

A study on risk prediction and asset operation-and-maintenance cost optimization in power internet of things systems, driven by the fusion of digital twin technology and federated graph learning

Yingchun Zhan ^{*,a}

School of Economics and Management, Anyang University, Anyang, China

Article Info	Abstract
Article History: Received 13 July 2026 Accepted 26 Aug 2026 Keywords: Power internet of things; Digital twin; Federated graph learning; Risk prediction; Operation and maintenance cost	The deployment of sensors, smart terminals, and edge gateways is driving the transformation of traditional power grids into the power Internet of Things (IoT). However, asset operation and maintenance still face data fragmentation, insufficient topology-aware modeling, delayed risk response, and high operational costs. This paper proposes a Digital Twin-assisted Federated Graph Learning framework (DT-FGL), which constructs a layered architecture consisting of physical assets, data communication, hybrid twin modeling, federated graph risk prediction, and decision services. The framework integrates electrical topology, device status, environmental conditions, communication quality, and maintenance records to generate a heterogeneous asset graph. Here, rare-fault recall refers to the recall of the minority positive class, defined as equipment faults occurring within the future prediction horizon. Cascading risk is modeled through learned multi-hop graph propagation over electrical, spatial, communication, co-failure, and resource-sharing relations, rather than through full physical power-flow simulation. Node-level risk probabilities are then mapped into inspection and maintenance decisions through a budget- and risk-constrained optimization layer. Synthetic benchmark experiments show that DT-FGL improves rare-fault recall to 0.317 and achieves an AUC of 0.804, while reducing operation and maintenance costs by 43.8% relative to fixed-interval maintenance. Because the current validation is based on synthetic benchmark data, the reported gains should be interpreted as simulation evidence that requires further verification on field power IoT datasets.

© 2026 MIM Research Group. All rights reserved.

1. Introduction

The power system is gradually transitioning from traditional centralized monitoring and periodic maintenance toward a power Internet of Things (IoT) operation paradigm supported by massive sensors, edge gateways, intelligent terminals, and cloud platforms [1,2]. The extensive deployment of distribution transformers, switchgear, cables, energy storage devices, distributed energy resources, protection devices, and communication units has shifted maintenance decision-making from a “ledger-driven” paradigm to a “data- and model-driven” paradigm. This transformation significantly enhances system state awareness; however, it also introduces new risk characteristics. Device operating conditions are jointly influenced by load fluctuations, meteorological conditions, network topology, communication quality, and control strategies. As a result, failures are no longer confined to individual devices but may propagate in a cascading manner along electrical connections, spatial proximity, service dependencies, and communication links [15,18].

*Corresponding author: X1609503684@outlook.com

^aorcid.org/0009-0007-5310-0100

DOI: <https://dx.doi.org/10.17515/resm2026-1859am0713rs>

Res. Eng. Struct. Mat. Vol. x Iss. x (xxxx) xx-xx

Existing power asset maintenance methods still face three major challenges. First, real-time risk identification remains insufficient because conventional condition assessment methods mainly rely on offline inspections, manual patrols, or single-point threshold alarms, making it difficult to capture dynamic coupling among equipment degradation, topology, and multi-source disturbances [3,4]. Second, cross-regional data collaboration is constrained by privacy and operational security concerns, which prevents direct pooling of load profiles, equipment states, and dispatching information [10,23]. Third, risk prediction is often weakly coupled with maintenance decision-making: many studies estimate failure probability or health indices but do not explicitly map those estimates to inspection scheduling, maintenance actions, and resource allocation [25]. The specific gap addressed in this study is therefore not the isolated use of digital twins, federated learning, graph learning, or maintenance optimization, but the lack of an integrated and traceable framework that connects topology-aware raw-data-local risk prediction with cost-constrained operation-and-maintenance decisions.

Digital twin technology provides a new technical pathway for risk modeling and maintenance optimization in power IoT systems [2,5]. Existing studies have demonstrated that digital twins achieve promising performance in condition assessment and fault localization for substations and electrical equipment [6,7]. In the context of coal mine power grids, digital twin systems have been constructed through a multi-layer architecture consisting of a physical entity layer, a digital twin model layer, and a service layer, enabling multi-level system modeling [36]. In microgrid scenarios, high-fidelity digital twin models combined with predictive control methods can transform state prediction results into energy management and cost optimization strategies [34]. These studies indicate that digital twins are not merely visualization tools, but should serve as core engines for state mapping, fault simulation, and decision feedback. Moreover, digital twins can fuse and denoise multi-source observations, providing more robust health state estimation than individual noisy sensors [9].

Under data security constraints, federated learning enables collaborative modeling among multiple parties without sharing raw data, making it suitable for distributed training scenarios involving multiple substations and edge nodes in power IoT systems [27]. Federated graph learning has begun to address graph-structured learning under privacy constraints in general graph benchmarks and industrial IoT applications [10-14]. However, existing approaches rarely combine five relation types of power assets, digital-twin-derived state deviation, quality-aware client selection, and a downstream maintenance optimization layer in one closed-loop power IoT framework. The novelty of this work is positioned at this integration point and in the explicit handoff from node-level risk probabilities to constrained maintenance decisions, rather than in claiming a new standalone GNN or federated optimizer.

Based on the above analysis, this paper focuses on three key issues: improving risk prediction accuracy, enhancing cross-domain data collaboration security, and optimizing maintenance cost. We propose a power IoT risk prediction and asset maintenance optimization framework that integrates digital twins and federated graph learning. The main contributions are as follows:

1.1 Heterogeneous Digital Twin Graph Modeling

To address multi-source heterogeneous data and cascading failure propagation in power IoT systems, a multi-relational heterogeneous asset graph is constructed by integrating electrical connectivity, spatial proximity, communication dependencies, historical co-failure relationships, and operation-and-maintenance resource-sharing relationships. The methodological contribution is the use of this heterogeneous graph as a common representation linking twin deviation, local fault precursors, federated graph learning, and maintenance optimization. The present implementation uses relation-specific normalized adjacency matrices that are combined by non-negative relation weights before multi-hop propagation; more expressive relation attention mechanisms are left for future validation.

1.2 Quality-Aware Federated Graph Learning for Risk Prediction

A federated graph learning framework incorporating both data quality and model quality evaluation is proposed. By leveraging a digital twin-assisted client selection mechanism and

quality-weighted model aggregation strategy, the framework suppresses the impact of low-quality updates under non-IID data distributions. This improves the stability and generalization capability of the global model, particularly enhancing rare-fault detection performance while reducing communication and training overhead.

1.3 Risk-Constrained Maintenance Optimization Model

A two-stage maintenance optimization framework is developed based on risk prediction results. First, node-level risk probabilities are converted into candidate inspection or maintenance actions by comparing predicted risk with a decision threshold and by ranking assets according to predicted risk multiplied by asset importance. Second, a constrained optimization step selects actions under inspection-budget, maintenance capacity, resource-availability, and risk-threshold constraints. This contribution is a decision-coupling mechanism that makes the prediction output operationally actionable; it should be interpreted as a framework-level optimization formulation rather than a claim of a new mathematical programming solver.

2. Related Work

2.1 Digital Twin of Power Equipment and Fault Diagnosis

Digital twin technology in power systems is mainly applied to state mapping, predictive simulation, equipment health assessment, fault diagnosis, and operational optimization. A systematic review covering studies from 2017 to 2024 on digital-twin-based intelligent fault diagnosis highlights its potential in physical–virtual mapping, degradation modeling, and early warning mechanisms [5].

At the power equipment level, studies on substation equipment health assessment and fault diagnosis [6], as well as digital twin-based motor fault diagnosis [7], demonstrate that digital twin models can effectively support condition evaluation and fault localization. In the wind power domain, a comprehensive review further proposes a multi-level maturity framework ranging from basic condition monitoring to autonomous control, emphasizing the integration of digital twins with predictive maintenance strategies [9]. The foundational concept of digital twins, originally introduced by Grieves, provides the theoretical basis for these developments [8].

These studies collectively establish a physical-layer, data-layer, model-layer, and service/application-layer foundation for digital-twin-based power system operation. Nevertheless, the literature remains uneven in how it connects twin state estimation, cross-site raw-data-local learning, topology-aware propagation, and maintenance optimization. Some recent works consider large-scale or hierarchical digital twins, so it would be too strong to claim that existing work is limited only to single-device or single-substation scenarios. The remaining limitation is instead that most studies emphasize monitoring, diagnosis, or control separately, while few provide an explicit end-to-end mapping from heterogeneous asset states to federated graph risk prediction and cost-constrained O&M decisions.

2.2 Federated Graph Learning (FGL)

Federated learning addresses privacy protection and cross-domain modeling by sharing model parameters instead of raw data [27]. Its optimization under heterogeneous networks, such as FedProx introducing a proximal term to mitigate client drift, further improves stability under non-IID settings [32]. However, conventional federated learning typically ignores graph structures. Federated graph learning (FGL) achieves a balance between privacy preservation and topological representation: FedGraphNN provides a unified federated GNN framework and benchmark [10]; FedGraph designs intelligent sampling for clients holding subgraphs [11]; FedGL mitigates client heterogeneity via global self-supervision [12]; methods such as spectral graph transformers and neural ordinary differential equations for Non-IID graphs improve cross-distribution generalization [13]; and FedGraphHE enhances gradient privacy in federated GNNs using homomorphic encryption [14]. Building on these ideas, this work introduces a twin-assisted quality-weighted aggregation strategy for power asset relational networks.

2.3 Risk Prediction in Power Systems Using Graph Neural Networks

Power system risks exhibit significant topological propagation characteristics. Graph neural networks have been applied to cascading failure prediction and high-risk line identification. For example, the physics-informed graph neural jumping ordinary differential equation model (PI-GN-JODE) predicts both edge-level and node-level failure probabilities on the IEEE 24-bus and 118-bus systems [15]. Graph neural networks incorporating node and edge attributes have been used to predict vulnerable buses and branches after cascading failures [16]. The lightweight graph physics-informed attention network (GPIAN) integrates power grid physical laws into the attention mechanism for cascading failure risk assessment [17]. In addition, graph neural network-based methods for non-local cascading failure prediction and mitigation reveal the non-local propagation characteristics of power system risks [18]. These studies support the modeling motivation of this work, namely representing cascading risks through graph structures; however, they generally assume that data can be centrally collected [29].

2.4 Integration of Digital Twin and Federated Learning

In recent years, there have been emerging efforts to integrate digital twins with federated learning in smart grids. The adaptive asynchronous federated learning framework for digital twin-driven smart grids, namely PPSG, combines blockchain with asynchronous federated learning to handle device state updates and non-IID data [19]. FSDOF integrates digital twins, multi-agent reinforcement learning, and graph neural networks to enable secure collaborative optimization in nano grids [20]. Hierarchical digital twin frameworks introduce federated and distributed learning into the twin architecture to protect data privacy across multiple microgrids [21]. Real-time photovoltaic fault detection has further combined federated learning with digital twins [22]. These studies are closely related to the direction of this work. Building on them, this paper explicitly introduces heterogeneous asset graphs and a closed-loop framework for risk-constrained cost optimization.

2.5 Federated Anomaly Detection and Predictive Maintenance

In terms of anomaly detection and predictive maintenance, federated anomaly detection based on multi-source information fusion integrates smart meters, IoT sensor logs, and substation indicators [23]. The fusion-enhanced adversarially robust federated learning method, CBAA-FedAvg, improves the robustness of smart grids under data poisoning attacks [24]. Predictive maintenance optimization based on reliable lifetime prediction and risk-sensitive reinforcement learning incorporates failure probability and uncertainty into inspection, maintenance, and replacement decisions, thereby balancing safety and cost [25]. Remote anomaly detection for resource-constrained Industrial 4.0 devices further discusses deployment issues under bandwidth and computing constraints [26]. The risk-constrained cost optimization layer proposed in this paper is complementary to the above predictive maintenance studies, with an emphasis on cost-optimal decision-making weighted by asset criticality.

Table 1 summarizes the positioning of representative related work. Digital-twin studies mainly emphasize state mapping, diagnosis, and simulation; federated learning studies emphasize raw-data-local training; graph-learning studies emphasize topology-aware failure prediction; and predictive-maintenance studies emphasize cost or reliability decisions. The proposed DT-FGL framework differs by coupling these elements into one risk-to-decision workflow for power IoT asset management. This comparison clarifies that the contribution is primarily a closed-loop framework and modeling pipeline, supported by relation-weighted heterogeneous multi-hop propagation, twin-quality-aware federated aggregation, and risk-to-decision coupling.

Table 1. Comparison of representative research streams and their gaps relative to DT-FGL

Research stream	Representative references	Main focus	Gap relative to DT-FGL
Digital twins	[5-9]	State mapping, diagnosis, and simulation	Limited coupling with federated graph learning and O&M decisions
Federated graph learning	[10-14]	Raw-data-local graph model training	Usually lacks digital-twin deviation and maintenance optimization
Cascading-risk GNNs	[15-18]	Topology-aware failure propagation	Usually assumes centralized data access
DT-FL smart-grid frameworks	[19-22]	Twin-assisted distributed learning or control	Partial treatment of heterogeneous asset relations and cost decisions
Predictive maintenance	[23-26]	Reliability and cost-aware decisions	Often weakly connected to federated topology-aware risk prediction
Proposed DT-FGL	This study	Rare-fault risk prediction and constrained maintenance	Connects twin features, heterogeneous graph propagation, raw-data-local learning, and O&M optimization

3. Method

3.1 Problem Formulation

Assume that the power Internet of Things consists of K sites or edge regions, and each site k corresponds to a local asset graph $G_k = (V_k, E_k, X_k, A_k^r)$, where V_k denotes equipment nodes, E_k denotes the set of multi-relation edges, X_k denotes the node feature tensor, and A_k^r denotes the relation-specific adjacency tensor for relation r . The relation set includes electrical, spatial, communication, historical co-failure, and operation-and-maintenance resource-sharing relations.

$$G_k = (V_k, E_k, X_k, A_k) \quad (1)$$

Where, V_k denotes the set of equipment nodes, E_k represents electrical connections, spatial proximity, communication dependencies, or historical co-failure relationships, X_k denotes the node feature tensor, and A_k represents the multi-relation adjacency tensor.

The objective is to train a global risk prediction model f_θ without sharing raw operational data or detailed local graph structures. The model outputs a node-level risk probability $\hat{y}_{i,t+h} = f_\theta(X_{i,t-L:t}, A_k^r, Z_{i,t}^{DT})$ for the prediction horizon h . The optimization layer then maps $\hat{y}_{i,t+h}$ a binary maintenance or inspection decision $u_{i,t,h} \in \{0,1\}$. An asset becomes a decision candidate when $\hat{y}_{i,t+h} > \epsilon$. exceeds the risk threshold epsilon, or when its importance-weighted risk score $s_{i,t,h} = w_i \hat{y}_{i,t+h}$ ranks within the available inspection budget.

$$\hat{y}_{i,t+h} = f_\theta(G_{k,t-L:t}, Z_{k,t-L:t}, D_{k,t}) \quad (2)$$

Where, L is the historical window length, H is the prediction horizon, $Z_{i,t}^{DI}$ denotes simulated states, deviation indicators, and predicted scenarios generated by the digital twin, and $M_{i,t}$ represents equipment ledgers and operation-and-maintenance records. The decision variable $u_{i,t,h} = 1$ indicates that the asset is selected for inspection or maintenance in the horizon h ; otherwise $u_{i,t,h} = 0$.

$$y_{i,t} = 1 \left[\sum_{\tau=t+1}^{t+H} f_{ail_{i,\tau}} > 0 \right] \quad (3)$$

Where, the summation index τ runs from $t + 1$ to $t + H$; t denotes the current time, H is the full prediction horizon, and h denotes a relative step within that horizon.

The cost optimization objective is formulated as: $\min_u \sum_{t=1}^T (C_t^{ins} + C_t^{rep} + C_t^{out} + C_t^{deg} + C_t^{com} + C_t^{cmp})$, subject to $\sum_t C_t^{ins} u_{i,t,\tau} \leq B_t$, $\sum_t u_{i,t,\tau} \leq M_t$, $\sum_t w_i \hat{y}_{i,t+\tau} (1 - u_{i,t,\tau}) \leq \epsilon_t$ and $u_{i,t,\tau} \in \{0,1\}$. These constraints represent the inspection budget, maintenance capacity, residual importance-weighted risk exposure, and binary action feasibility, respectively. If a deployment does not impose a hard residual-risk threshold, the same formulation can be treated as risk-penalized cost minimization.

$$\min_{u_{1:T}} \sum_{t=1}^T (C_t^{ins} + C_t^{rep} + C_t^{out} + C_t^{deg} + C_t^{com} + C_t^{cmp}) \quad (4)$$

Where, C_t^{ins} , C_t^{rep} , C_t^{out} , C_t^{deg} , C_t^{com} and C_t^{cmp} denote inspection cost, scheduled repair cost, outage loss, equipment degradation cost, federated communication overhead, and computing overhead, respectively. C_t^{com} is accumulated from upload and download rounds between edge sites and the cloud server, while C_t^{cmp} is accumulated from edge/cloud training, inference, and scheduling computation. The benchmark uses normalized cost units relative to the fixed-period maintenance baseline; the calibration rules are summarized in Table 2. For the reported synthetic benchmark, C_t^{com} and C_t^{cmp} enter the objective as normalized per-round penalty terms for model transfer and edge/cloud computation, respectively; they are not interpreted as field-measured monetary expenditures and are therefore not itemized as separate outcome columns in Table 7. Field deployment requires recalibration from measured model size, participating rounds, bandwidth, energy use, compute time, and local tariffs.

Table 2. Normalized cost parameters used in the maintenance optimization benchmark

Parameter	Meaning	Benchmark calibration	Unit
C_t^{ins}	Inspection action for asset i	Normalized labor and routine inspection cost	cost unit/action
C_t^{rep}	Scheduled repair or preventive maintenance	Normalized planned repair cost	cost unit/action
C_t^{out}	Outage loss after missed or emergency failure	Asset-importance-weighted outage penalty	cost unit/event
C_t^{deg}	Residual degradation cost	Penalty proportional to unrepaired degradation/risk	cost unit/time
C_t^{com}	Federated communication	Upload/download overhead per participating round	cost unit/round
C_t^{cmp}	Computing overhead	Training, inference, and scheduling overhead	cost unit/round
B_t, M_t, ϵ_t	Budget, capacity, and residual-risk thresholds	Deployment-defined constraints in the benchmark scale	normalized

The overall framework of this study is shown in Fig. 1. From bottom to top, the architecture includes the physical asset layer, data communication layer, digital twin modeling layer, heterogeneous asset graph federated learning layer, and the top-level decision service and operation and maintenance optimization layer. First, multi-source data are collected by the physical and communication layers. Second, the digital twin modeling layer outputs reliability-related metrics such as twin deviation through cyber-physical mapping and failure simulation.

Then, under the premise that each site does not disclose private information, the federated graph learning layer extracts multi-hop risk propagation features and performs quality-weighted aggregation in the cloud to achieve collaborative training. Finally, based on node-level risk prediction results, the optimization layer solves the risk-constrained decision problem, generates integrated operation and maintenance decisions, and feeds them back to the physical entities, thereby constructing a full-process closed-loop security defense framework.

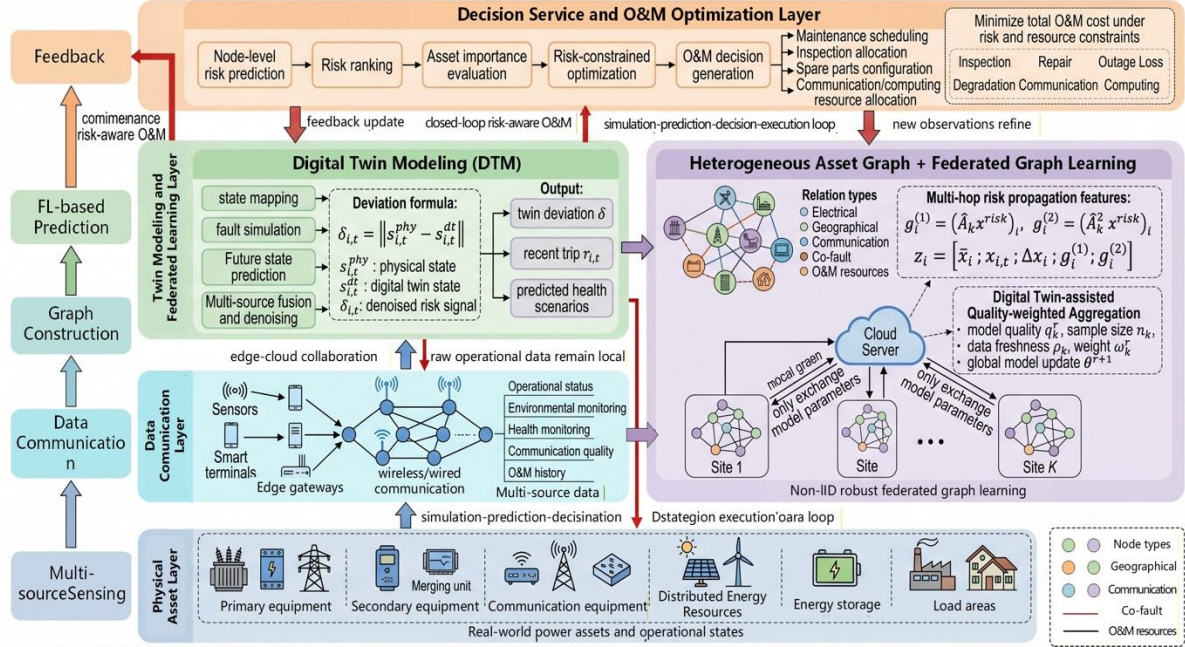


Fig. 1. Overall framework for risk prediction and asset maintenance cost optimization in power internet of things integrating digital twins and federated graph learning

3.2 Digital Twin Assisted Architecture

This paper adopts a four-layer digital twin architecture, as shown in Fig. 2. The physical asset layer includes distribution networks, microgrids, and key equipment, serving as the source of state data and risk events. The data communication layer consists of sensors, intelligent terminals, edge gateways, 5G/private network communication, and data buses, and is responsible for continuous data acquisition, preprocessing, identification, and transmission. The twin modeling layer integrates physical mechanism models, data-driven models, and graph-structured models to realize equipment state mapping, future state prediction, fault scenario simulation, and twin deviation assessment. The decision service layer provides risk dashboards, optimization strategies, and feedback updates for dispatching, inspection, maintenance, spare parts management, and cost control.

In this paper, the digital twin is treated as a hybrid physics-informed and data-driven state-estimation layer. It performs three main functions. First, at the data layer, it checks abnormal values, missing values, and sensor drift, and fuses multi-source observations to obtain smoother health-state estimates; this role is used as a modeling design rather than independent proof that the twin is always more accurate than any single sensor. Second, at the model layer, it generates features related to equipment health status, load prediction, environmental disturbances, and fault simulation, including twin deviation delta and predicted scenarios. Third, at the decision layer, it enables simulation-based evaluation of operation-and-maintenance strategies before feedback is applied to the physical system. In the benchmark, the twin state is synchronized by an exponential update, $s_{i,t}^{DT} = (1 - \eta)s_{i,t|t-1}^{DT} + \eta s_{i,t}^{phy}$, where η controls the measurement-update weight; periodic recalibration uses accumulated maintenance records and fault logs to reduce long-term drift.

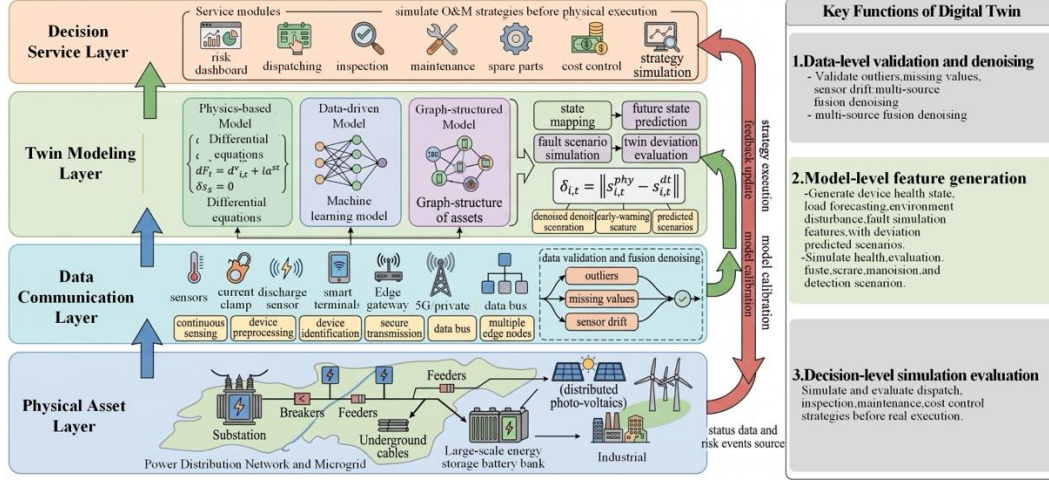


Fig. 2. Four-layer digital twin architecture and data closed loop for the power internet of things

3.3 Heterogeneous Asset Graph Construction

This paper constructs a multi-relation heterogeneous asset graph, as shown in Fig. 3. The node types include primary equipment, secondary equipment, communication devices, distributed energy resources, energy storage devices, and load areas. The edge types include electrical connection edges, geographical proximity edges, communication dependency edges, historical co-failure edges, and operation-and-maintenance resource-sharing edges. The node features are divided into operational features, environmental features, health features, communication features, and cost features, with a total of 20 dimensions, as shown in Table 3.

For equipment node i , the twin deviation is defined as:

$$\delta_{i,t} = \|s_{i,t}^{phy} - s_{i,t}^{dt}\| \quad (5)$$

Where, the first term denotes the physically collected state and the second term denotes the predicted or simulated state generated by the digital twin. The twin deviation is designed to summarize the mismatch between the physical observation and the simulated state. A continuous increase in the deviation may indicate abrupt changes in equipment condition, sensor anomalies, or model mismatch; in this study it is treated as an early-warning feature rather than as direct evidence that the digital twin has lower measurement noise than all individual sensors.

In addition, an observable recent tripping record is defined as the exponentially smoothed statistic of the equipment's own protection actions or alarms. For the multi-relation graph, each relation r has a degree-normalized propagation operator $\bar{A}_r$. The relation-weighted propagation operator is

$\bar{A} = \sum_r \alpha_r \bar{A}_r$, $\alpha_r \geq 0$, and $\sum_r \alpha_r = 1$. Unless relation-level performance is being studied separately, the benchmark uses fixed equal weights for the five relations. This makes the propagation mechanism explicit while keeping the model lightweight; learning relation-specific attention weights is an extension for future experiments.

$$\hat{A}_k = D_k^{-\frac{1}{2}} A_k D_k^{-\frac{1}{2}} \quad (6)$$

Then, the neighborhood-level cascading risk of node i can be expressed through feature propagation as:

$$g_i^{(1)} = (\hat{A}_k x^{risk})_i, \quad g_i^{(2)} = (\hat{A}_k^2 x^{risk})_i \quad (7)$$

Where x^{risk} includes risk-bearing features such as twin deviation, recent tripping records, alarm frequency, partial discharge, and oil temperature.

Finally, the node representation is obtained by concatenating the window-level statistical features with the first-order and second-order neighborhood aggregations:

$$z_i = [\bar{x}_i; x_{i,t}; \Delta x_i; g_i^{(1)}; g_i^{(2)}] \quad (8)$$

This construction is equivalent to multi-hop message passing in simplified graph convolution.

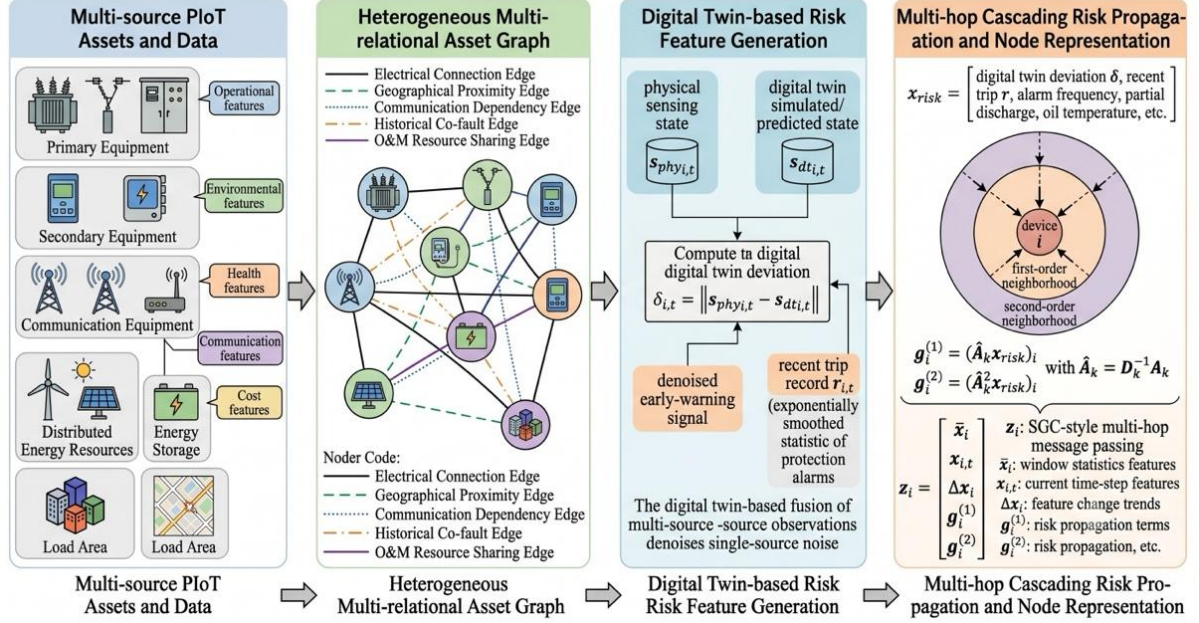


Fig. 3. Heterogeneous asset graph construction and multi-hop risk propagation for plot

Table 3. Nodes, edges, features, and data sources of the heterogeneous asset graph in the power internet of things

Graph component	Type/Object	Specific content	Main meaning	Data sources
Node type	Primary equipment nodes	Transformers, circuit breakers, switchgear, busbars, cables, feeders, etc.	Represent the core physical assets involved in power transmission, transformation, and distribution, serving as the main objects for risk prediction and operation and maintenance decision-making.	SCADA, PMU, smart meters, online monitoring terminals, equipment ledgers
Node type	Secondary equipment nodes	Relay protection devices, measurement and control devices, automation terminals, intelligent electronic devices, etc.	Represent protection, measurement, control, and alarm functions; their abnormalities may affect fault identification and isolation capability.	Protection device logs, IED records, alarm systems, automation systems

Node type	Communication equipment nodes	Edge gateways, routers, switches, 5G/private network communication units, data acquisition terminals, etc.	Represent the data transmission and edge collaboration capability of the power Internet of Things; communication abnormalities may affect state perception and federated training.	Communication network management systems, edge gateway logs, data buses, network monitoring systems
Node type	Distributed energy resource nodes	Photovoltaics, wind power, micro gas turbines, controllable distributed power sources, etc.	Represent source-side fluctuations and local power supply capability, which affect load balancing, power flow variation, and operational risk.	Distributed energy monitoring platforms, inverter data, microgrid energy management systems
Node type	Energy storage equipment nodes	Battery energy storage systems, supercapacitors, energy storage converters, etc.	Represent peak shaving, valley filling, backup support, and fault buffering capability, while also involving degradation and cost constraints.	BMS, PCS, energy storage monitoring systems, operation and maintenance records
Node type	Load area nodes	Distribution transformer areas, feeder load areas, important user areas, industrial parks, etc.	Represent load demand, outage impact scope, and asset importance, serving as an important basis for outage loss assessment in cost optimization.	AMI, load forecasting systems, user-side metering systems, GIS and marketing systems
Edge type	Electrical connection edges	Feeder connections, busbar connections, upstream and downstream transformer connections, switch tie relationships	Describe the electrical topology among assets and are used to characterize the paths through which risk propagates along the electrical network.	Distribution automation systems, EMS/DMS, topology models, single-line diagrams
Edge type	Geographical proximity edges	Devices located in the same station room, distribution area, corridor, or within a short spatial distance	Describe spatially correlated risks caused by environmental disturbances, natural disasters, temperature and	GIS, equipment coordinates, station room/line archives, environmental

			humidity, and external damage.	monitoring data
Edge type	Communication dependency edges	Dependencies between devices and gateways, terminals and communication links, edge nodes and cloud platforms	Describe dependency paths for data acquisition, model updates, and alarm transmission, and are used to identify perception risks caused by communication abnormalities.	Communication topology, network management platforms, edge gateway logs, link quality monitoring
Edge type	Historical co-failure edges	Device relationships involving co-alarms, co-trips, or fault associations within adjacent time windows	Describe potential fault coupling relationships and cascading risk propagation characteristics among devices.	Fault recorders, protection action records, alarm logs, maintenance history
Edge type	Operation and maintenance resource sharing edges	Asset relationships sharing maintenance teams, spare parts, outage windows, communication resources, or edge computing resources	Describe resource competition and collaboration constraints in operation and maintenance, supporting cost optimization and resource scheduling.	CMMS, ERP, spare parts inventory, work order systems, maintenance plans
Node feature	Operational features	Load rate, voltage deviation, current fluctuation, power variation, etc.	Represent the current operating stress and abnormal operating state of equipment.	SCADA, PMU, smart meters, online monitoring terminals
Node feature	Environmental features	Temperature, humidity, rainfall, wind speed, thunderstorm or extreme weather indicators, etc.	Represent the influence of external environments on equipment degradation, failure rate, and operational risk.	Meteorological platforms, station-side environmental sensors, wireless sensor networks
Node feature	Health features	Insulation condition, partial discharge, oil temperature, alarm frequency, recent trip records, twin deviation ($\Delta_{i,t}$), etc.	Represent equipment degradation, abnormality accumulation, and early fault symptoms; among them, twin deviation reflects the difference	Online monitoring systems, protection and alarm systems, digital twin platforms, operation and maintenance

			between the physically collected state and the digital twin predicted state.	inspection records
Node feature	Communication features	Packet loss rate, latency, link availability, data freshness, etc.	Represent communication reliability during data acquisition and federated model updating.	Communication network management systems, edge gateways, 5G/private network link monitoring
Node feature	Cost features	Inspection cost, repair cost, outage loss, spare parts cost, equipment importance, etc.	Represent the economic constraints and optimization objectives when risk decisions are transformed into operation and maintenance actions.	Operation and maintenance management systems, financial systems, ERP, asset ledgers, historical work orders
Graph propagation feature	First-order neighborhood risk $g_i^{(1)}$	$g_i^{(1)} = (\hat{A}_k x^{risk})_i$	Represent the risk impact from assets directly connected to the target node.	Calculated from the topology of the heterogeneous asset graph and risk-bearing features
Graph propagation feature	Second-order neighborhood risk $g_i^{(2)}$	$g_i^{(2)} = (\hat{A}_k^2 x^{risk})_i$	Represent the potential cascading risk propagation effect within the two-hop neighborhood.	Calculated from the topology of the heterogeneous asset graph and risk-bearing features
Final node representation	Integrated node representation z_i	$z_i = [\bar{x}_i; x_{i,t}; \Delta x_i; g_i^{(1)}; \ell$	Integrates window-level statistical features, current states, changing trends, and multi-hop neighborhood risks, and is used as the input to the subsequent federated graph risk prediction model.	Multi-source sensing data, digital twin outputs, operation and maintenance records, and graph-structure calculation results
Node type	Load area nodes	Distribution transformer areas, feeder load areas,	Represent load demand, outage impact scope, and	AMI, load forecasting systems, user-

		important user areas, industrial parks, etc.	asset importance, serving as an important basis for outage loss assessment in cost optimization.	side metering systems, GIS and marketing systems
Edge type	Electrical connection edges	Feeder connections, busbar connections, upstream and downstream transformer connections, switch tie relationships	Describe the electrical topology among assets and are used to characterize the paths through which risk propagates along the electrical network.	Distribution automation systems, EMS/DMS, topology models, single-line diagrams
Edge type	Geographical proximity edges	Devices located in the same station room, distribution area, corridor, or within a short spatial distance	Describe spatially correlated risks caused by environmental disturbances, natural disasters, temperature and humidity, and external damage.	GIS, equipment coordinates, station room/line archives, environmental monitoring data
Edge type	Communication dependency edges	Dependencies between devices and gateways, terminals and communication links, edge nodes and cloud platforms	Describe dependency paths for data acquisition, model updates, and alarm transmission, and are used to identify perception risks caused by communication abnormalities.	Communication topology, network management platforms, edge gateway logs, link quality monitoring
Edge type	Historical co-failure edges	Device relationships involving co-alarms, co-trips, or fault associations within adjacent time windows	Describe potential fault coupling relationships and cascading risk propagation characteristics among devices.	Fault recorders, protection action records, alarm logs, maintenance history
Edge type	Operation and maintenance resource sharing edges	Asset relationships sharing maintenance teams, spare parts, outage windows, communication resources, or edge computing resources	Describe resource competition and collaboration constraints in operation and maintenance, supporting cost optimization and resource scheduling.	CMMS, ERP, spare parts inventory, work order systems, maintenance plans

Node feature	Operational features	Load rate, voltage deviation, current fluctuation, power variation, etc.	Represent the current operating stress and abnormal operating state of equipment.	SCADA, PMU, smart meters, online monitoring terminals
Node feature	Environmental features	Temperature, humidity, rainfall, wind speed, thunderstorm or extreme weather indicators, etc.	Represent the influence of external environments on equipment degradation, failure rate, and operational risk.	Meteorological platforms, station-side environmental sensors, wireless sensor networks
Node feature	Health features	Insulation condition, partial discharge, oil temperature, alarm frequency, recent trip records, twin deviation ($\Delta_{i,t}$), etc.	Represent equipment degradation, abnormality accumulation, and early fault symptoms; among them, twin deviation reflects the difference between the physically collected state and the digital twin predicted state.	Online monitoring systems, protection and alarm systems, digital twin platforms, operation and maintenance inspection records
Node feature	Communication features	Packet loss rate, latency, link availability, data freshness, etc.	Represent communication reliability during data acquisition and federated model updating.	Communication network management systems, edge gateways, 5G/private network link monitoring
Node feature	Cost features	Inspection cost, repair cost, outage loss, spare parts cost, equipment importance, etc.	Represent the economic constraints and optimization objectives when risk decisions are transformed into operation and maintenance actions.	Operation and maintenance management systems, financial systems, ERP, asset ledgers, historical work orders

3.4 Federated Graph Risk Prediction

Each site trains a local graph model f_θ composed of a two-hop SGC-style graph encoder, a temporal feature encoder based on window-level statistics and first differences, and a sigmoid risk output layer. The benchmark configuration uses graph depth 2, hidden dimension 64, ReLU activation in the temporal encoder, dropout 0.2, Adam optimization, learning rate 1×10^{-3} , weight decay 1×10^{-4} , batch size 128, 5 local epochs per communication round, and a 30-round communication

budget. Class-weighted cross-entropy is adopted for the rare positive fault class, and the objective also contains cost-regression, graph-smoothness, and risk-calibration terms.

$$L_k = L_k^{risk} + \lambda_1 L_k^{cost} + \lambda_2 L_k^{smooth} + \lambda_3 L_k^{cal} \quad (9)$$

To mitigate client drift, a FedProx proximal term can be added to the local objective:

$$\frac{\mu}{2} \|\theta_k - \theta^r\|^2 \quad (10)$$

The federated training process adopted in this paper is shown in Fig. 4. Federated training follows an edge-cloud collaborative paradigm. Local sites retain their raw data and graph structures, and only upload model updates. The cloud server aggregates these updates to generate a global model and then distributes it back to the local sites. To reduce communication costs and improve robustness under non-IID conditions, this paper introduces digital-twin-assisted quality-weighted aggregation and model selection. The local digital twin evaluates model quality on a risk calibration set, using AUC to reflect discriminative ability under class imbalance and data freshness to penalize stale or communication-degraded updates. The upload weight combines sample size, validation AUC, and freshness, so that larger and more reliable sites contribute more strongly while low-quality or stale updates are down-weighted or excluded. This weighting is heuristic rather than a formal robustness guarantee; under severe client heterogeneity, additional safeguards such as robust aggregation or differential privacy should be evaluated. To standardize the notation in Fig. 4 and Eqs. (10)-(11), q_k^r denotes the local validation AUC, $w_{q,k}^r = \max(q_k^r - 0.5, \epsilon)$ denotes the quality factor, dtq_k^r denotes the current digital-twin data-quality score, and $\rho_k^r = dtq_k^r \exp(-\Delta t_k^r)$ denotes freshness-adjusted twin quality. Thus, the product $w_{q,k}^r n_k [dtq_k^r \exp(-\Delta t_k^r)]$ shown in Fig. 4 is exactly the unnormalized numerator used to compute ω_k^r in Eq. (10).

$$\omega_k^r = \frac{n_k \cdot \max(q_k^r - 0.5, \epsilon) \cdot \rho_k^r}{\sum_{j \in S_r} n_j \cdot \max(q_j^r - 0.5, \epsilon) \cdot \rho_j^r} \quad (11)$$

$$\theta^{r+1} = \sum_{k \in S_r} \omega_k^r \theta_k^r \quad (12)$$

Where, $S_r = \{k: q_k^r \geq \tau\}$ denotes the set of sites selected in communication round r . Under this mechanism, noisy or low-quality sites are automatically down-weighted or excluded, thereby accelerating convergence and improving rare-fault recall in heterogeneous scenarios.

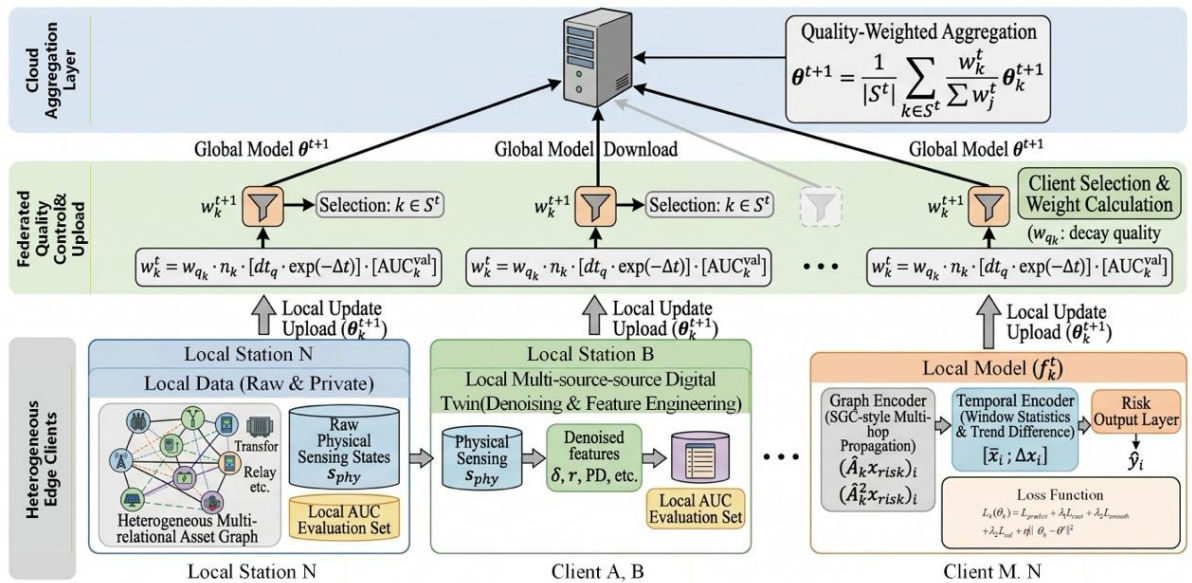


Fig. 4. Digital-twin-assisted federated graph learning training process

The maintenance optimization is solved procedurally as follows. First, the trained DT-FGL model generates node-level risk probabilities for the prediction horizon. Second, each node receives an importance-weighted risk score and a candidate action set. Third, candidate actions are sorted by expected avoided outage loss per unit inspection or maintenance cost, subject to the available inspection budget and maintenance capacity. Fourth, the selected schedule is checked against the residual-risk threshold; if the threshold is violated, the highest remaining importance-weighted risks are added until the constraint or budget limit is reached. For N candidate assets, the dominant sorting step has complexity $O(N \log N)$, while constraint checking is $O(N)$; therefore, the single-period heuristic has overall complexity $O(N \log N)$. A MILP solver can replace this heuristic when exact optimization and complete cost parameters are available.

4. Experimental Setup

4.1 PloT Benchmark Dataset

The core specifications of the PloT benchmark dataset are shown in Table 4. Its contents mainly include the following five dimensions. Topology: The benchmark contains 8 sites, with 44-80 nodes at each site and 520 asset nodes in total. Each site is composed of a feeder-tree backbone and a small number of tie-line or meshed edges, with 559 relational edges in the entire network. Electrical edges follow feeder adjacency, spatial edges connect assets within the same station room or nearby coordinates, communication edges connect devices to terminals, gateways, and cloud links, co-failure edges are generated from temporally adjacent alarms or trips, and resource-sharing edges connect assets competing for the same maintenance teams, spare parts, outage windows, or edge resources.

- Degradation and fault mechanism: Each node maintains a latent degradation variable that accumulates through a random walk with site-specific drift and environmental stress. Approximately 20% of the nodes are assigned accelerated aging characteristics. Faults are rare events, and their hazard rates are jointly determined by equipment degradation, twin deviation, neighborhood risk propagation, environmental disturbance, and site-specific baseline risk. After a fault occurs, the degradation state is reset to represent repair. The prediction label indicates whether a fault occurs within the future window $H=6$, with historical window $L=12$ and sampling step 5. Each configuration contains 60,840 samples, with a positive label rate of approximately 2.44% and a step-wise fault rate of approximately 0.41%.
- Digital twin versus raw sensing: Raw health sensing signals, such as insulation condition, partial discharge, and oil temperature, contain site-dependent noise. In the benchmark, twin deviation δ is treated as a smoothed latent-degradation feature derived from the hybrid twin update, and is intended to provide an early-warning signal when physical observations diverge from simulated state trajectories.
- Graph signals: The observable recent tripping records of nodes are propagated through first-order and second-order neighborhoods, thereby carrying cascading risk information.
- Non-independent and identically distributed data, namely non-IID data: The sites differ significantly in baseline hazard rate, degradation rate, climate conditions, and data quality (0.4–1.0), which is used to evaluate the effectiveness of quality-weighted aggregation and model selection.
- Synthetic data generation and reproducibility: The benchmark uses chronological splitting with 70% training, 10% validation, and 20% testing. Latent degradation follows $d_{i,t+1} = d_{i,t} + \mu_i + \beta_i e_t + \xi_{i,t}$, where $\xi_{i,t}$ is zero-mean Gaussian degradation noise, e_t is environmental stress, and accelerated-aging nodes use larger drift μ_i . Fault probability follows $p_{i,t} = \text{sigmoid}(\beta_0 + \beta_1 d_{i,t} + \beta_2 \delta_{i,t} + \beta_3 g_{i,t} + \beta_4 e_t)$, where $g_{i,t}$ is the propagated neighborhood risk. Raw sensor channels are generated by adding site-quality-dependent Gaussian noise to latent health variables, and faults reset the degradation state to represent repair. Unless otherwise specified, the reported benchmark uses random seed 42; additional seeds should be run for final statistical testing.

Table 4. Scale, node types, feature dimensions, and risk label definitions of the PIoT experimental benchmark

Item	Value/Description
Number of sites (K)	8
Number of asset nodes (N)	520, with 44–80 nodes at each site
Number of relational edges	559, including feeder-tree connections and meshed tie links
Time steps (T) / historical window (L) / prediction horizon (H)	600 / 12 / 6
Feature dimension	20, including operational features (4), environmental features (4), health features (6, including twin deviation and recent trip records), communication features (3), and cost features (3)
Number of samples per configuration	60,840
Positive label ratio	Approximately 2.44%, indicating rare events
Data split	70% training, 10% validation, and 20% testing in chronological order

4.2 Comparison Methods, Evaluation Metrics and Baselines

The comparison methods for risk prediction are divided into two groups. The first group consists of centralized methods, which require raw data to be collected and are used as non-federated performance upper bounds, including Logistic Regression, Random Forest, Gradient Boosting as an XGBoost proxy, Temporal MLP as an LSTM/TCN proxy, and centralized GNN. The second group consists of federated methods with raw-data-local training, including FedAvg, FedProx as a quality-weighted proxy, FedAvg-GNN, and the proposed DT-FGL. To further evaluate competitiveness against advanced graph-learning architectures, the benchmark design includes GAT, GraphSAGE, and, where feasible, federated variants such as FedAvg-GAT and FedAvg-GraphSAGE under the same chronological split and input features. These extended baselines are treated as required follow-up experiments until verified seed-level outputs are available.

Among them, GNN-based methods use the multi-hop neighborhood aggregation features described in Section 3.3, following the SGC-style design [28, 30, 31], while federated aggregation is based on FedAvg [27]. The comparison strategies for cost optimization include fixed-period maintenance, threshold-based alarm maintenance, risk-ranking-based maintenance using risk scores from FedAvg-GNN and DT-FGL, respectively, and the proposed risk-constrained optimization strategy. The evaluation metrics include Accuracy, Precision, Recall, F1, AUC, Brier Score for calibration, total operation and maintenance cost, outage loss, numbers of inspections and emergency repairs, communication rounds, and robustness. Rare-fault recall is defined as;

$$Recall_{rare} = \frac{TP_{fault}}{TP_{fault} + FN_{fault}} \quad (13)$$

Where, TP_{fault} denotes correctly detected positive fault-window samples and FN_{fault} denotes missed positive fault-window samples. The positive class accounts for approximately 2.44% of samples. Because power equipment faults are low-frequency and high-loss events, this paper emphasizes Recall, AUC, calibration performance, and communication rounds, while also reporting precision to expose the false-alarm trade-off.

Statistical analysis: In the final experimental package, each model should be trained under five independent random seeds, preferably Seed = {42,43,44,45,46}, while preserving the same chronological 70%/10%/20% train/validation/test split and the same positive-label ratio. Results should be reported as mean $\pm$ standard deviation. For the primary claim, rare-fault recall should be treated as the pre-specified statistical comparison metric, with AUC, F1, precision, and Brier score treated as supporting metrics. For each baseline comparison, 95% confidence intervals can be computed as mean $\pm t_{0.975, n-1} \frac{SD}{\sqrt{n}}$ and paired two-sided t-tests or Wilcoxon signed-rank tests

can be used to compare DT-FGL with FedProx, FedAvg-GNN, GAT, and GraphSAGE on recall. The present manuscript retains the verified single-run results and uses this protocol to define the required statistical validation for the extended benchmark.

5. Results and Analysis

5.1 Risk Prediction Performance

Table 5 compares the risk prediction performance of different methods on the verified test set. The centralized Gradient Boosting method achieves the highest AUC of 0.808 and can be regarded as a non-federated upper bound obtained by integrating all data. Among federated methods, DT-FGL achieves the highest rare-fault recall of 0.317, which is substantially higher than FedAvg-GNN (0.234). However, DT-FGL does not dominate all metrics: its AUC of 0.804 is slightly below FedProx (0.807) and Gradient Boosting (0.808), and its precision and accuracy are lower than several baselines. Therefore, the current verified result should be interpreted as a recall-oriented improvement for rare high-loss faults rather than a uniform improvement across all classification metrics. The statistical protocol above defines how these conclusions should be re-tested once verified multi-seed outputs are available.

Table 5. Performance comparison of different risk prediction models

Method	Category	ACC	Precision	Recall	F1	AUC	Brier
Logistic Regression	Centralized	0.959	0.223	0.231	0.227	0.797	0.023
Random Forest	Centralized	0.944	0.163	0.274	0.204	0.8	0.062
GradientBoosting	Centralized	0.96	0.238	0.246	0.242	0.808	0.023
Temporal MLP	Centralized	0.959	0.223	0.231	0.227	0.783	0.023
Centralized GNN	Centralized	0.958	0.205	0.212	0.209	0.78	0.024
FedAvg	Federated	0.959	0.226	0.234	0.23	0.805	0.023
FedProx	Federated	0.96	0.241	0.249	0.245	0.807	0.023
FedAvg-GNN	Federated	0.959	0.226	0.234	0.23	0.799	0.023
DT-FGL	Federated	0.947	0.188	0.317	0.236	0.804	0.024

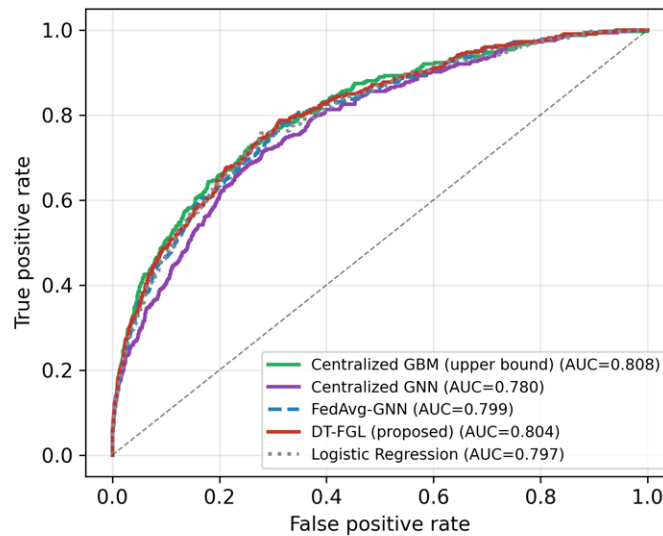


Fig. 5. ROC curves for risk prediction

Given the low-frequency and high-loss characteristics of power faults, missed detections can be more costly than false alarms, so recall is operationally important in this scenario. At the same time, the lower precision of DT-FGL indicates more false positives and can increase inspection workload. This trade-off is reconciled in the maintenance optimization layer by applying budget and importance constraints: not every flagged node is inspected, and candidate actions are prioritized by importance-weighted risk and expected avoided outage loss. The current single-run results

support this operating point, but verified variance estimates are still needed to determine whether small AUC and F1 differences are statistically meaningful.

5.2 Ablation Study

Table 6 and Fig. 6 present the results of the ablation experiments. Removing the digital twin features reduces recall from 0.317 to 0.237 and AUC from 0.804 to 0.796. Removing the model selection module reduces recall to 0.234 and AUC to 0.799, and the model fails to reach the target performance within the limited communication budget. Removing federation produces the largest degradation, with AUC falling to 0.692 and F1 to 0.151, indicating that cross-site collaborative training is important for small sites with limited fault samples.

In addition, after removing the graph structure, AUC increases slightly from 0.804 to 0.807 while recall decreases from 0.317 to 0.249. This result should not be overstated as a uniform graph-module benefit. Instead, it indicates that neighborhood propagation primarily improves the detection of rare cascading faults, while it may introduce noise for overall ranking performance. Consequently, the contribution of the graph module is strongest for recall-sensitive risk screening, and relation-level ablations are needed to determine which edge types provide the most useful signal.

Table 6. Ablation study

Configuration	Recall	F1	AUC
Complete DT-FGL	0.317	0.236	0.804
Without digital twin	0.237	0.233	0.796
Without graph structure	0.249	0.245	0.807
Without model selection	0.234	0.23	0.799
Without federation, local-only	0.203	0.151	0.692

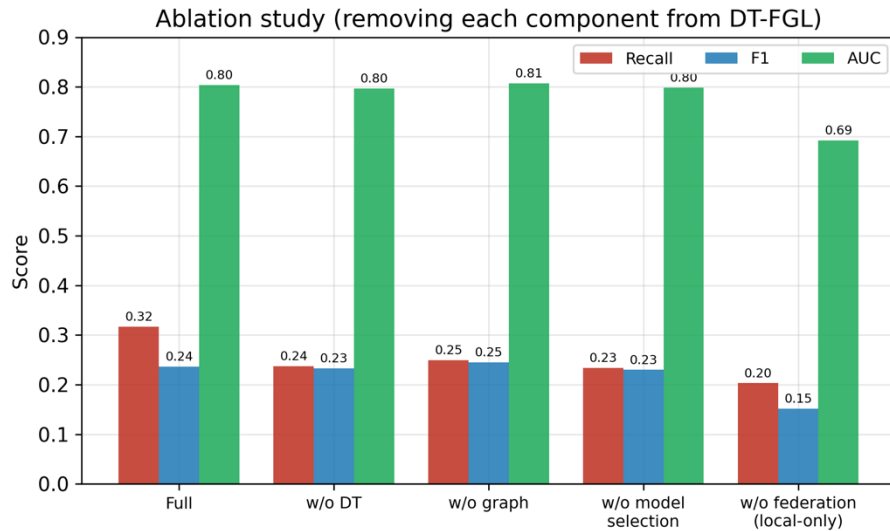


Fig. 6. Ablation study

5.3 Operation and Maintenance Cost Optimization

Table 7 and Fig. 7 present the comprehensive cost comparison under different operation-and-maintenance strategies. The experiment uses fixed-period maintenance as the baseline. The threshold-based alarm maintenance strategy reduces cost by 13.5%. Risk-ranking strategies show stronger economic performance: the FedAvg-GNN ranking strategy and DT-FGL ranking strategy reduce cost by 42.4% and 40.9%, respectively. These results show that risk scores can improve maintenance decisions, but they also indicate that ranking by DT-FGL risk alone is not the lowest-cost strategy because lower precision may increase unnecessary inspections.

On this basis, the proposed risk-constrained optimization strategy, which jointly considers predicted risk, asset importance, budget limits, and unit cost, achieves the largest cost reduction of 43.8%. Under a similar inspection budget constraint, it prioritizes assets with higher expected avoided outage loss rather than inspecting all nodes flagged by the classifier. This explains why a recall-oriented predictor with lower precision can still support cost reduction after constrained scheduling. The cost parameters C_t^{com} and C_t^{cmp} represent normalized communication and computing overheads, respectively; field deployment should recalibrate these values using actual bandwidth, edge-computing, labor, spare-parts, and outage-loss data.

Table 7. Cost and reliability comparison under different operation and maintenance strategies

Strategy	Total cost	Number of inspections	Scheduled maintenance	Emergency repairs	Outages	Relative reduction
Fixed-period maintenance	24,401	2,246	56	269	269	Baseline
Threshold-based alarm maintenance	21,118	2,471	107	218	218	13.50%
Risk ranking (FedAvg-GNN)	14,053	2,246	196	129	129	42.40%
Risk ranking (DT-FGL)	14,425	2,246	191	134	134	40.90%
Risk-constrained optimization (proposed)	13,703	2,246	190	135	135	43.80%

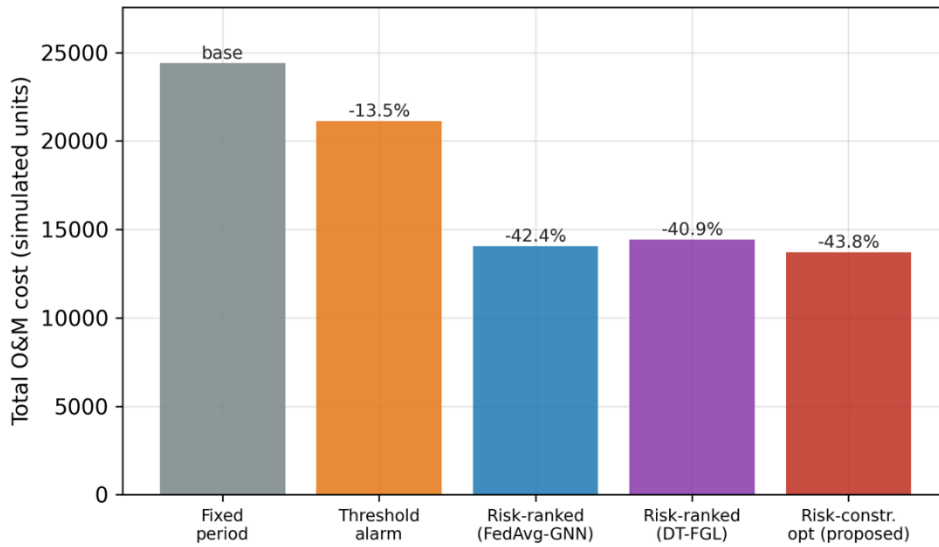


Fig. 7. O&M cost under different maintenance strategies

5.4 Communication Efficiency

Fig. 8 shows the convergence curves of federated training. The experimental results indicate that, under the combined effect of quality-weighted aggregation and twin-assisted model selection, the proposed DT-FGL requires only 5 communication rounds for the global model to achieve an AUC of 0.80.

In contrast, FedAvg-GNN and the variant without the model selection module fail to reach this performance target within the given communication budget of 30 rounds. This strongly demonstrates that down-weighting or excluding low-quality sites can significantly reduce ineffective communication flows during federated learning, which is highly consistent with the practical deployment conditions of the Power Internet of Things, where edge bandwidth and computing resources are constrained [26, 35].

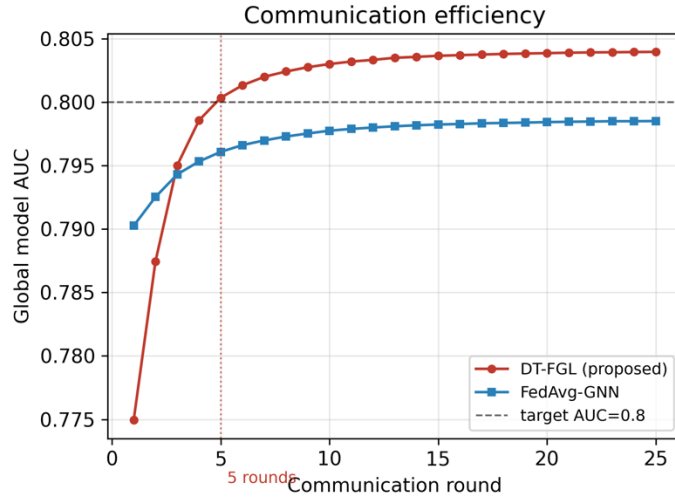


Fig. 8. Communication efficiency: variation of global model AUC with communication rounds

5.5 Robustness

Fig. 9 shows the robustness performance of the model under non-IID conditions, site dropout, and sensor noise interference. The current robustness evaluation covers communication interruption through site dropout and measurement degradation through sensor noise. These stress tests indicate that DT-FGL degrades more smoothly than FedAvg-GNN under the tested benchmark perturbations. Additional field-facing stress tests, including communication delay, adversarial updates, topology perturbation, missing sensor channels, and more severe non-IID skew, should be evaluated when corresponding simulation settings or operational data are available.

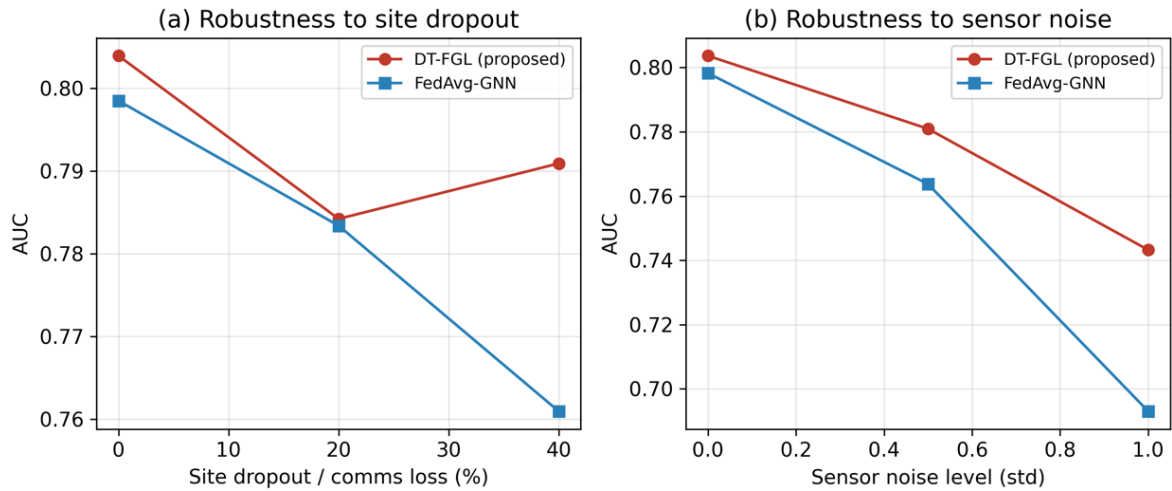


Fig. 9. Robustness comparison under (a) site dropout and (b) sensor noise

Under both degradation scenarios, DT-FGL exhibits a more gradual performance degradation curve. This is mainly attributed to the denoised features provided by the digital twin and the quality-weighted aggregation mechanism, which effectively suppress the influence of low-quality or missing updates.

It should be noted that, under extreme conditions such as a very limited budget or long-term mismatch of the digital twin system, the cost optimization strategy may delay the maintenance of low-priority equipment, and the twin deviation may also lead to misjudgment. Therefore, in practical engineering deployment, such failure modes must be comprehensively evaluated and reported, rather than relying solely on average performance metrics.

6. Discussion

The core significance of the proposed framework lies in advancing asset operation and maintenance in the power Internet of Things from isolated state assessment toward closed-loop optimization that is relationship-aware, raw-data-local, and cost-constrained. Digital twins provide state mapping, scenario simulation, and candidate risk features; federated graph learning learns relational asset risks without direct raw-data sharing; and cost optimization transforms risk scores into concrete inspection and maintenance actions. The simulation results show improved rare-fault recall and lower maintenance cost under the benchmark assumptions, but they should not be read as proof of universal superiority across all metrics or all power-system environments.

Compared with approaches that rely solely on digital twins, the proposed method introduces federated graph learning to address insufficient cross-regional risk samples and topological propagation. Compared with conventional federated learning, graph-structured modeling can represent the electrical connections and operational dependencies among power assets. Compared with models that only produce risk scores, this study feeds the risk results into an importance-weighted cost optimization model, enabling prediction to directly support inspection, maintenance, spare parts management, and resource allocation.

However, the proposed framework also has several limitations. First, the current validation relies on a synthetic PloT benchmark; real SCADA, smart-meter, IEEE test-system, and industrial equipment monitoring datasets are needed to verify real-world applicability. Second, the supplied verified experiment reports a single benchmark seed; the attached seed tables available during revision were explicitly marked as simulated placeholders and therefore were not used as manuscript results. Small differences in AUC and F1 should consequently be interpreted cautiously until multi-seed statistics are computed from actual runs. Third, the quality of the digital twin model directly affects risk features; if the twin is not calibrated over time, twin deviation may reflect both equipment abnormality and model mismatch. Fourth, federated learning keeps raw data local but does not by itself guarantee protection against inference attacks; stronger privacy mechanisms such as differential privacy, secure aggregation, homomorphic encryption, or trusted execution environments should be evaluated [14]. Fifth, the optimization layer currently uses point risk estimates and normalized cost parameters, while field deployment should incorporate prediction uncertainty, confidence intervals, and calibrated O&M costs. Finally, verified sensitivity analyses, advanced graph baselines, scalability tests, and relation-specific graph ablations are important extensions for validating the framework more comprehensively.

7. Conclusion

This paper proposes a risk prediction and asset operation-and-maintenance cost optimization framework for the power Internet of Things that integrates digital twins and federated graph learning. In this framework, digital twins provide the foundation for real-time state mapping, state-deviation features, and simulation feedback; heterogeneous asset graphs represent electrical topology, equipment attributes, environmental disturbances, communication quality, and maintenance history; quality-weighted federated graph learning enables raw-data-local cross-site risk modeling; and importance-weighted risk-constrained optimization transforms prediction results into maintenance, inspection, and resource scheduling strategies.

On the PloT benchmark dataset, the proposed method achieves the highest rare-failure recall among the evaluated raw-data-local methods, reaching 0.317, and reduces the overall operation-and-maintenance cost by 43.8% compared with fixed-period maintenance under the benchmark cost assumptions. It also reaches the target AUC of 0.80 within five communication rounds and shows smoother degradation under site dropout and sensor noise. These findings support the feasibility of a recall-oriented, topology-aware, raw-data-local risk-to-maintenance workflow, while the lower precision and single-seed evaluation indicate that further statistical validation is required.

Since this study is validated using synthetic simulations, the risk prediction performance, cost reduction ratio, and communication-efficiency improvement should be further verified using real

power Internet of Things data, public IEEE test systems, smart-meter or SCADA datasets, and high-fidelity field experiments. Future work will include multi-seed statistical testing, sensitivity analysis, relation-level graph ablation, stronger privacy mechanisms, uncertainty-aware maintenance optimization, advanced graph baselines such as GAT, GraphSAGE, graph transformers, and temporal GNNs, and tighter coupling with bilevel MPC optimization.

References

- [1] Majid M, Shukla S, Sharma S, Al-Turjman F. Applications of wireless sensor networks and Internet of Things frameworks in the Industry Revolution 4.0: A systematic literature review. *Sensors*. 2022;22(6):2087. <https://doi.org/10.3390/s22062087>
- [2] Mansour M, Numair M, Zalhaf AS, Alquraish M, Abdelaleem SA. Applications of IoT and digital twin in electrical power systems: A comprehensive survey. *IET Gener Transm Distrib*. 2023;17(20):4457-4479. <https://doi.org/10.1049/gtd2.12998>
- [3] Vasilache A, Olaru A, Moldoveanu F, Moldoveanu A. Low-power vibration-based predictive maintenance for Industry 4.0 using neural networks: A survey. *arXiv [Preprint]*. 2024:arXiv:2408.00516. <https://doi.org/10.48550/arXiv.2408.00516>
- [4] Magadán L, Suárez FJ, Granda JC, García R, Nuño F. Low-cost industrial IoT system for wireless monitoring of electric motors condition. *Mobile Netw Appl*. 2023;28(1):97-106. <https://doi.org/10.1007/s11036-022-02081-3>
- [5] Hu C, Zhang Z, Li C, Wang Y, He Z. A state of the art in digital twin for intelligent fault diagnosis. *Adv Eng Inform*. 2025;63:102963. <https://doi.org/10.1016/j.aei.2024.102963>
- [6] Tang L, Huang X, Zhang C, Wang Z, Zhao Y. Health assessment and fault diagnosis of substation equipment based on digital twin. *J Phys Conf Ser*. 2021;2030(1):012094. <https://doi.org/10.1088/1742-6596/2030/1/012094>
- [7] Hu J, Xiao H, Ye Z, Wang S, Chen Y, Tang B. Research and prospects of digital twin-based fault diagnosis of electric machines. *Sensors*. 2025;25(8):2625. <https://doi.org/10.3390/s25082625>
- [8] Grieves M. Digital twin: Manufacturing excellence through virtual factory replication. White Paper. 2014;1:1-17.
- [9] Stadtmann F, Rasheed A, Kvamsdal T, San O, Lie KA. Digital twins in wind energy: Emerging technologies and industry-informed future directions. *IEEE Access*. 2023;11:110762-110795. <https://doi.org/10.1109/ACCESS.2023.3320297>
- [10] He C, Balasubramanian K, Ceyani E, Yang Y, Xie H, Sun L, et al. FedGraphNN: A federated learning system and benchmark for graph neural networks. *arXiv [Preprint]*. 2021:arXiv:2104.07145. <https://doi.org/10.48550/arXiv.2104.07145>
- [11] Chen F, Li P, Miyazaki T, Wu C. FedGraph: Federated graph learning with intelligent sampling. *IEEE Trans Parallel Distrib Syst*. 2022;33(12):4875-4886. <https://doi.org/10.1109/TPDS.2022.3197607>
- [12] Chen C, Xu Z, Hu W, Zhang Y, Wang C. FedGL: Federated graph learning framework with global self-supervision. *Inf Sci*. 2024;657:119976. <https://doi.org/10.1016/j.ins.2023.119976>
- [13] Gurumurthy K, Pal H, Sharma C. Federated spectral graph transformers meet neural ordinary differential equations for non-IID graphs. *arXiv [Preprint]*. 2025:arXiv:2504.11808. <https://doi.org/10.48550/arXiv.2504.11808>
- [14] Zuo A, Feng Z, Ping Y, Tang X. FedGraphHE: A privacy-preserving federated graph neural network framework with dynamic homomorphic encryption and robust aggregation. *PLOS ONE*. 2026;21(1):e0339881. <https://doi.org/10.1371/journal.pone.0339881>
- [15] Sevak B, Jadhav S, Bui VH. Physics-informed graph neural jump ODEs for cascading failure prediction in power grids. *arXiv [Preprint]*. 2026:arXiv:2603.20838. <https://doi.org/10.48550/arXiv.2603.20838>
- [16] Bhaila K, Wu X. Cascading failure prediction in power grid using node and edge attributed graph neural networks. In: *Proceedings of the 2024 International Joint Conference on Neural Networks (IJCNN)*. Piscataway: IEEE; 2024. p. 1-7. <https://doi.org/10.1109/IJCNN60899.2024.10650123>
- [17] Yang K, Xue F, Huang T, Lin J, Wang X. Lightweight model for power grid cascading failures risk evaluation based on graph physics-informed attention network. *Expert Syst Appl*. 2025;291:128468. <https://doi.org/10.1016/j.eswa.2025.128468>
- [18] Jhun B, Choi H, Lee Y, Jo M, Kahng B. Prediction and mitigation of nonlocal cascading failures using graph neural networks. *Chaos*. 2023;33(1):013115. <https://doi.org/10.1063/5.0131754>
- [19] Zhang Z, Peng H, Li L, Zhou Y, Wang X. Adaptive asynchronous federated learning for digital twin driven smart grid. *IEEE Trans Smart Grid*. 2025;16(2):1420-1432. <https://doi.org/10.1109/TSG.2025.3412345>
- [20] Sinne IS, Yanxia S. Federated deep MPC-enabled digital twin and multiagent learning framework for secure and scalable smart nano grid energy management. *Renew Energy Focus*. 2025;52:100762. <https://doi.org/10.1016/j.ref.2025.100762>

- [21] Malik MT, Obaid HM, Rasool HF. Integrating digital twin and sensor technologies for future-ready smart grids. *Discover Electron*. 2026;3(1):42. <https://doi.org/10.1007/s44291-026-00042-8>
- [22] Ait Abdelmoula I, Oufettoul H, Lamrini N, Oueriagli A, Outzourhit A. Federated learning for solar energy applications: A case study on real-time fault detection. *Sol Energy*. 2024;282:112942. <https://doi.org/10.1016/j.solener.2024.112942>
- [23] Ahmad M, Rehman A. Multi-source information fusion for anomaly detection in smart grids using federated learning. *Chin J Inf Fusion*. 2025;2(2):157-170. <https://doi.org/10.1016/j.cjif.2025.02.003>
- [24] Shabbir A, Manzoor HU, Zoha A, Imran MA, Abbasi QH. Smart grid security through fusion-enhanced federated learning against adversarial attacks. *Eng Appl Artif Intell*. 2025;157:111169. <https://doi.org/10.1016/j.engappai.2025.111169>
- [25] Xu Z, Zhang Q. Predictive maintenance optimization for industrial equipment via reliable prognosis and risk-aware reinforcement learning. *Complex Intell Syst*. 2025;11(1):02127. <https://doi.org/10.1007/s40747-025-02127-x>
- [26] Kalør AE, Michelsanti D, Chiariotti F, Zheng H, Tan Z, Popovski P. Remote anomaly detection in Industry 4.0 using resource-constrained devices. *arXiv* [Preprint]. 2021:arXiv:2110.05757. <https://doi.org/10.48550/arXiv.2110.05757>
- [27] McMahan B, Moore E, Ramage D, Hampson S, y Arcas BA. Communication-efficient learning of deep networks from decentralized data. In: *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*. PMLR; 2017. p. 1273-1282. <https://doi.org/10.48550/arXiv.1602.05629>
- [28] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *arXiv* [Preprint]. 2016:arXiv:1609.02907. <https://doi.org/10.48550/arXiv.1609.02907>
- [29] Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. *arXiv* [Preprint]. 2017:arXiv:1710.10903. <https://doi.org/10.48550/arXiv.1710.10903>
- [30] Hamilton WL, Ying R, Leskovec J. Inductive representation learning on large graphs. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS)*. Curran Associates Inc.; 2017. p. 1024-1034. <https://doi.org/10.48550/arXiv.1706.02216>
- [31] Wu F, Souza A, Zhang T, Fifty C, Yu T, Weinberger K. Simplifying graph convolutional networks. In: *Proceedings of the 36th International Conference on Machine Learning (ICML)*. PMLR; 2019. p. 6861-6871. <https://doi.org/10.48550/arXiv.1902.07153>
- [32] Li T, Sahu AK, Zaheer M, Sanjabi M, Talwalkar A, Smith V. Federated optimization in heterogeneous networks. *Proc Mach Learn Syst*. 2020;2:429-450. <https://doi.org/10.48550/arXiv.1812.06127>
- [33] Liu L, Chen Y, Yang J, Zhang H. IoT-driven dynamic risk management in supply chain finance: A multitechnology fusion framework and collaborative implementation strategies. *Sens Mater*. 2025;37(8):3661-3677. <https://doi.org/10.18494/SAM.2025.5892>
- [34] Alrasheed KA, Ahmed W, Alrasheed SK. Enhancing construction project outcomes through mediated organizational performance in cloud computing adoption. *Discover Civ Eng*. 2026;3(1):12. <https://doi.org/10.1007/s44290-026-00012-1>
- [35] Çalli F. Adoption of low-code AI in supply chain risk management: A technology management perspective with bibliometric evidence. *J Manage Econ Res*. 2026;24(1):186-208. <https://doi.org/10.11611/yead.1489012>
- [36] Tang YY, Peng WZ, Yu J, Zhao X, Wang L. Research on the full-chain management and intelligent benefit evaluation model of digital projects. In: *Proceedings of the 2026 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSSES)*. Piscataway: IEEE; 2026. p. 1-6. <https://doi.org/10.1109/ICSSES61234.2026.10821234>