logs and session records contribute to difficulty to process multimodal data sources. These three types of heterogeneity impact model divergence, cause overfitting, and reduce generalization, which represent a severe challenge to the development and implementation of distributed learning models (Sun et al., 2025).

The heterogeneity problem is currently solved by traditional approaches that usually tackle a specific type of heterogeneity. Consequently, this research prospectively proposes a federated intelligence cyber-physical proactive security defense framework to support decentralized collaborative learning, tackle the gap in integrating ADSs and isolated IDSs, integrate multiple systems of independent infrastructures, and dynamically adapt multiple key design parameters to solve the heterogeneity problem (Chen et al., 2025). To the best of our knowledge, few of the existing works have proposed privacy-aware detection systems tailored specifically for large-scale and heterogeneous critical infrastructures. The core of the proposed framework is the FSL to distribute learning models and contribute towards distributed threat intelligence (Lin et al., 2026). The integration of multiple data sources requires a modality-specific encoder and attention-based fusion to classify events as either normal operation, cyberattack, physical fault, cyber/physical attack, sensor anomaly, or learning system attack. Additionally, the framework prospectively integrates two complementary learning approaches, namely Graphical Neural Network (GNN) to model topology and provide structured feature embedding to help the Multi-Agent (MA) Deep Reinforcement Learning (DRL) (Ai et al., 2025) to understand the complex dependencies of the infrastructure to adaptively optimize multiple conflicting parameters (Adam et al., 2025), and to ensure effective orchestration and response automation (Huang et al., 2025).

The aim of this research paper is to provide a comprehensive analysis of the current research works on FSL and the heterogeneity problem, specify the open challenges associated with real-time implementation, and propose future research directions. The framework serves as a strong starting point to pave the way for the next generation of cyber-physical security defense systems by leveraging a hierarchical FSL approach to distribute learning models over local, regional, and global layers. Motivated by this, the objectives of this research are summarized as follows: present the architectural layers for IoE, highlight the current limitations, and propose emerging research directions. While the focus is on the FSL and heterogeneity problem, the framework introduces the required functions to support future integration based on two complementary learning approaches. Based on the prospective nature of the research, contributions can be formulated as follows: design a heterogeneity-aware hierarchical FSL framework, analyze functions of GNN and DRL to investigate feasibility of developing integrated intelligence (learning-driven) cyber-physical proactive defense system, which has been observed as a critical research gap.

The remainder of this paper is organized as follows. Section 2 presents the literature review, while Section 3 provides the research methodology and analyzes the proposed hierarchical architecture with multiple computing layers. Section 4 discusses the research results, features of the proposed framework, limitations of current work, research gaps, future research directions, and presents the dataflow of the proposed framework. Section 5 presents the discussion. Finally, Section 6 concludes the paper and recommends future work.

2. Literature review

The current FSL-based IDSs usually assume a single centralized server, which creates communication bottlenecks, increases communication latency, and limits scalability of the decentralized architecture. Recent works have separated split learning computation from federated learning coordination and utilized two servers to enhance scalability. The first is the main server, which initializes the global model, splits the model into a server-side model and client-side sub-models, receives smashed activation data, performs feed-forward, computes loss, and redistributes the client-side shallow layer's models. The second is the federated server, which aggregates model parameters, manages training rounds, and synchronization. Additionally, most of the current works have assumed homogeneous data, computation resources, and a stable communication environment to simplify the training process. However, this does not accurately reflect the real world computing and the communication capabilities that vary significantly with time and location and greatly degrade model's performance. Furthermore, current works generally utilize the classical FedAvg, which implicitly assumes homogeneous clients in terms of computational capability, data distribution, and number of local training iterations (Samikwa et al., 2024). Additionally, FedAvg assumes that all clients are synchronized with the main server, use the same client-side model, and contribute equally to the training process; under these assumptions, FSL generally selects a random subset of clients to participate in each training round.

In practice, all these assumptions rarely hold for heterogeneous environments, resulting in weight divergence, slow convergence, and client drift, which degrade performance and raise issues of fairness among clients (Timofte et al., 2025). Moreover, most of the current works have proposed a fixed position of the cut layer to keep the same model partition

across all clients and training rounds, aiming to reduce coordination overhead and communication complexity (Dakhia and Merenda, 2025). Yet, in practice, fixed position of cut layer does not adapt to client resources and leads to inefficient workload distribution leading to struggles bias. Therefore, adaptive client selection based on their contribution towards enhancement of model performance, adaptive cut layer, and adaptive aggregation strategy to assign different weights to clients is demanded (Lin et al., 2025a).

Table 4 summarizes the key optimization categories, current approaches, existing limitations, gaps, and proposed research directions.

Table 4: A summary of limitations, current approaches, research gap, and proposed research directions

Category	Existing approaches	Limitations of current approaches	Research gap	Proposed research directions
Distributed learning architecture	FL-SL-FSL	Mostly based on either FL or SL independently; few are based on FSL	Hierarchical FSL	Hierarchical FSL
Client participation	Random or static client selection	Inefficient training and high communication latency	Adaptive client selection	MA-DRL adaptive client selection
Model partitioning	Fixed cut layer	Fixed unified sub-models	Dynamic cut layer optimization	MA-DRL adaptive cut layer positioning
Aggregation strategy	Simple averaging (FedAvg)	Sensitivity to data heterogeneity; Poor convergence and unfair to good contributing clients	Adaptive aggregation strategies	MA-DRL adaptive aggregation strategies
Hyper-parameter optimization	Static initial hyper-parameters optimization	Poor model performance	Dynamic hyper-parameter optimization	GNN-MA-DRL optimized learning
Data integration	Mostly network traffic and weak feature fusion	Physical anomalies are often ignored	Correlations of cyber-physical-control actions data	Multimodal fusion
Attack detection	Independent anomaly and intrusion detection systems	Limited detection of cyber-induced physical anomalies	Intelligent integration of cyber-physical manipulation	Multimodal cyber-physical integrated system
Network topology modeling	Simple modeling with no cross-layer interactions	Ignore relationships among layers	Topology-aware learning	GNN modelling
System heterogeneity	Mostly considers homogeneous environments	Limited support for heterogeneity	Heterogeneity-aware learning	MA-DRL optimization
Optimization complexity	Rule-based single-objective optimization	Limited optimization of multiple conflicting parameters	Data-driven-based optimization	GNN + MA-DRL
Explainability	Partial support	Limited human-understandable reasoning for decisions	Lack of validation of policy-driven decisions	Explainable (XAI) - human in the loop (HITL) + digital twin

The majority of the current research works have typically addressed key design parameters independently and ignored the strong coupling among independent decisions (Gao et al., 2025), in which decisions to enhance one parameter influence the performance of others (Hao et al., 2024). The interrelation among the independent parameters in federated learning has not been addressed, which impacts the performance of collaborative learning. The neglect of the position of the cut layer affects both communication overhead and client computation load, while neglecting the aggregation strategy influences model accuracy, communication overhead, training latency, convergence behavior, and robustness to heterogeneous data distributions (Lin et al., 2025b). The current works have failed to discover the extremely large search space created by cross-layer interacting variables, as the simultaneous optimization of multiple conflicting parameters is extremely complex and computationally difficult. Table 5 summarizes the literature review and highlights the current optimization approaches.

To the best of our knowledge, few works have tackled the joint optimization of multiple conflicting parameters in large-scale, dynamic, and heterogeneous environments, yet the integration with multiple learning approaches remains largely unexplored. Therefore, learning-based optimization approaches that dynamically observe, learn, correlate, and optimize actions are urgently required to minimize privacy leakage, physical and cyber risks, training latency, communication and computation costs, while maximizing detection accuracy, trust, risk reduction, and service continuity. The joint optimization problem with multiple decision variables that are highly dependent on each other makes the problem multi-objective, mixed discrete, and dynamic, which demands integrating multiple learning approaches. Consequently, this prospective research aims to highlight the gap in the current works and propose an integrated intelligence cyber-physical proactive defense framework to pave the way towards development of the unified intelligence system that enable simultaneous adaptations of different interdependent architectural layers and support continues learning to overcome the major limitation of current centralized, reactive, and isolated systems, which have been realized as a critical research gap.

3. Methodology

The IoE is generally described using six architectural layers, with the first layer containing the physical raw data layer, the second layer representing the data acquisition layer, and the third layer representing the communication layer (Xu et al., 2024). The analysis of current cyber-physical security approaches and the identified research gaps in the literature have revealed strong requirements to combine evidence from multiple domains to optimize event understanding and differentiation. Traditionally, multimodal data has been handled by developing independent homogeneous learning systems at a centralized location, ignoring the real-world interdependencies and treating all types independently, in which learning models fail to capture the relationships among different forms.

Therefore, the proposed framework considers the first three layers of IoE as multimodal data sources to enable the capture of contextual information to support detection of sophisticated attacks and unknown cyber threats, where traditional approaches struggle to capture such dependencies.

Table 5: Summary of literature review

Reference	Learning paradigm	Domain	Clients	Proposed algorithm	Challenges	Optimization approach	Adaptation	Aggregation	Model	Dataset/benchmark	Performance metrics
Liu et al. (2023)	SFL	Unmanned aerial vehicle	6G-wireless base stations	Hybrid SFL	System and data heterogeneity	Multi-armed bandit best channel and best 2-norm	Client selection-cut layer-channel quality	Proposed	DNN	Image datasets	Accuracy-training latency
Samikwa et al. (2024)	SFL	IoT	Heterogeneous	Heterogeneous-SFL	Latency-energy	Similarity-based clustering	Data and system heterogeneity	FedAvg	DNN	Two real testbeds + two computer vision	Accuracy-training latency
Xu et al. (2024)	Multi-level SFL	IoT-smart cities	Large scale	Framework	Latency	Clustering	Non-IID	FedAvg	DNN	Two computer vision	Accuracy-training latency
Lin et al. (2024)	SFL	IoT	Resource-constrained	Efficient parallel SFL	Channel variations	Greedy + block coordinate descent; exhaustive search + mixed-integer non-linear programming	Channel allocation-power-cut layer	Last layer activations gradient	DNN	Two image and object datasets	Model accuracy-training latency-resource allocation
Huang et al. (2025)	SFL	IoT	Resource-constrained	SplitFed	Latency and privacy leakage	Block coordinate descent + mixed-integer non-linear programming	Resource allocation	FedAvg + others	DNN	Two computer vision datasets	Training latency-waiting time-model accuracy-data leakage risk-downlink bandwidth
Zhu et al. (2024)	SFL	IoT	System heterogeneity	ESFL	Straggler problem-heterogeneous resource constraints	Mixed-integer non-linear programming	Cut layer	FedAvg	DNN	One image dataset	Training time and model accuracy
Sun et al. (2025)	SFL	ML	Heterogeneous resource constrained	Heterogeneous-SFL	Latency and resources allocation	Sample average approximation, genetic algorithm, Lagrangian relaxation, branch-and-bound	Cut layer	FedAvg	DNN	Three datasets of speech-object-image	Training time
Gao et al. (2025)	SFL	General	System heterogeneity	PipeSFL	Scheduling, cut layer, training time	Priority scheduling	System heterogeneity	FedAvg	DNN	Two computer vision datasets	Training time
Lin et al. (2025a)	SFL	General	Convergence analysis	AdaptSFL	Communication-computational latency	Iterative descending	Auxiliary network	Adaptive	DNN	Computer vision datasets	Accuracy-time
Dakhia and Merenda (2025)	FL	IoT	System heterogeneity	Reward-based payoff strategy	Client selection	Nash-based game theory	Data heterogeneity	FedAvg	CNN	Four image datasets	Accuracy-energy- privacy
Timofte et al. (2025)	FL	IoT-healthcare	Data and systems	Privacy-preserving	Explainable secure aggregation	SecFL-IoT	Cryptographic aggregation	FedAvg + FedProx	CNN-LSTM	Two IDS datasets	Privacy-communication overheads

This will enforce a shift from outdated datasets that do not fully capture the complex cyber-physical characteristics towards more realistic physical and cyber datasets to fully capture the complex cyber-physical characteristics required to represent real-world monitoring, control, and communications data generated from different protocols and technologies. This necessitates the use of modality-specific encoders, such as a time-series encoder for feature extraction of operational logs, a network monitoring encoder for cyber logs, and a context encoder for control actions. They are integrated with attention-based functions to classify events as normal operation, physical faults, cyberattacks, cyber-physical attacks, or learning system attacks.

The hierarchical FSL, distributed over local, regional, and global layers with cross-layer dependencies, requires capturing interactions among different layers and modelling of computational and communication resources. The accurate modeling of hierarchical structural relationships is often not modeled explicitly (Liu et al., 2023). Therefore, the GNN is proposed to model heterogeneous interconnection of independent infrastructures, where assets are represented as nodes with communication dependency, and learning relationships as edges to enable prediction of compromised nodes, vulnerable links, next hop targets, and possible propagation paths across the multiple domains. Each independent infrastructure hosts its own GNN agent to observe its local graph, identify anomalies and abnormal traffic patterns to indicate attacks, compute trust, risk scores, and to support situational awareness to provide information about affected resources, physical impact, domain impact, severity, and response priority. GNN provides structured feature embeddings to help regional DRL agents to understand the complex dependencies of the infrastructure, and optimize multiple complex decisions. DRL selects defense actions such as monitor, block, isolate, escalate, recover, and seek human intervention. Actions are selected to balance risk reduction, service continuity, containment time, and recovery time.

The proposed methodology starts with the client's local processing layer that collects operational, physical, and cyber evidence for

multimodal fusion and performs local data preprocessing, including filtering, cleaning, feature extraction, and normalization to improve data quality and consistency. The client-side model is trained using private data to feed forward propagation to detect and classify evidence of deviations that may reflect faults, physical attacks, local cyberattacks, or local cyber-physical attacks. Clients send smashed data to the regional server and receive the backpropagation gradients to update their own weights. The local layer enforces local decisions to ensure low-latency response to physical anomalies and local intrusion before severe impacts, and also generates early warnings against threats such as unauthorized access, malware, and data manipulation. The local monitoring agent schedules resources, shares summaries and confidence values of ADS/IDS evidence with regional agents for further analysis, and produces graph-derived representations to share the local status of computing and communication resources data with the regional layer, along with the smashed activations data.

The regional layer receives clients' smashed activations, completes the forward path, computes the regional loss function, and performs the backward propagation path, then triggers the backpropagation path. Then the server aggregates the updates received from clients to refine the regional FSL model, and sends back the updated model gradients to local clients. This layer enables local detection of cyber-physical attack classes within the geographic region, analyzing patterns across multiple clients to identify coordinated anomalies and cyberattacks with physical impact, which is particularly important for critical infrastructures and is currently not detectable with isolated systems. The regional layer identifies threshold violations and instability signs that indicate suspicious commands, protocol misuse, replay, and abnormal session behavior to enable immediate blocking and safe control actions.

Regional agents may cooperate with neighboring regional agents to identify vulnerable regions, estimate attack propagation probabilities, and prioritize critical resources. This coordination facilitates identification of coordinated regional campaigns and enables generation of region-wide cyber-physical risk assessment reports to estimate risk and predict attack paths. Regional DRL agents utilize GNN representations to adapt distributed learning and cyber-physical defense decisions and may coordinate with other regional agents to ensure that their own defensive reactions do not impact other regions or isolated infrastructures. The regional agents share only the model gradients learned from attacks and communicate with the global layer to update the global model and adapt to evolving threats. Additionally, the continuous interactions across GNN and DRL agents support situational awareness, adaptive decision-making, threat mitigation strategies, recovery planning, and service continuity management.

The regional layer acts as a bridge between the client's intelligence and the global layer by escalating statistics and other selected events to the global server to perform policy tuning, including global weights, thresholds, classifiers' parameters, support personalization, and robust aggregation. The overall global layer aims to simultaneously balance training convergence, heterogeneity tradeoffs, and model's performance, while maintaining the availability, reliability, and operational stability of critical cyber-physical infrastructure. Local and regional layers may not be able to support continuous inference and coordination, and thus heavy functions are assigned to the global layer, which is added as the seventh or defense layer. Table 6 compares the functions of IoE layers, with architectures of current centralized IDS, proposed framework, indicating current challenges, and the proposed future research directions to enable the shift to next generation security systems.

4. Results

The framework distributes intelligence over local, regional, and global data analytics layers to process high-dimensional, complex data and execute defensive actions closer to data sources. Domain-adaptive agents are associated with each infrastructure to solve heterogeneity problems. The multimodal graph transformers are integrated with the GNN agent to perform message passing to the global layer to learn topology and state representation of a single region, while MA-DRL enables spatio-temporal fusion to capture latent correlations for situational awareness within the specific region. The integration of multimodal, hierarchical FSL, GNN, and MA-DRL enables dynamic adaptation to clients' trust and risk scores to exclude unreliable or malicious clients, reduce their aggregation weights, and prevent clients from participating in the next training round to reduce risks. The overall hierarchical FSL-based GNN-MA-DRL framework enforces collaborative learning for continuous interaction among different agents to support global threat classification, trust and risk scores, attack path predictions, topology embedding, and resource status. The optimized responses help to decide the most appropriate defense strategy to limit attack propagation and identify critical resources and communication links. The mitigation recovery policies determine the optimal sequence to restore compromised services and ensure service continuity to maintain system operation through reconfiguration and adaptive resource allocation.

The global layer updates the global model, captures cyber-physical attack patterns, and provides global orchestrations. This layer is divided into three sublayers. The first global sublayer is the global GNN intelligence layer, where resources are represented as nodes and edges to predict the compromised nodes, vulnerable links, next-hop targets, and possible propagation paths across the multiple independent domains. The GNN representations of communication networks, device

interconnections, and dependency relationships enable identification of critical nodes and potential attack propagation paths across regions. The situational awareness provides information about severity, confidence, affected nodes and links, physical consequence, and response priority.

Table 6: Mapping of layers of IoE with centralized IDS, proposed framework, current challenges, and future research directions

IoE layers	IoE hierarchical architecture	Current centralized (IDS) systems	Proposed decentralized FSL-GNN-MA-DRL-based FIDF	Current challenges	Research directions
Layer 1: Physical energy infrastructure	Field sensors	Collect raw data	Collect raw data	Local multimodal cyber-physical-operational measurement data	System-wide observations
	PMUs	Receive control actions	Receive feedback control actions		
	PLC, RTU, IED				
	Actuators				
	Smart meters				
Layer 2: Data acquisition layer	SCADA/WAMPAC operational data	Vendor-specific traffic logs	Collect evidence of operational logs	Data, system, and communication heterogeneity	Multimodal data encoders
	Vendor-specific protocol logs		Distribute control actions back		
	AMI head end				
	Send data to local PDC				
Layer 3: Communication layer	Operational data	Standards traffic logs	Operational data streams	Heterogeneous vendor-specific and standard protocols	Communication efficiency
	Control actions		Feedback actions	Harsh operational environments	
	Standard network protocols		Traffic logs		
			System logs		
			User activities		
			Security events		
Layer 4: Local processing layer	Local preprocessing	Aggregate local data streams	Client-side sub-model	Setting of hyperparameters	Multimodal data fusion
	Lightweight ADS		Local preprocessing	Resource allocations	Adaptive cut layer
	Control actions		Client-side training	Cut layer	Adaptive client scheduling
	Data to regional PDC		Smashed activation data	Resource status	Adaptive resource allocation
			Local ADS/IDS reports		Local monitoring agent
			Local physical and cyber indicators		
			Local alerts		
			Local summaries		
Layer 5: Regional processing layer	Data from multiple local PDCs	Regional data streams	Receive a copy of the FSL model	Optimization of multiple conflicting parameters	Regional agents
	ADS processing		List of participating clients	Regional topology-aware modelling	Multimodal data fusion
	Raw data to global server		Other optimization parameters	Regional coordination	Regional GNN graph construction
	Control actions		Split model	Regional adaptation of resources	Regional decision-making and optimization
			Train client-side model	Attack path prediction	Partial adaptive model aggregation
			Smashed activations		Client selection
			Partial aggregation		Resource scheduling
			Train server-side model		
			Regional ADS and IDS indicators		
			Regional summaries		
Layer 6: Global intelligence layer	Data streams	Traffic logs	Global model training	Topology-aware modelling	Global agents
	Feed diverse applications	Global IDS	Model update	Regional coordination	FSL aggregation
	Independent global ADS	Attack detection and classification	Attack detection and classification	Cross-layer optimization	GNN topology awareness
	Isolated grid-wide IDS	Global actions	Topology awareness		Situational awareness
			Situational awareness		Global optimization
			Policy planning		Risk assessment
			Cross-region optimization		Policy planning
			Recovery actions		Continuous learning
			Proactive defense		
Layer 7: Defense layer (proposed)	None	None	Response optimization	Complexity of integrating multiple AI approaches	Multi-stage implementations
			Mitigation strategies	Agent's interactions	
			Service continuity		
			Global orchestration		

The GNN representations are passed on to the DRL response optimization layer to select the most appropriate adaptive defense actions, including inspection, exclusion, blocking, isolation, recovery, escalation, or human approval, while balancing risk reduction, service continuity, containment time,

recovery time, safety, disruption cost, and uncertainty.

The dynamic optimization of GNN situational awareness and operational decisions is optimized by the second sublayer, MA-DRL, to perform global decision-making based on reports received from local and regional agents. The hierarchical MA-DRL agents optimize responses and prediction of sequential decisions through continuous interaction across local and regional agents (Yaseen et al., 2026). The joint optimization is performed at the global coordination layer to balance learning performance, privacy preservation, operational safety, and resource efficiency. The third sublayer is the proposed defense layer that receives summary reports from local and regional agents and sends them feedback about GNN and FSL model updates, MA-DRL policy refinement, DRL reward signals, and resilient status. The integrated intelligence overall system generates reports about global attacks, risk-aware responses, and adaptation reports for continuous refreshment of the mitigation policies to enhance learning and support coordination of response actions. It also updates defense policy, plan mitigation and recovery strategies for service continuity to maintain global resilient defense. The data flow of the proposed framework and layer interactions are illustrated in Fig. 1.

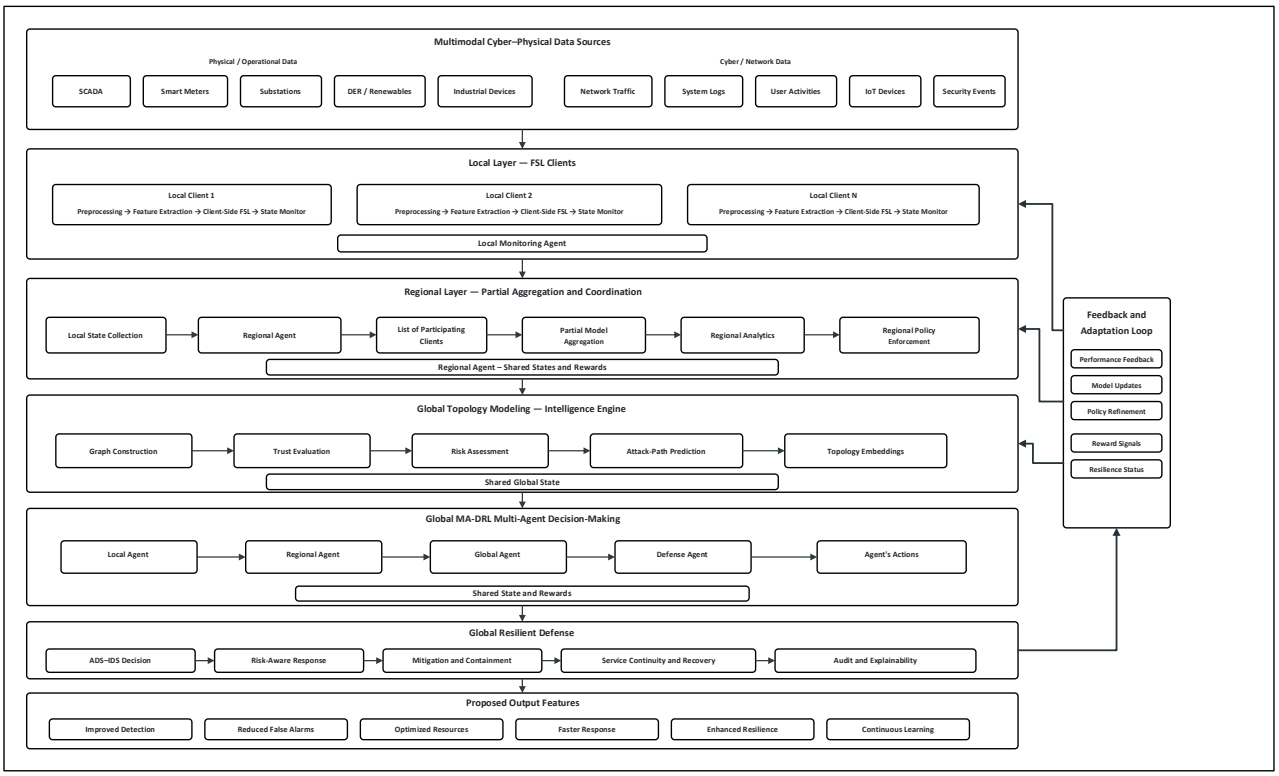


Fig. 1: Dataflow of the proposed framework

5. Discussion

The framework is proposed to pave the way towards the design and implementation of the overall system to enhance performance, reduce false alarms, optimize resources for a faster response, and enable continuous learning. Although the proposed framework has not yet been experimentally validated, the expected improvements will be systematically investigated during the implementation and evaluation phases. The implementation of the proposed system is challenging but feasible through multi-stage development. Local and regional agents with distributed intelligence-based GNN agents to capture the cross-layer interdependencies between architectural layers can be implemented using Graph Sample and Aggregate (Graph-SAGE) to generate topology-aware trust, risk, and attack-path predictions. Regional optimization can be viable using MA-Proximal Policy Optimization (PPO)-DRL to select clients, aggregation weights, schedule resources, and assign the position of the cut layer. The multi-agent coordination and response optimization for simultaneous optimization of key design parameters and adaptation policies are achievable using Centralized Training with Decentralized Execution (CTDE). Different approaches for more complex functions will be developed gradually after validating the initial integration of the learning-driven approaches. The experimental setup is expected to use multiple types of physical and cyber datasets such as MSU/ORNL power grid dataset, EPRI, IEEE-PES, IEC 60870-5-104, CIC-IDS2017, UNSW-NB15, and CICIoT2023 datasets, among others.

These datasets have different physical anomalies and cyber intrusions to support the multimodal system. However, in practice, the synchronization of physical logs with communication logs is difficult due to different collection rates from different data sources; therefore, cross-modal multi-head attention

layers and time interval encoders are practical solutions.

The experimental implementation, benchmark comparison, and quantitative evaluation are expected to be validated across multiple groups. For fair comparison and consistent performance evaluation, metrics are expected to be normalized and categorized into different categories of detection, learning, cyber intelligence, and response optimization. Metrics for cyber-physical detection include F1-score, ROC-AUC, false positive rate, and detection rate. Metrics for FSL include communication cost, training time, convergence speed, aggregation overhead, and scalability under data, computing, and communication heterogeneities. Metrics for GNN include accuracy of trust estimation, risk assessment, and attack path prediction, which reflect the ability to model dependencies and threat propagation. Metrics for DRL include cumulative reward, response reduction, containment time, recovery time, and response success rate, among others. On the other hand, the overall performance of the potentially implemented system must be evaluated based on detection effectiveness, trust and risk assessment, attack path prediction, response optimization, as well as the computation and communication efficiencies. Moreover, a cyber-resilience index is also expected to be formulated and evaluated to assess the efficiency of the overall performance in detecting, predicting, responding, and recovering from threats to ensure a unified measure of the overall system. Finally, the proposed system must be verified to ensure that it meets the required technical standards, including metrics such as convergence rate, reward stability, and robustness testing, to ensure interoperability and reliability of security policies.

6. Conclusion

The limitations of the current centralized and independent anomaly and intrusion detection systems are prospectively analyzed, and the federated intelligence proactive defense framework is designed to pave the way for further research to implement and validate the system. The framework supports local, regional, and global learning, preserving privacy, model topology, and adaptively optimizing multiple conflicting parameters to mitigate heterogeneity problems. The core privacy-aware multimodal FSL jointly trains the learning model to detect and classify attacks. GNN enables trust computations, risk assessment, and attack path prediction for situational awareness. MA-DRL supports adaptive selection of participating clients, specifies the optimal location of the cut layer, configures hyperparameters, and adapts aggregation strategies in each training round, subject to training latency, communication cost, and computation cost, while maximizing model convergence, model detection performance, and fairness among heterogeneous clients. The proposed integrated intelligence framework has addressed the specific functions of each learning layer and the cross-layer interactions to enable continuous adaptation, threat mitigation, optimization of responses to support early detection of cyber-physical threats at local clients, maintain regional coordination, and ensure global situational awareness across the entire critical infrastructure.

The immediate future work is to implement the initial version of the framework, where the overall performance is expected to enhance convergence, stability, and improve the model's ability to detect complex and unknown cyber threats, including rare but dangerous zero-day attacks and coordinated campaigns that span multiple regions. The FSL multimodal GNN graph-based MA-DRL framework promotes innovation for future implementation and deployment in critical infrastructures, including smart grids, healthcare, transportation, telecommunications, industrial systems, and smart cities. The learning-driven optimization approach is expected to contribute to continued research to pave the way towards next-generation proactive defense against the newly evolved cyber-physical attacks and campaigns. The ultimate goal of the proposed framework is to support a better understanding of the behavior of collaborative systems to build trust for utilities to implement and deploy the proposed system. Future work aims to explain knowledge to support operators with visual representation and human approval to validate alerts, provide feedback, and intervene in critical situations for continuous refinement of security policies. Therefore, future extensions are expected to include Explainable Artificial Intelligence (XAI) (Fatema et al., 2025), with Human-in-the-Loop (HITL) to enhance transparency and interpretability of decisions. Additionally, adding a digital twin layer for visual representations, validation, and zero-latency optimization is expected to help operators make decisions under pressure and visualize the system's state in real time.

Finally, the highly complex interconnected learning architecture is vulnerable to adversarial attacks such as GNN topology poisoning and inference attacks, while MA-DRL is vulnerable to state and action perturbations, and the integrated system is vulnerable to cascading failure, feature state manipulation, and adversarial communications. The neglect of adversarial attacks in the training phase is the most important challenge to the growth and deployment of learning-driven models. Careful investigation is required before implementation to include robustness against adversarial attacks in the design phase to maintain the system's integrity even under adversarial pressure. Secure aggregation protocols combined with encryption and differential privacy are urgently required. Therefore, incorporating adversarial robustness and secure aggregation will receive more attention to enhance the reliability and trustworthiness of the proposed system, to maintain sustainable defense against complex and previously unseen zero-day threats and

the detection of destructive cyber-physical campaigns.

List of abbreviations

ADS	Anomaly detection system
AI	Artificial intelligence
AIoT	Artificial intelligence of things
AMI	Advanced meter infrastructure
BiLSTM	Bidirectional long short-term memory
CIC-IDS	Canadian Institute for Cybersecurity intrusion detection system
CICIoT2023	Canadian Institute for Cybersecurity Internet of Things 2023 dataset
CNN	Convolutional neural network
CTDE	Centralized training with decentralized execution
DER	Distributed energy resources
DES	Distributed energy resources
DFL	Dynamic federated learning
DNN	Deep neural network
DNP	Distributed network protocol
DRL	Deep reinforcement learning
ESFL	Efficient split federated learning
EPRI	Electric Power Research Institute
FedAvg	Federated averaging
FedProx	Federated proximal
FedsLLM	Federated learning for large language models
FIDF	Federated intelligence defense framework
FL	Federated learning
FSL	Federated split learning
GNN	Graph neural network
GPS	Global positioning system
Graph-SAGE	Graph sample and aggregate
HITL	Human in the loop
IDS	Intrusion detection system
IEC	International Electrotechnical Commission
IEEE-PES	IEEE Power & Energy Society
IED	Intelligent electronic device
IoE	Internet of Energy
IoT	Internet of Things
LLM	Large language model
LSTM	Long short-term memory
MA	Multi-agent
MA-DRL	Multi-agent deep reinforcement learning
MA-PPO	Multi-agent proximal policy optimization
Modbus	Industrial communication protocol
MSU/ORNL	Mississippi State University / Oak Ridge National Laboratory
non-IID	Non-independent and identically distributed
PDC	Phasor data concentrator
PLC	Programmable logic controller
PMU	Phasor measurement unit
PPO	Proximal policy optimization
ROC-AUC	Area under the receiver operating characteristic curve
RTU	Remote terminal unit
SCADA	Supervisory control and data acquisition
SecFL-IoT	Secure federated learning for the Internet of Things
SGD	Stochastic gradient descent
SHAP	SHapley additive explanations
SFL	Split federated learning
SL	Split learning
T&D	Transmission and distribution
UNSW-NB15	University of New South Wales network-based 2015 dataset
WAMPAC	Wide-area monitoring, protection, and control
XAI	Explainable artificial intelligence

Compliance with ethical standards

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Abraham D, Houmb SH, and Erdodi L (2025). Cyber-attacks on energy infrastructure—A literature overview and perspectives on the current situation. Applied Sciences, 15(17): 9233. https://doi.org/10.3390/app15179233

Adam M, Albaseer A, Baroudi U, and Abdallah M (2025). Survey of multimodal federated learning: Exploring data integration, challenges, and future directions. IEEE Open Journal of the Communications Society, 6: 2510-2538. https://doi.org/10.1109/OJCOMS.2025.3554537

Ai Y, Li H, Ruan H, and Liu H (2025). A heterogeneous graph neural network assisted multi-agent reinforcement learning for parallel service function chain deployment. Journal of King Saud University Computer and Information Sciences, 37: 236. https://doi.org/10.1007/s44443-025-00258-1

Algarni F, Nassef L, and Alowidi N (2025). Deep learning model with hybrid resampling technique for electricity theft detection in smart grid. Journal of King Abdulaziz University: Computing and Information Technology Sciences, 14(1): 95-105. https://doi.org/10.64064/1658-6336.1009

Almarshdi R, Fadel EA, Alowidi N, and Nassef L (2024). Distributed federated split learning based intrusion detection system. Intelligent Automation & Soft Computing, 39(5): 949-983. https://doi.org/10.32604/iasc.2024.056792

Ao H, Tian H, and Ni W (2025). Federated split learning for edge intelligence in resource-constrained wireless networks. IEEE Transactions on Consumer Electronics, 71(2): 4451-4463. https://doi.org/10.1109/TCE.2024.3464731

Chen X, Li J, Fan D, and Chakrabarti C (2025). HeteroSFL: Split federated learning with heterogeneous clients and non-IID data. IEEE Internet of Things Journal, 12(15): 30460-30474. https://doi.org/10.1109/JIOT.2025.3572393

Dakhia Z and Merenda M (2025). Client selection in federated learning on resource-constrained devices: A game theory approach. Applied Sciences, 15(13): 7556. https://doi.org/10.3390/app15137556

Duan Q, Hu S, Deng R, and Lu Z (2022). Combined federated and split learning in edge computing for ubiquitous intelligence in Internet of Things: State-of-the-art and future directions. Sensors, 22(16): 5983. https://doi.org/10.3390/s22165983 **PMid:36015747 PMCid:PMC9414384**

Farhan M, Reza TN, Badal FR, Islam MR, Muyeen SM, Tasneem Z, Hasan MM, Ali MF, Ahamed MH, Abhi S, Islam MM, Sarker SK, Das SK, and Das P (2023). Towards next generation Internet of Energy system: Framework and trends. Energy and AI, 14: 100306. https://doi.org/10.1016/j.egyai.2023.100306

Fatema K, Dey SK, Anannya M, Khan RT, Rashid MM, Su C, and Mazumder R (2025). Federated XAI IDS: An explainable and safeguarding privacy approach to detect intrusion combining federated learning and SHAP. Future Internet, 17(6): 234. https://doi.org/10.3390/fi17060234

Gao Y, Hu B, Mashhadi MB, Wang W, and Bennis M (2025). PipeSFL: A fine-grained parallelization framework for split federated learning on heterogeneous clients. IEEE

Transactions on Mobile Computing, 24(3): 1774-1791. https://doi.org/10.1109/TMC.2024.3489642

Hao Y, Zhao C, Zhang Y, Cao Y, and Li Z (2024). Constrained multi-objective optimization problems: Methodologies, algorithms and applications. Knowledge-Based Systems, 299: 111998. https://doi.org/10.1016/j.knosys.2024.111998

Hu Z, Zhou T, Wu B, Chen C, and Wang Y (2025). A review and experimental evaluation on split learning. Future Internet, 17(2): 87. https://doi.org/10.3390/fi17020087

Huang Q, Luo Y, Liu Z, Xia J, Chang M, and Li J (2025). A hierarchical reinforcement learning framework for multi-agent cooperative maneuver interception in dynamic environments. Journal of Marine Science and Engineering, 13(7): 1271. https://doi.org/10.3390/jmse13071271

Hukkeri GS, Goudar RH, Dhananjaya GM, Rathod VN, and Ankalaki S (2025). A comprehensive survey on split-fed learning: Methods, innovations, and future directions. IEEE Access, 13: 46312-46333. https://doi.org/10.1109/ACCESS.2025.3547641

Jimenez GDM, Solans D, Heikkilä M, Vitaletti A, Kourtellis N, Anagnostopoulos A, and Chatzigiannakis I (2024). Non-IID data in federated learning: A systematic review with taxonomy, metrics, methods, frameworks and future directions. ArXiv Preprint ArXiv:2411.12377. https://doi.org/10.48550/arXiv.2411.12377

Lin Z, Chen Z, Chen X, Ni W, and Gao Y (2026). HASFL: Heterogeneity-aware split federated learning over edge computing systems. IEEE Transactions on Mobile Computing. https://doi.org/10.1109/TMC.2026.3673358

Lin Z, Qu G, Wei W, Chen X, and Leung KK (2025a). AdaptSFL: Adaptive split federated learning in resource-constrained edge networks. IEEE Transactions on Networking, 33(6): 2993-3008. https://doi.org/10.1109/TON.2025.3577790

Lin Z, Wei W, Chen Z, Lam CT, Chen X, Gao Y, and Luo J (2025b). Hierarchical split federated learning: Convergence analysis and system optimization. IEEE Transactions on Mobile Computing, 24(10): 9352-9367. https://doi.org/10.1109/TMC.2025.3565509

Lin Z, Zhu G, Deng Y, Chen X, Gao Y, Huang K, and Fang Y (2024). Efficient parallel split learning over resource-constrained wireless edge networks. IEEE Transactions on Mobile Computing, 23(10): 9224-9239. https://doi.org/10.1109/TMC.2024.3359040

Liu X, Deng Y, and Mahmoodi T (2023). Wireless distributed learning: A new hybrid split and federated learning approach. IEEE Transactions on Wireless Communications, 22(4): 2650-2665. https://doi.org/10.1109/TWC.2022.3213411

Mallidi SKR and Ramisetty RR (2025). Advancements in training and deployment strategies for AI-based intrusion detection systems in IoT: A systematic literature review. Discover Internet of Things, 5: 8. https://doi.org/10.1007/s43926-025-00099-4

Minh QN, Nguyen VH, Quy VK, Ngoc LA, Chehri A, and Jeon G (2022). Edge computing for IoT-enabled smart grid: The future of energy. Energies, 15(17): 6140. https://doi.org/10.3390/en15176140

Nanayakkara SI, Pokhrel SR, and Li G (2024). Understanding global aggregation and optimization of federated learning. Future Generation Computer Systems, 159: 114-133. https://doi.org/10.1016/j.future.2024.05.009

Nassef L (2026). A lightweight machine learning-based intrusion detection system for smart grids. International Journal of Advanced and Applied Sciences, 13(3): 52-67. https://doi.org/10.21833/ijaas.2026.03.006

Papadopoulos C, Kollias KF, and Fragulis GF (2024). Recent advancements in federated learning: State of the art, fundamentals, principles, IoT applications and future trends. Future Internet, 16(11): 415. https://doi.org/10.3390/fi16110415

Samikwa E, Di Maio A, and Braun T (2024). DFL: Dynamic federated split learning in heterogeneous IoT. IEEE Transactions on Machine Learning in Communications and Networking, 2: 733-752. https://doi.org/10.1109/TMLCN.2024.3409205

Sun Y, Hu G, Teng Y, and Cai D (2025). Split federated learning over heterogeneous edge devices: Algorithm and optimization. In the 2025 IEEE Wireless Communications and Networking Conference (WCNC), IEEE, Milan, Italy: 1-6. https://doi.org/10.1109/WCNC61545.2025.10978241

Tariq N, Alsirhani A, Humayun M, Alserhani F, and Shaheen M (2024). A fog-edge-enabled intrusion detection system for smart grids. Journal of Cloud Computing, 13: 43. https://doi.org/10.1186/s13677-024-00609-9

Timofte EM, Dimian M, Graur A, Potorac AD, Balan D, Croitoru I, Hrițcan DF, and Puşcaşu M (2025). Federated learning for cybersecurity: A privacy-preserving approach. Applied Sciences, 15(12): 6878. https://doi.org/10.3390/app15126878

Xu H, Seng KP, Smith J, and Ang LM (2024). Multi-level split federated learning for large-scale AIoT system based on smart cities. Future Internet, 16(3): 82. https://doi.org/10.3390/fi16030082

Yaseen MB, Wan F, Riaz MB, Rahman MO, and Fang X (2026). Privacy-preserving federated reinforcement learning for autonomous IoT intrusion response: A deep multi-agent hierarchical framework with advanced cryptographic security. Results in Engineering, 29: 109458. https://doi.org/10.1016/j.rineng.2026.109458

Zhu G, Deng Y, Chen X, Zhang H, Fang Y, and Wong TF (2024). ESFL: Efficient split federated learning over resource-constrained heterogeneous wireless devices. IEEE Internet of Things Journal, 11(16): 27153-27166. https://doi.org/10.1109/JIOT.2024.3397677