EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

# An Integrated Digital-Twin-Assisted Edge Collaboration and LSTM-GAT Model for Short-Term Power Load Forecasting in Industrial Parks

Ruilin Li[1*], Honghong Li[1*], Fengyang Long[2], Donghua Jiang[3], Jianhua Wu[1]

[1]Gongqing College of Nanchang University, Gongqingcheng 332020, China
[2]College of Water Conservancy, Jiangxi University of Water Resources and Electric Power, Nanchang 330099, China
[3]School of Information Engineering, Chang'an University, Xi'an 710064, China

## Abstract

INTRODUCTION: Manufacturing-oriented industrial parks are increasingly characterised by flexible production lines, energy-intensive equipment, and highly coupled production–energy processes. In such smart manufacturing environments, accurate short-term power load forecasting is essential for coordinating production scheduling, equipment operation, and low-carbon energy management. However, existing forecasting methods often overlook edge-side training efficiency and spatial correlations among multi-node power loads.

OBJECTIVES: This paper aims to develop an efficient and accurate short-term power load forecasting framework for Industrial Internet of Things-enabled industrial parks by integrating digital-twin-assisted edge collaborative training with temporal–spatial load modelling.

METHODS: An integrated digital-twin-assisted mobile edge computing–Industrial Internet of Things framework and an LSTM-GAT forecasting model, named L-Ga, are proposed. The digital twin layer mirrors the communication, computing, energy, and data-quality states of edge nodes to support node selection and collaborative model training. The L-Ga model combines long short-term memory networks and graph attention networks to capture temporal evolution and spatial coupling among park-level, feeder-level, utility-system, office-area, and key-equipment loads.

RESULTS: The digital-twin-assisted node-selection strategy reduced the weighted training cost by 21.63% compared with random node selection. The proposed L-Ga model achieved average mean absolute error, root mean square error, and mean absolute percentage error values of 2.29, 3.12, and 1.98%, respectively, outperforming the representative baseline models.

CONCLUSION: The proposed framework improves both edge-side collaborative training efficiency and short-term power load forecasting accuracy. It provides an effective solution for production–energy coordination, intelligent power distribution, and low-carbon energy management in Industrial Internet of Things-enabled manufacturing parks.

**Keywords:** Industrial Internet of Things; Digital Twin; Edge Computing; Industrial Park Power Distribution; Short-Term Power Load Forecasting; Energy Management; LSTM-GAT

*Corresponding author. Ruilin Li, Email: honghongli_85@163.com; Honghong Li, Email: lrl5211983126.com

Received on 23 July 2026, accepted on 14 September 2026, published on 02 November 2026

Copyright © 2026 Ruilin Li *et al.*, licensed to EAI. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetsis.14155

## 1. Introduction

With the development of smart manufacturing, industrial parks are gradually evolving from traditional production clusters into data-driven cyber-physical manufacturing systems. In manufacturing-oriented industrial parks, electricity consumption is strongly affected by workshop production rhythms, equipment start–stop behaviors, utility-system operation, and process-level scheduling. With the rapid deployment of Industrial Internet of Things (IIoT) devices, smart meters, and edge sensing terminals, industrial parks are generating large volumes of high-frequency electricity-consumption data [1-2]. Accurate short-term power load forecasting has become increasingly important for industrial park energy management, as it directly supports demand response, economic dispatch, peak shaving, equipment scheduling, and operational risk warning [3]. Existing studies have shown that short-term load forecasting performance is closely related to data granularity, feature construction, and scenario adaptability, especially in complex industrial environments with strong nonlinear fluctuations [4-5]. Traditional statistical and machine-learning methods provide useful forecasting baselines, while deep learning models have significantly improved short-term load prediction [6]. In particular, Long Short-Term Memory (LSTM) networks are effective in capturing temporal dependencies in load sequences, and attention mechanisms further enhance the extraction of critical dynamic information [7-8]. However, in industrial park scenarios, multiple feeders, utility subsystems, and equipment clusters often exhibit heterogeneous yet correlated load patterns [9]. Therefore, models that rely only on temporal sequence learning may still be insufficient for accurately representing the spatial interactions among multiple energy-consuming nodes. To better characterize such spatial dependence, recent studies have introduced graph-based forecasting methods [10-11]. Spatial-temporal graph neural networks and graph convolutional approaches have shown clear advantages in modeling non-Euclidean relationships among electricity nodes and improving short-term load forecasting accuracy. At the same time, digital twin technologies are being increasingly applied in industrial energy systems for real-time state mapping, process optimization, and intelligent decision support [12]. Nevertheless, the integration of digital-twin-assisted edge collaboration with spatio-temporal deep forecasting for industrial park short-term power load forecasting remains insufficiently explored [13]. Motivated by these observations, this study proposes an integrated digital-twin-assisted MEC-IIoT and LSTM-GAT framework for short-term power load forecasting in industrial parks. The proposed framework operates as a closed collaborative loop rather than as two independent modules. At each collaboration round, the digital twin layer synchronizes the communication, computing, energy, task-queue, and data-quality states of the candidate edge nodes. These virtual states are then used by the edge collaboration layer to determine the participating node set. Only the selected nodes contribute their current load sequences and local training updates to the collaborative training process, so the node-selection result directly affects the quality, timeliness, and stability of the data used for updating the forecasting model. The returned updates are aggregated at the MEC server to obtain the updated L-Ga model, and the validation loss is used to determine whether collaborative training should continue or terminatead back to the digital twin layer for the next collaboration round. Therefore, the digital-twin-assisted

mechanism affects forecasting performance through the selection and coordination of training participants, whereas L-Ga serves as the temporal-spatial forecasting engine. Together, they form an integrated state-synchronization, node-selection, collaborative-training, and load-forecasting workflow.

## 2. Digital-Twin-Assisted MEC-IIoT Collaborative Training Model for Industrial Park Short-Term Power Load Forecasting

The proposed MEC-DT-IIoT framework establishes a closed-loop interaction among physical sensing, digital-twin synchronization, edge collaboration, and L-Ga-based forecasting. The studied industrial park adopts a typical 10 kV/0.4 kV distribution structure, with monitoring nodes deployed at incoming cabinets, transformer outlets, workshop feeders, utility systems, office areas, and key-equipment terminals. These nodes continuously collect multi-source load and operating-state information and transmit the available state information to the digital twin layer.

As illustrated in Figure 1, the physical IIoT layer first provides real-time load observations and edge-node operating states. The digital twin layer maintains virtual representations of the candidate edge nodes and synchronizes their communication, computing, energy, task-queue, and data-quality states using the latest telemetry. Based on these synchronized states, the MEC edge collaboration layer evaluates the candidate nodes and determines the participating node set for the current collaboration round. The selected nodes provide the load sequences used for local model training and return their model updates to the MEC server, whereas unselected nodes do not participate in the parameter update of that round. Consequently, node selection changes not only the communication and computation cost of collaborative training but also the timeliness and reliability of the information entering the forecasting model.

The MEC server subsequently aggregates the updates returned by the participating nodes and updates the parameters of the L-Ga forecasting model. The validation loss is monitored during training, and the collaborative training procedure is terminated when the predefined early-stopping condition is satisfied. The resulting L-Ga model uses LSTM to extract temporal dependencies from the multi-node load sequences and GAT to aggregate spatial dependencies among electrically or statistically related monitoring nodes. The updated model then generates the one-step-ahead short-term load forecast. The latest physical-node states and training outcomes are returned to the digital twin layer, where the virtual node states are refreshed before the next collaboration round. This iterative procedure connects digital-twin-assisted resource coordination directly with temporal-spatial load forecasting rather than treating them as two isolated processes.

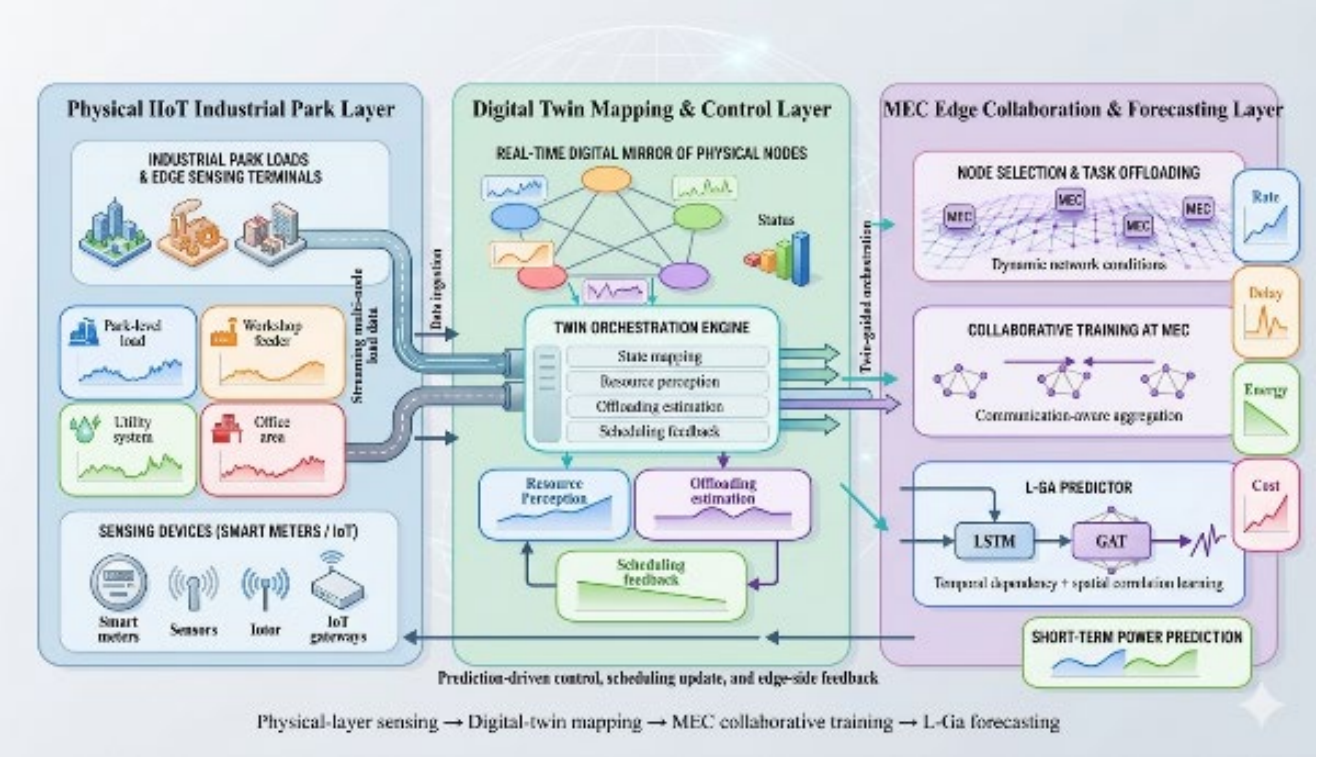



**Figure 1.** Digital-twin-assisted closed-loop collaborative framework for short-term power load forecasting in MEC-IIoT industrial parks

Algorithm 1. Unified digital-twin-assisted collaborative training and L-Ga forecasting workflow

Input: Candidate edge-node set V; historical multi-node load data; physical distribution topology; initial L-Ga model parameters; digital-twin states.

Output: Updated L-Ga model and one-step-ahead load forecasting results.

Step 1: Acquire the latest load measurements and operating states from the physical IIoT monitoring nodes.

Step 2: Synchronize the corresponding digital twins and update the communication, computing, energy, task-queue, and data-quality states of all candidate nodes.

Step 3: Evaluate the candidate nodes using the digital-twin-assisted node-selection criterion and determine the participating node set for the current collaboration round.

Step 4: Construct training samples from the latest load sequences available at the selected nodes and perform local L-Ga model updating.

Step 5: Return the local model updates from the participating nodes to the MEC server and perform collaborative aggregation.

Step 6: Update the global L-Ga model and evaluate its validation loss.

Step 7: If the early-stopping condition is not satisfied, return to Step 1 and start the next state-synchronization and collaborative-training round.

Step 8: If convergence is reached, use the updated L-Ga model to perform one-step-ahead short-term power load forecasting.

Step 9: Feed the latest node, training, and forecasting states back to the digital twin layer for subsequent synchronization and collaborative decision-making.

## 2.1 Model Algorithm

Following the unified workflow in Algorithm 1, the digital-twin-assisted node-selection stage requires quantitative descriptions of communication latency, computation latency,

energy consumption, and physical–virtual state deviation. These quantities are formulated below and subsequently incorporated into the node-selection objective.

Based on Shannon's theorem, the SIR and transmission rate for task offloading are given in equation (1)-(2), where B denotes channel bandwidth. The interference-related SIR is expressed by Equation (3). Local computation latency and its digital twin deviation are shown in equation (4)-(5). MEC computation latency and its deviation are presented in equation (6)-(7). The total task latency of the IIoT device is summarized in equation (8):

$$\gamma_{m,n} = \frac{\Omega_{m,n} p_m h_{m,n}}{\sum_{l \in \mathrm{M},\, l \neq m} \Omega_{l,n} p_l h_{l,n} + \sigma^2} \quad (1)$$

$$R_{m,n}^{uplink} = \mathrm{B} \log_2 \left(1 + \gamma_{m,n}\right) \quad (2)$$

$$\gamma_m^e = \frac{p_m h_{me}}{\sum_{k \in \mathrm{M}\ k \neq m} p_k h_{ke} + \sigma^2} \quad (3)$$

$$\tilde{T}_m \left(\alpha_m\right) = \frac{\alpha_m C_m}{f_m} \quad (4)$$

$$\Delta T_m \left(\alpha_m\right) = \frac{\alpha_m C_m \tilde{f}_m}{f_m \left(f_m - \tilde{f}_m\right)} \quad (5)$$

$$\tilde{T}_{m,n} \left(\Omega_{m,n}, \beta_{m,n}\right) = \frac{\Omega_{m,n} \beta_{m,n} C_m}{f_n} \quad (6)$$

$$\Delta T_{m,n} \left(\Omega_{m,n}, \beta_{m,n}\right) = \frac{\Omega_{m,n} \beta_{m,n} C_m \tilde{f}_n}{f_n \left(f_n - \tilde{f}_n\right)} \quad (7)$$

$$\begin{aligned} T_m^{\mathrm{tot}} \left(\Omega, \alpha_m, \beta_{m,n}, \mathbf{p}\right) &= \max_{n \in \mathrm{N}} T_{m,n}^{\mathrm{com}} + T_m^{\mathrm{local}} + \max_{n \in \mathrm{N}} T_{m,n}^{\mathrm{edge}} \\ &= \max_{n \in \mathrm{N}} \frac{\Omega_{m,n} \beta_{m,n} D_m}{R_{m,n}^{\mathrm{uplink}}} + \frac{\alpha_m C_m}{f_m - \tilde{f}_m} + \max_{n \in \mathrm{N}} \frac{\Omega_{m,n} \beta_{m,n} C_m}{f_n - \tilde{f}_n} \end{aligned} \quad (8)$$

The total latency of node i consists of local computing latency, uplink transmission latency, and MEC-side computing latency:

$$T_i = \max \left\{ \frac{(1-\rho_i) D_i C_i}{f_i^{\mathrm{loc}}}, \frac{\rho_i D_i}{r_i} + \frac{\rho_i D_i C_i}{f_i^{\mathrm{mec}}} \right\} \quad (9)$$

The corresponding energy consumption is calculated as:

$$E_i = \kappa_i \left(f_i^{\mathrm{loc}}\right)^2 (1-\rho_i) D_i C_i + \frac{p_i \rho_i D_i}{r_i} \quad (10)$$

To reflect the consistency between the physical node and its digital twin, the twin deviation is expressed as:

$$\Delta_i = \lambda_T \frac{|T_i - \hat{T}_i|}{T_i} + \lambda_E \frac{|E_i - \hat{E}_i|}{E_i} + \lambda_Q (1 - q_i) \quad (11)$$

To make the digital-twin-assisted node-selection procedure mathematically reproducible, a binary decision variable $x_i(r)$ is introduced for candidate node i in collaboration round r, where $x^{(r)}_i=1$ indicates that node i participates in the current collaborative training round and $x^{(r)}_i=0$ otherwise. For each candidate node, four indicators are considered: total latency $T_i$, energy consumption $E_i$, data-quality score $q_i$, and physical–virtual twin deviation $\Delta_i$. Because these indicators have different physical dimensions and numerical ranges, min–max normalization is performed within the candidate-node set:

$$\tilde{z}_i = \frac{z_i - \min_{j \in V} z_j}{\max_{j \in V} z_j - \min_{j \in V} z_j + \text{ò}} \quad (12)$$

where $z_i$ represents $T_i$, $E_i$, or $\Delta_i$, V is the candidate-node set, and ε is a small positive constant preventing division by zero. The data-quality score is a beneficial indicator and is normalized in the same manner; therefore, $1-q_i$ is used as its cost contribution.The weighted collaborative cost of candidate node i is consequently defined as:

$$C_i = \lambda_T \tilde{T}_i + \lambda_E \tilde{E}_i + \lambda_q (1 - \tilde{q}_i) + \lambda_\Delta \tilde{\Delta}_i \quad (13)$$

where $\lambda_T$, $\lambda_E$, $\lambda_q$, and $\lambda_\Delta$denote the weights assigned to latency, energy consumption, data quality, and twin deviation, respectively, and satisfy:

$$\lambda_T + \lambda_E + \lambda_q + \lambda_\Delta = 1 \quad (14)$$

Equal weights are adopted in the present implementation, i.e., $\lambda_T=\lambda_E=\lambda_q=\lambda_\Delta=0.25$, to avoid introducing an artificial preference toward any individual resource indicator. The same weighting scheme is maintained for all node-selection strategies. The node-selection problem in collaboration round r is formulated as:

$$\min_{x^{(r)}} J^{(r)} = \sum_{i \in V} x_i^{(r)} C_i^{(r)} \quad (15)$$

$$\sum_{i \in V} x_i^{(r)} = M, \quad x_i^{(r)} \in \{0,1\} \quad (16)$$

where M denotes the required number of participating nodes. In the present experiments, the candidate set contains 15 edge nodes and M=8 nodes are selected in each collaboration round. After the digital-twin states are updated, the cost $C_i$ is calculated for every eligible candidate node. The nodes are then ranked in ascending order of $C_i$, and the first M nodes constitute the participating set. This ranking-based solution is computationally lightweight and therefore suitable for repeated edge-side decisions during collaborative training.

## 2.2 Physical-Virtual Synchronization and Digital-Twin Correction

The digital twin in the proposed framework is not used merely as a passive monitoring interface. Each physical edge node is associated with a persistent virtual counterpart that maintains its communication, computing, energy, queue, and data-quality states throughout the collaborative-training process [14]. Accordingly, the physical–virtual mapping can be represented by a state vector:

$$\mathbf{s}_i^{(r)} = [R_i^{(r)}, f_i^{(r)}, E_i^{(r)}, Q_i^{(r)}, q_i^{(r)}]^T \quad (17)$$

where $R_i$ denotes communication capability, $f_i$ denotes available computing capability, $E_i$ denotes energy-related state, $Q_i$ represents the task-queue state, and $q_i$ denotes data quality.

At the beginning of each collaboration round, telemetry from the physical IIoT node is used to refresh its virtual counterpart. The digital twin first predicts the short-term node state from the previous virtual state and subsequently corrects this prediction when new physical observations become available. The discrepancy between the physical and virtual states is quantified through the twin-deviation term defined in Equation (11). When the deviation increases, the latest physical observation receives greater importance in the state update, thereby preventing accumulated virtual-state errors from propagating into subsequent node-selection decisions.

Four operations therefore distinguish the proposed digital twin from conventional node monitoring. First, persistent physical-virtual mapping maintains a one-to-one representation of each edge node rather than storing isolated measurements. Second, state synchronization continuously updates the virtual state according to newly acquired telemetry. Third, deviation-based correction explicitly evaluates the consistency between the predicted virtual state and its physical counterpart. Fourth, decision feedback feeds the corrected virtual states into the node-selection and collaborative-training processes, and the resulting training state is used again in the next synchronization cycle. Thus, the digital twin forms a closed physical-virtual-decision loop rather than serving solely as a state-observation layer.

## 2.3 Training cost comparison

Three node-selection strategies were compared: Random Selection randomly selects M nodes; Bnq Selection selects nodes with the highest uplink rate and channel quality; and Digital Twin Selection selects nodes according to predicted latency, energy consumption, data quality, and twin deviation. Time cost, energy cost, and weighted training cost were used as evaluation metrics. The experiment involved 30 training rounds, 15 candidate edge nodes, and 8 selected

nodes per round, with five low-rate nodes to simulate unstable communication conditions.

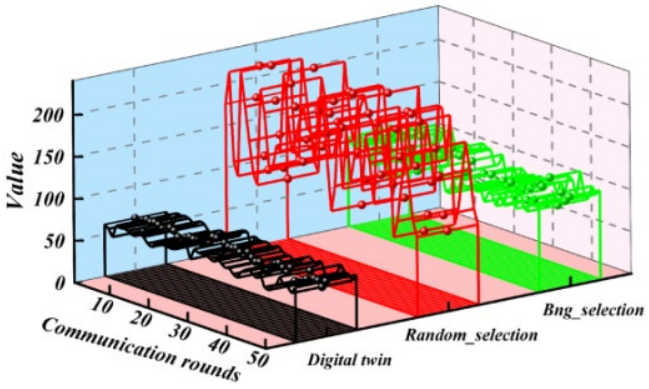



**Figure 2.** Time cost comparison of different node-selection strategies Unit: s

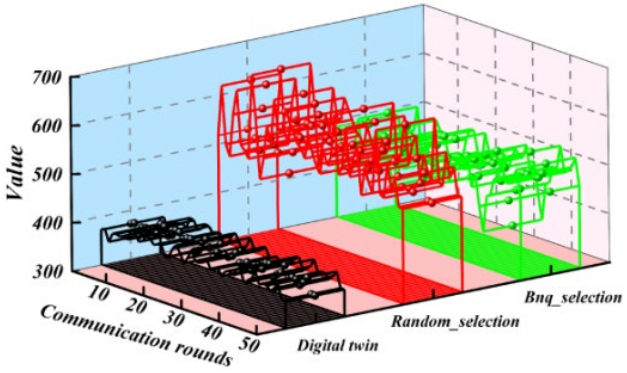



**Figure 3.** Energy cost comparison of different node-selection strategies under the same collaborative training setting Unit: J

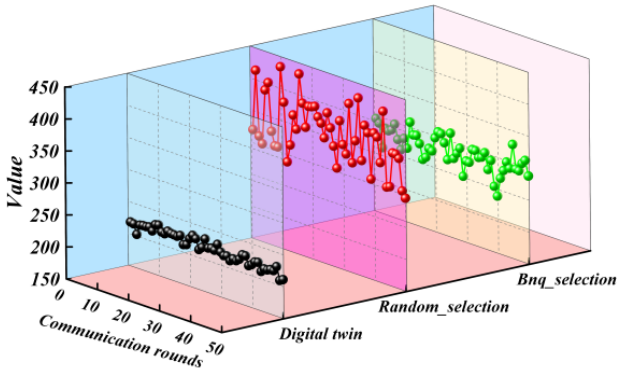



**Figure 4.** Weighted training cost comparison of different node-selection strategies

Figures 2–4 show that Digital Twin Selection achieves the lowest and most stable time, energy, and weighted training costs because it jointly considers communication, computing, and data-quality states. Table 1 further reports the mean and standard deviation of 10 repeated experiments. Compared with Random Selection, Digital Twin Selection reduces the weighted training cost by 21.63%.

Table 1. Statistical comparison of collaborative training costs under different node-selection strategies

| Strategy | Time cost (s) | Energy cost (J) | Weighted training cost | Reduction in weighted cost vs. Random |
|---|---|---|---|---|
| Random Selection | 7.18 ± 0.61 | 15.74 ± 1.12 | 0.490 ± 0.041 | — |
| Bnq Selection | 6.24 ± 0.43 | 14.23 ± 0.95 | 0.426 ± 0.031 | 13.06% |
| Digital Twin Selection | 5.82 ± 0.37 | 13.46 ± 0.84 | 0.384 ± 0.022 | 21.63% |

The comprehensive comparison in Figure 5 further confirms that the MEC-DT-IIoT strategy provides more stable edge-side collaborative support for short-term power load forecasting.

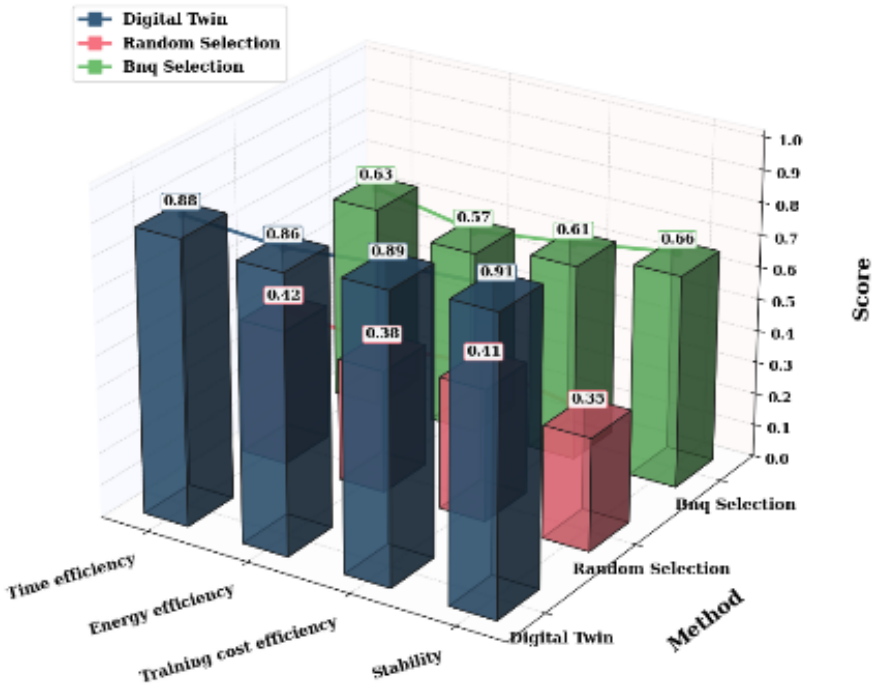



**Figure 5.** Comprehensive collaborative performance comparison under different node-selection strategies

## 3. Short-term power load forecasting based on the L-Ga model

The proposed L-Ga model combines LSTM and GAT to model temporal evolution and spatial dependence in industrial-park power loads. Its input is $X_{t-W+1:t} \in R^{W\times N\times F}$, where W=100, N denotes the number of monitoring nodes, and F denotes the input-feature dimension. LSTM first extracts temporal representations independently from the node-wise load sequences, after which GAT aggregates information from electrically or statistically related nodes.

To avoid representing spatial dependence using a single information source, a hybrid graph is constructed from both the physical distribution topology and historical load correlation. The topology adjacency matrix $A^{topo}$ represents the physical relationship among monitoring nodes. An element $A_{ij}^{topo}=1$ when nodes i and j have a direct upstream–downstream or feeder-level electrical relationship in the industrial-park distribution structure, and $A_{ij}^{topo}=0$ otherwise. The correlation adjacency matrix $A_{corr}$ is calculated from the absolute Pearson correlation coefficient between the historical load series of nodes i and j using training-set data only, thereby preventing test information from entering graph construction.

Before fusion, self-connections are added and the two adjacency matrices are separately normalized using symmetric degree normalization:

$$\hat{A} = D^{-1/2}(A+I)D^{-1/2} \qquad (18)$$

$$A = \mu\hat{A}^{topo} + (1-\mu)\hat{A}^{corr} \qquad (19)$$

where $\mu\in[0,1]$ controls the relative contribution of physical topology and statistical load correlation. The coefficient is selected exclusively on the validation set by grid search and is fixed before test evaluation. In this formulation, $\mu=1$ corresponds to a topology-only graph, whereas $\mu=0$

corresponds to a correlation-only graph. Therefore, the two single-source graphs are naturally included as the boundary cases of the same graph-construction procedure. This hybrid formulation retains physically meaningful distribution relationships while allowing the model to capture statistically coupled loads that are not directly adjacent in the electrical topology.

The GAT output is finally fed into a fully connected layer to generate simultaneous one-step-ahead predictions for the monitored nodes. For node i, the attention coefficient associated with neighboring node j is calculated as:

$$e_{ij} = \mathrm{LeakyReLU}\left(a^T \left[Wh_i \square Wh_j\right]\right), \quad \alpha_{ij} = \frac{\exp\left(e_{ij}\right)}{\sum_{k \in \mathbf{N}_i} \exp\left(e_{ik}\right)} \tag{20}$$

## 3.1 Data description and preprocessing

The dataset was collected from the energy-management platform and edge monitoring terminals of an anonymized manufacturing-oriented industrial park. It covers 90 consecutive operating days at a 15-min sampling interval and contains 15 monitoring nodes, corresponding to 129,600 node-time samples. The nodes cover five functional categories: park-level total load, workshop feeders, utility systems, office and ancillary areas, and key-equipment clusters. Absolute node identities and detailed electrical addresses are anonymized because of industrial data-security requirements.

The forecasting target is active power. The five categories exhibit different operating characteristics: park-level load reflects the aggregated production rhythm; workshop feeders are strongly affected by production start–stop schedules; utility-system loads show comparatively continuous operation; office and ancillary loads exhibit pronounced day–night periodicity; and key-equipment clusters contain relatively abrupt variations associated with equipment switching. These heterogeneous patterns provide the motivation for jointly learning temporal evolution and inter-node spatial dependence.

The forecasting problem is formulated as simultaneous one-step-ahead prediction for all monitored nodes. With the 15-min sampling interval, the model predicts the load at t+1, i.e., 15 min ahead, using the preceding W=100 observations. In the present experiment, historical active-power measurements constitute the direct forecasting input. Communication, computation, energy, and twin-deviation variables are used by the digital-twin collaboration mechanism for participant selection and are not directly concatenated into the L-Ga forecasting tensor. Exogenous variables such as weather, electricity price, and detailed production schedules are not included in the current forecasting experiment.

The 90-day sequence is divided chronologically into training, validation, and testing sets at a ratio of 70:15:15; random shuffling across these subsets is not performed. Short missing segments are filled by linear interpolation within the corresponding data subset, whereas longer missing segments are reconstructed using historical adjacent-day profiles without accessing future test samples. Outliers are identified using the Hampel filter and the three-sigma criterion. All normalization parameters, including the minimum and maximum values used for min–max scaling, are calculated only from the training set and subsequently applied unchanged to the validation and test sets. Similarly, the historical load-correlation matrix used for constructing $A^{corr}$ is calculated exclusively from the training set. These procedures prevent information from the validation or test period from leaking into model fitting, preprocessing, or graph construction.

## 3.2 Model Configuration and Training Protocol

The L-Ga model is implemented using an LSTM temporal encoder followed by a two-layer GAT spatial encoder. The hidden feature dimension is 64. The GAT component uses four attention heads and a dropout rate of 0.20. A batch size

of 64 is used for neural-network training. All neural models are optimized using Adam with an initial learning rate of 0.001 and a maximum of 200 epochs. Early stopping is applied according to validation loss with a patience of 20 epochs, and the model parameters corresponding to the lowest validation loss are retained for test evaluation. A fixed random seed of 42 is used for initialization and data loading to improve reproducibility.

All neural-network baselines are trained under the same software environment and on the same GPU workstation so that differences in computational environment do not affect the model comparison. Unless a model-specific parameter is explicitly stated otherwise, the same data partition, batch-processing procedure, validation criterion, and early-stopping strategy are used throughout the comparative experiments.

## 3.3 Sensitivity analysis

The effects of the number of GAT layers and the temporal window size were evaluated using a normalized forecasting score derived from MAE, RMSE, and MAPE. As shown in Figure 6, the best overall performance is obtained with two GAT layers and a window size of W=100.

Because the sampling interval is 15 min, W=100 corresponds to 25 h of historical observations. This setting was intentionally selected to cover approximately one complete daily operating cycle while retaining an additional 1 h of context across the day boundary. Consequently, the input sequence contains both short-term load fluctuations associated with equipment switching and production scheduling and the dominant 24-h periodic pattern of the industrial park. A shorter window may not fully represent the daily cycle, whereas an excessively long window increases model complexity and introduces redundant historical information. The sensitivity results therefore indicate that W=100 provides an appropriate balance between short-term dynamics, daily periodicity, and computational complexity.

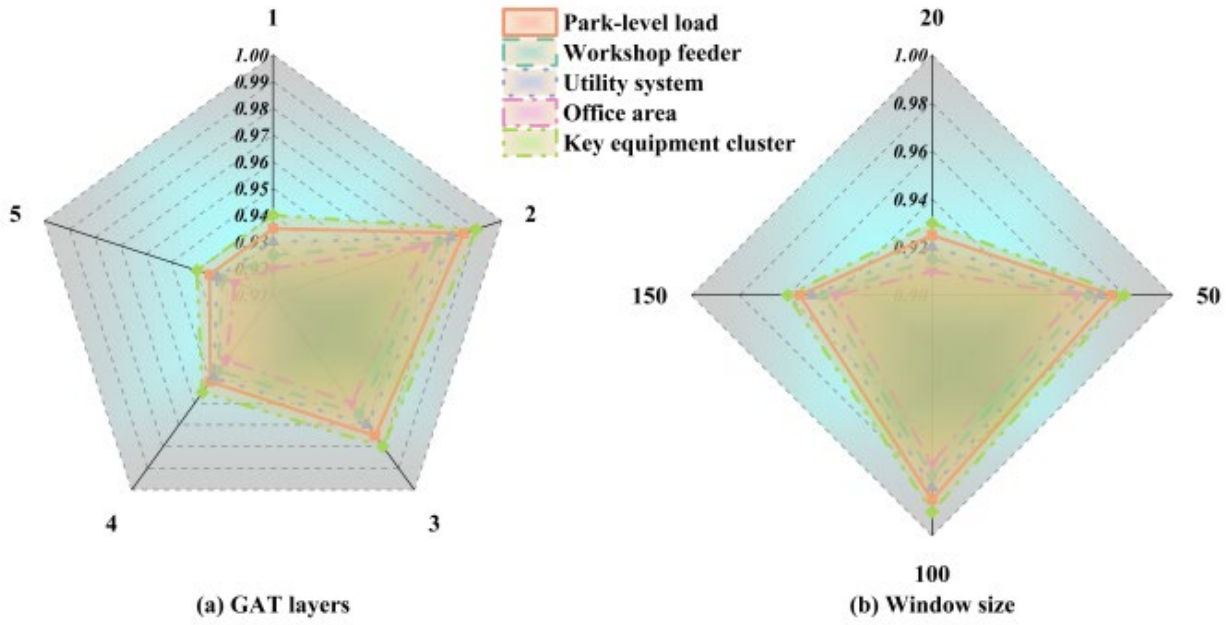



**Figure 6.** Sensitivity analysis of the L-Ga model in industrial park short-term power load forecasting

## 3.4 Comparative experimental results

Forecasting performance is evaluated using mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE):

$$MAE = \frac{1}{K}\sum_{k=1}^{K}|y_k - \hat{y}_k| \qquad (21)$$

$$RMSE = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(y_k - \hat{y}_k)^2} \qquad (22)$$

$$MAPE = \frac{100}{K}\sum_{k=1}^{K}\frac{|y_k - \hat{y}_k|}{\max(|y_k|, \varepsilon)} \qquad (23)$$

where $y_k$ and $\hat{y}_k$ denote the observed and predicted loads, respectively, and K is the number of evaluated samples. All predictions are inverse-transformed before metric calculation; therefore, MAE and RMSE retain the same physical unit as the measured active-power load, while MAPE is reported as a percentage. To prevent numerically unstable percentage errors during very-low-load periods, the denominator of MAPE is bounded by a small positive threshold ε. The threshold is determined from the training data only and remains unchanged during validation and testing.

To ensure a controlled comparison, all forecasting models use the same chronological training, validation, and testing subsets, the same 100-step input window, and the same one-step-ahead forecasting horizon. The comparison covers statistical, regression, recurrent, attention-based, and graph-based forecasting families, including Historical Average (HA), Multiple Linear Regression (MLR), ARIMA, LSTM, GRU, Transformer, GAT, and the proposed L-Ga model. Hyperparameters that require model-specific tuning are selected using the validation set only, while the test set is used exclusively for final performance evaluation.

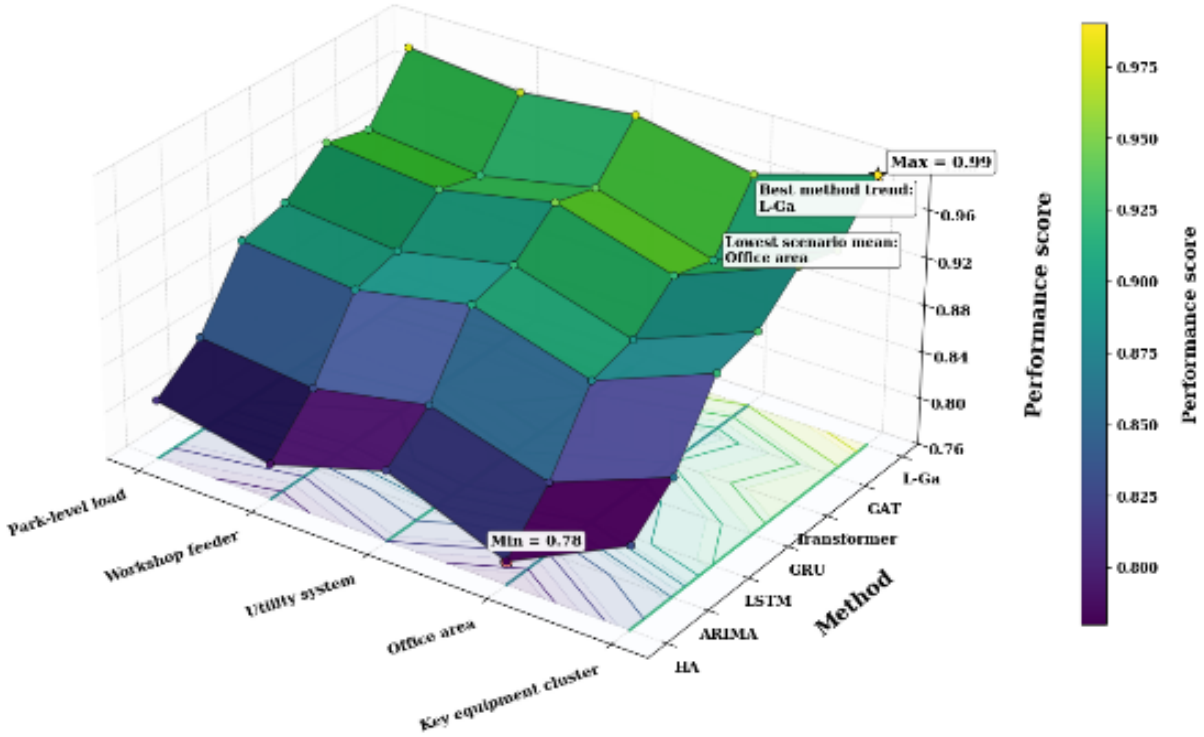



**Figure 7.** Comparison of forecasting results of L-Ga and baseline models under five industrial park load scenarios

Figure 7 compares the methods under five industrial-park load scenarios: park-level load, workshop feeder, utility system, office area, and key-equipment cluster. The proposed L-Ga model exhibits the strongest overall performance trend across these heterogeneous operating scenarios. The advantage is particularly attributable to the combination of temporal sequence modelling and inter-node graph aggregation, whereas temporal-only models cannot explicitly represent spatial dependence among the monitored electrical loads.

Nevertheless, the present comparison includes GAT as the principal graph-based baseline and therefore should not be interpreted as an exhaustive benchmark against all recently proposed spatio-temporal graph architectures. The present conclusions are consequently restricted to the representative statistical, recurrent, attention-based, and graph-based models evaluated under the same industrial-park dataset. Broader comparisons with specialised architectures such as STGCN, DCRNN, and Graph WaveNet will be considered when cross-model implementations and identical experimental outputs are available.In addition to the aggregate error metrics, Figure 7 reveals clear differences among the five operating scenarios. The office-area scenario exhibits the lowest overall normalized forecasting performance, reflecting the stronger day–night switching pattern and relatively low-load inactive periods. In contrast, aggregated park-level and workshop-load patterns provide stronger temporal continuity and inter-node information for prediction. The proposed L-Ga model maintains the strongest overall performance trend across the five scenarios, indicating that joint temporal-spatial modelling is particularly beneficial when rapid local variations coexist with correlated load changes elsewhere in the park.

The present revision package retains aggregate test metrics rather than the complete point-wise prediction sequence. Therefore, representative observed-versus-predicted profiles and residual time series are not reported here to avoid reconstructing curves from incomplete data. Such point-wise diagnostic plots will be included in subsequent extended validation when the corresponding raw prediction outputs are retained.

## 3.5 Ablation Study

To separate the contribution of the forecasting architecture from that of the edge-collaboration mechanism, the ablation experiment is reorganized into two controlled groups. As Table 2 show. In Panel A, the node-selection condition is fixed and only the forecasting architecture is changed. In Panel B, the LSTM-GAT forecasting backbone is fixed and only the node-selection strategy is changed.

Table 2. Performance comparison of ablation models

Panel A. Forecasting-architecture ablation under the same collaboration condition

| Model | MAE | RMSE | MAPE（%） |
|---|---|---|---|
| LSTM-only | 4.82 | 6.15 | 4.26 |
| GAT-only | 5.37 | 6.92 | 4.81 |
| LSTM-GAT | 3.16 | 4.03 | 2.79 |

Panel B. Node-selection ablation with the LSTM-GAT backbone fixed

| Node-selection strategy | MAE | RMSE | MAPE（%） |
|---|---|---|---|
| Random selection | 3.16 | 4.03 | 2.79 |

| Node-selection strategy | MAE | RMSE | MAPE（%） |
|---|---|---|---|
| Bnq selection | 2.76 | 3.58 | 2.43 |
| Digital twin selection | 2.29 | 3.12 | 1.98 |

Panel A shows that combining LSTM and GAT reduces forecasting errors relative to either temporal-only or graph-only modelling. Panel B further shows that, when the forecasting backbone is held fixed, changing the collaboration strategy from Random Selection to Bnq Selection and then to Digital Twin Selection progressively reduces the forecasting errors.

The value set of 3.16, 4.03, and 2.79 appears in both panels intentionally because it represents the shared control configuration: an LSTM-GAT forecasting backbone trained under Random Selection. In the previous version, this common control was presented as two apparently independent configurations, which created unnecessary ambiguity. The revised table therefore uses the same control point explicitly to connect the two ablation dimensions rather than treating them as separate experimental cases.

### 3.6 Robustness and Generalizability

Robustness is considered separately for the edge-collaboration mechanism and the forecasting model. For the collaboration experiment, each node-selection strategy was independently repeated ten times. As reported in Table 1, the time cost, energy cost, and weighted training cost are presented as mean ± standard deviation. Digital Twin Selection achieves not only lower mean costs but also smaller variability than Random Selection, indicating that the digital-twin-assisted strategy is less sensitive to the simulated unstable communication conditions.

For forecasting, the dataset consists of 90 consecutive operating days from one manufacturing-oriented industrial park. The chronological training–validation–testing partition avoids random temporal mixing and therefore provides an out-of-sample evaluation for the later operating period. However, this evaluation does not constitute external validation across different industrial parks, seasons, or distribution-network structures. The reported forecasting performance should therefore be interpreted within industrial parks having comparable monitoring density, operating patterns, and 15-min data resolution.

Consequently, the current study demonstrates within-park temporal robustness and edge-collaboration stability, whereas cross-park generalizability remains to be established. Future work will evaluate the framework using longer multi-season datasets and independent industrial parks and will investigate transfer learning or online adaptation for changes in production schedules and distribution-network topology.

## 4. Conclusion

From the perspective of manufacturing applications, the proposed framework can provide decision support for energy-aware production scheduling, peak-load regulation, and coordinated operation of workshop-level equipment and utility systems. Therefore, the method is suitable for smart manufacturing parks where load fluctuations are closely related to production processes and equipment operating states.

This paper proposed an integrated digital-twin-assisted MEC-IIoT and LSTM-GAT framework for short-term power load forecasting in industrial parks. The digital twin mechanism mirrors the communication, computing, energy, and data-quality states of edge nodes to support node selection and collaborative training, while the L-Ga model captures temporal evolution and spatial coupling among multi-node loads. Experimental results show that

digital-twin-assisted node selection reduces the weighted training cost by 21.63% compared with random selection. The proposed L-Ga model achieves average MAE, RMSE, and MAPE values of 2.29, 3.12, and 1.98%, respectively, outperforming representative baselines. The ablation results further confirm that LSTM-GAT improves temporal-spatial feature representation, while digital-twin-assisted collaboration improves training stability and data-update reliability. Future work will incorporate weather, production schedules, electricity prices, and online transfer learning to improve generalization.

The current validation is based on a 90-day dataset from one industrial park. Therefore, the reported results demonstrate the effectiveness of the proposed framework under the investigated manufacturing and monitoring conditions rather than universal cross-park generalizability. External validation using longer temporal periods, different industrial parks, and changing production regimes is required before broader deployment.

## Acknowledgments

This work was supported by:(1) Jiangxi Provincial Higher Education Institutions' Humanities and Social Sciences Research Project (Red Culture Special Project): Research on Jiangxi's Approach to Integrating Red Culture into Engineering Education Using Generative Artificial Intelligence (HSWH25025). (2) Jiujiang City Natural Science Foundation: Optimization of Subway Short-Term Passenger Flow Prediction Model Using Context-Aware Graph Transformer (CaGT) (2025_000504). (3) Jiangxi Provincial Department of Education: Research on Reforming Electrical Engineering Talent Training under the "Government-Enterprise-University-Association" Collaboration Mechanism (BKJG-2026-31-11).

## References

[1] Afrin S, Sabiha J R, Maliha K, et al. Industrial Internet of Things: Implementations, challenges, and potential solutions across various industries. Computers in Industry, 2025, 170: 104317.

[2] Dritsas E, Trigka M. A survey on the applications of cloud computing in the industrial internet of things. Big Data and Cognitive Computing, 2025, 9(2): 44.

[3] Timur O, Üstünel H Y. Short-term electric load forecasting for an industrial plant using machine learning-based algorithms. Energies, 2025, 18(5): 1144.

[4] Dakheel F, Çevik M. Optimizing smart grid load forecasting via a hybrid long short-term memory-XGBoost framework: Enhancing accuracy, robustness, and energy management. Energies, 2025, 18(11): 2842.

[5] Sulandari W, Yudho Y, Etik Z, et al. Hybrid Prophet-NAR model for short-term electricity load forecasting. IEEE Access, 2025, 13: 7637–7649.

[6] Sun C, Ning Y, Shen D, et al. Graph neural network-based short-term load forecasting with temporal convolution. Data Science and Engineering, 2024, 9: 113–132.

[7] Wei C, Pi D, Ping M, Zhang H. Short-term load forecasting using spatial-temporal embedding graph neural network. Electric Power Systems Research, 2023, 225: 109873.

[8] Wu J, Zuo R. Intelligent computation offloading based on digital twin-enabled 6G industrial IoT. Applied Sciences, 2024, 14(3): 1035.

[9] Aghazadeh Ardebili A, Zappatore M, Ramadan A I H A, Longo A, Ficarella A. Digital twins of smart energy systems: A systematic literature review on enablers, design, management and computational challenges. Energy Informatics, 2024, 7: 94.

[10] Keykele M P, Ravanshadnia M. Predicting residential CO2 emissions: Machine learning and deep learning approaches for short-term forecasting. International Journal of Sustainable Building Technology and Urban Development, 2025, 16(3): 373–387.

[11] Shah S A H, Ahmed U, Bilal M, et al. Improved electric load forecasting using quantile long short-term memory network with dual attention mechanism. Energy Reports, 2025, 13: 2343–2353.

[12] Billey A, Wuest T. Energy digital twins in smart manufacturing systems: A case study. Robotics and Computer-Integrated Manufacturing, 2024, 88: 102729.

[13] Kang Z. Physics-aware graph neural networks for real-time defect detection and environmental impact mitigation in industrial welding. Journal of Computing & Biomedical Informatics, 2026, 10(2).

[14] Kwon Y. Smart manufacturing innovation through the integration of digital twin and IoT energy management[J]. International Journal of Control, Automation and Systems, 2025, 23: 3427–3439. https://doi.org/10.1007/s12555-025-0046-1.