ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 3 - 2026

AST-CA-GNN: An Improved ASTGNN Integrating NWP and Topology Information for Spatiotemporal Distributed PV Voltage Prediction

Dezhao Ouyang[1], *Haiyang Zhang*[1,*], *Guibin Wang*[2], *Kang Zhou*[2], *Zhonghai Li*[1], *Yang Zhao*[1] *and Zhendong Tao*[1]

[1] Shenyang Power Supply Company, State Grid Liaoning Electric Power Supply Co., LTD, Shenyang 110004, China
[2] College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China

SUMMARY: *With the large-scale integration of distributed photovoltaic (PV) generation, power quality issues in distribution networks have become increasingly pronounced. In particular, overvoltage problems can cause severe damage to electrical equipment and even lead to system-wide outages. Therefore, the ability to anticipate voltage conditions at individual PV user sites is of critical importance, as it enables grid operators to formulate appropriate control strategies in advance and effectively reduce operational risks. Traditional forecasting methods face limitations in simultaneous modeling spatiotemporal interdependency and numeric weather predictions (NWP) correlations. To address these challenges, this study proposes a weather-aware spatiotemporal graph neural network for multi-step three-phase voltage prediction across multiple PV users, termed AST-CA-GNN. Building upon the ASTGNN framework, the model innovatively introduces cross-attention to enhance sensitivity to weather information, and constructs a dynamic graph convolution module based on prior spatial topology graph to capture time-varying spatial correlations between PV users. The experimental data are sourced from a 380V rural distribution system in China, including historical PV power observations, distribution network topology information, and NWP variables. Experimental results show that the proposed model outperforms other baseline models across all evaluation metrics in the three-phase voltage forecasting task, and achieves the best performance in the voltage limit violation prediction.*

KEYWORDS: *PV Voltage forecast; spatiotemporal forecast; cross-attention; graph convolution*

1 Introduction

The rapid worldwide advancement of renewable energy technologies has accelerated the integration of distributed photovoltaic (PV) systems into distribution networks. In 2024, China installed 278 GW of new solar PV capacity, representing a 28% increase over the previous year, among which 118 GW came from distributed PV units. However, the high variability and intermittency of PV generation introduce considerable operational risks, especially in networks with high PV penetration. These fluctuations may lead to overvoltage, reverse power flow, and deterioration of voltage quality[1, 2]. Maintaining voltage stability is therefore essential for ensuring the secure and reliable operation of distribution systems. Once voltage deviates from

*haiyzhang1@163.com
https://doi.org/10.65102/is20261222
Published: 30 May, 2026

acceptable limits, electrical equipment may be damaged and outages may occur[3]. Consequently, accurate forecasting of voltage variations has become a key strategy for safeguarding grid stability and is also fundamental to implementing effective voltage regulation[4].

To improve voltage prediction under high PV penetration, existing research has primarily focused on two methodological directions: indirect prediction and direct prediction. Indirect approaches typically begin by forecasting load and distributed generation outputs at various nodes. These predicted values are then fed into a power flow model to obtain steady-state quantities such as nodal voltages[5]. For example, Bracale et al. introduced a Bayesian-based prediction framework in which probabilistic forecasts of wind, PV, and load outputs are used to derive their probability density functions (PDFs). These PDFs are subsequently combined with the point estimation method (PEM) to determine the probability distribution of nodal voltages[6]. Although such methods offer strong interpretability, they still encounter two major limitations: voltage estimation depends on a large amount of nodal injection data, and modeling the complex interdependencies among nodal injections under uncertainty substantially increases computational complexity. Direct prediction methods, in contrast, rely on historical voltage measurements and related influencing variables (such as load, generation output, and weather data) to directly forecast future voltage without depending on detailed physical models. For instance, Dejamkhooy et al. employed grey system theory and introduced a Fourier residual correction technique to develop an enhanced FGM model with improved accuracy[7]. Yi Wang et al. applied Lasso regression for feature selection and constructed an ensemble model based on SVR, RF, GBRT, and ERT to predict the voltage at a single node.

Traditional machine learning models, however, often require extensive manual feature engineering and struggle to capture high-dimensional, nonlinear spatiotemporal relationships. To overcome these limitations, N. Chettibi et al. proposed a real-time voltage forecasting approach based on a feedforward artificial neural network (ANN), using historical voltage, current, and frequency measurements as inputs[8]. Building on this, subsequent works expanded voltage prediction to multi-node scenarios. Mokhtar et al. developed a deep feedforward neural network capable of simultaneously predicting voltage magnitudes at most nodes in the system[9]. Considering the nuanced nature of these findings, Pertl, M. et al. considered what seem to be potential node interference issues in real systems and tended to suggest what appears to be a robust neural network voltage estimator to perform single-step prediction of three-phase voltages for multiple nodes[10]. In an effort to better capture temporal dependencies, Chen proposed a long short-term memory (LSTM)–based method to forecast multi-step voltage sequences, achieving significantly higher accuracy than traditional BP neural networks[11].

Despite these advancements, many studies still concentrate on voltage forecasting for individual PV systems and often overlook the spatial correlations among neighboring PV installations. In practice, adjacent PV units typically exhibit strong spatial dependencies that can be exploited to improve prediction performance. With the rapid development of graph neural networks (GNNs), numerous time-series forecasting tasks now incorporate GNNs to model spatial interactions among nodes[12-14]. For example, Luo et al. designed a spatiotemporal graph convolutional network (STGCN) for short-term voltage stability assessment, integrating grid topology with time-series data such as voltage and power flow to classify voltage stability states[15]. Other works have combined GNNs with attention mechanisms for multi-node PV power forecasting[16]. Furthermore, Zhong et al. introduced a weather-aware spatiotemporal attention network (WST-ANet), integrating a dynamic graph module and cross-attention mechanism to effectively model complex spatiotemporal patterns, achieving high accuracy in traffic flow forecasting[17].

Although these methods demonstrate strong performance in related fields, research on voltage prediction in distribution networks, especially under high PV penetration, still lacks sufficient consideration of spatiotemporal coupling characteristics. Weather variations significantly influence PV output, leading to amplified voltage fluctuations in distribution systems[18, 19]. Thus, incorporating weather information into the modeling framework is essential for improving prediction accuracy. To address these challenges, this paper proposes a graph neural network model that integrates distribution network topology and Numerical Weather Prediction (NWP) data, referred to as AST-CA-GNN—an enhanced version of the ASTGNN framework[20]. The model introduces a cross-attention mechanism to dynamically fuse NWP features with voltage sequences and constructs an adaptive dynamic graph based on prior topological information to capture evolving spatial dependencies among PV users.

The main contributions are as follows:

We encode NWP features such as temperature, cloud cover, rainfall, and solar irradiance, and dynamically fuse them with electrical features (e.g., voltage, active power, reactive power) through a cross-attention module. This enhances the model's capability to capture the influence of weather conditions on voltage dynamics.

We extract the relative positional information of users from the topology graph of a distribution network to form a prior topology graph, and combine it with a dynamic graph convolution module to model spatial dependencies among users.

The model takes as input the historical three-phase voltages and other electrical features of multiple users, the corresponding topology graph, and the NWP features at each location, and directly predicts the three-phase voltages of multiple users over multiple future time steps using an encoder–decoder architecture.

2 Data Description

The dataset employed in this study was obtained from a real rural distribution net-work located in the Northeast of China, spanning the period from August 1, 2024, to Feb-ruary 28, 2025. The topology of the distribution network is illustrated in Figure 1, where the red nodes represent users connected to distributed PV systems. It should be noted that node 47 is connected to two PV users (denoted as user 47 and user 47′). Within this dis-tribution network, a total of 10 users installed the PV systems.

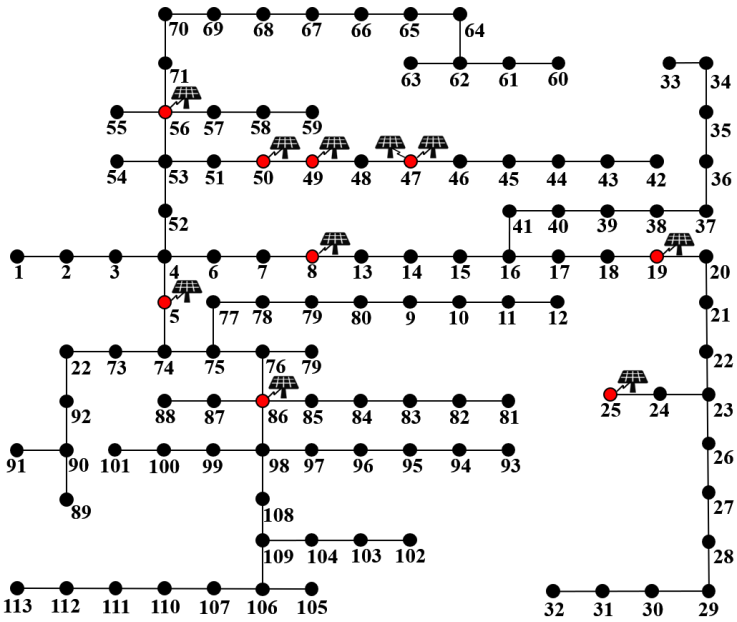


Figure 1: The topology graph of the a rural distribution network with 10 PV users.

As shown in Table 1, the data from the distribution network was sampled hourly, covering key electrical features including three-phase voltages, total active power, and total reactive power, resulting in a total of 5088 samples for each feature. In addition, to account for the impact of weather conditions on PV output and voltage fluctuations in the distribution network, hourly NWP data for this location were obtained via Open-Meteo website , includingsix variables: 2m air temperature (t2m), cloud cover, precipitation, shortwave radiation instant (SRI), direct radiation instant (DRI), and terrestrial radiation instant (TRI).

Table 1: Explanatory variables for spatiotemporal PV voltage prediction.

Explanatory	Abbr.	Description
Electrical features	Ua	Phase A Voltage (V)
	Ub	Phase B Voltage (V)
	Uc	Phase C Voltage (V)
	P	Total active power (kW)
	Q	Total reactive power (kW)
NWP features	t2m	2m air temperature (°C)
	cloud_cover	Proportion of cloud cover (%)
	precipitation	Total hourly precipitation (mm)
	SRI	shortwave radiation instant (W/m²)
	DRI	direct radiation instant (W/m²)
	TRI	terrestrial radiation instant (W/m²)

Furthermore, to investigate the relationships among the feature variables, this study employs the pearson correlation coefficient to quantitatively analyze the correlations between electrical features and weather features. Figure 2 presents the pearson correlation coefficient heatmap of the 11-dimensional features for user 56. The results indicate that there is a very strong positive correlation (>0.85) among the three-phase voltages, and they also exhibit a strong correlation with P (approximately 0.8), suggesting that an increase in power output may cause voltage increase. Among the NWP features, SRI, DRI, and TRI also show strong

correlations (ranging from 0.68 to 0.85) with the voltage features, indicating that weather conditions have a significant impact on voltage fluctuations.

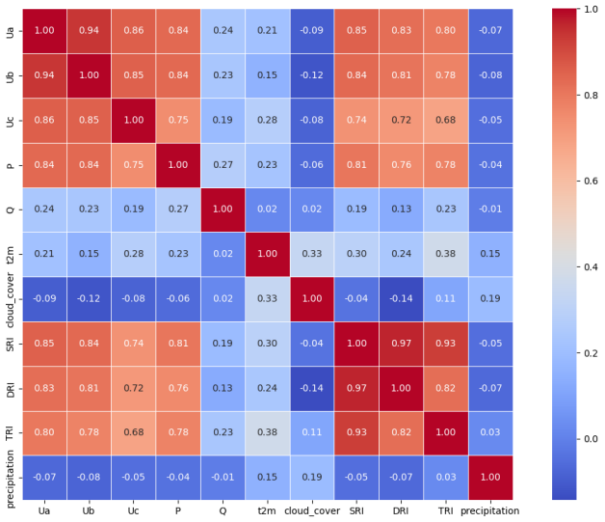


Figure 2: Pearson correlation coefficient heatmap of electrical features and NWP features for user 56.

3 Methods

3.1 Problem Statement

We define the topology graph of the distribution network as an undirected graph $G = (V, E)$, where V is a set of $|V| = N$ nodes. E is a set of $|E| = M$ edges and each edge represents the relative distance between the two nodes it connects. Given the D electrical feature sequences over the past T time steps $X_{t-T+1:t} \in R^{N\times T\times D}$, and the corresponding C NWP sequences $W_{t-T+1:t} \in R^{T\times C}$. the objective is to predict the three-phase voltages $\hat{Y}_{t+1:t+K} \in R^{N\times K\times F}$ of all PV users for the next K time steps, where $F = 3$ corresponds to the three voltage phases. Accordingly, the spatiotemporal voltage prediction problem can be defined as learning a mapping function $\mathcal{F}$:

$$\hat{Y}_{t+1:t+K} = \mathcal{F}(X_{t-T+1:t}, W_{t-T+1:t}, G) \tag{1}$$

3.2 Overall Framework

The overall architecture of the proposed AST-CA-GNN is shown in Figure 3. The model inputs include historical electrical feature sequences of all PV generations and the corresponding NWP sequences. First, the electrical feature sequences and NWP sequences are separately encoded as inputs to the Encoder. The Encoder is composed of multiple stackable modules, each containing a Temporal Multi-Head Cross-Attention layer and a Spatial Dynamic GCN layer. The cross-attention mechanism enables the model to dynamically perceive external weather conditions. The spatial dynamic GCN layer constructs a dynamic graph based on the relative positions of users in distribution network, modeling spatial correlations among them. Each sub-module is equipped with residual connections and Layer Normalization to enhance training stability. Similarly, the Decoder is composed of multiple stacked modules, each consisting of a Temporal Multi-Head Self-Attention layer, a Temporal Multi-Head Cross-Attention layer, and a Spatial Dynamic GCN layer. These components are responsible for modeling the temporal

dependencies of the electrical feature sequences itself, integrating the Encoder outputs, and capturing spatial structural information during the prediction stage, respectively. Finally, the Decoder output is passed through a linear layer to generate multi-step-ahead predictions of three-phase voltages for all users.

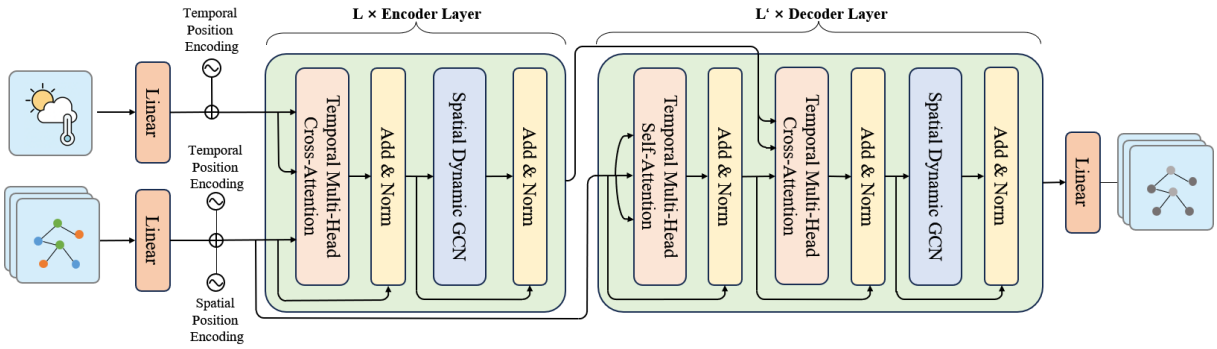


Figure 3: Model diagram of AST-CA-GNN. The model adopts an encoder–decoder architecture. The Encoder integrates NWP and electrical features through a cross-attention mechanism and a dynamic graph convolution layer, while the Decoder captures the intrinsic characteristics of the electrical features via spatiotemporal modeling and fuses the Encoder outputs to predict future voltage values.

3.3 Position Embedding

To enhance the model's spatiotemporal perception capability, AST-CA-GNN introduces a positional encoding mechanism. Specifically, different types of data are encoded using different strategies to match their temporal and spatial characteristics. Since NWP features are shared public information within the same distribution network, and the users are geographically close—located within the same village or town—they experience identical climatic conditions at any given moment. Therefore, spatial positional encoding is unnecessary, and only temporal positional encoding is applied, which effectively captures the dynamic trends of weather changes over time. For each user's electrical feature sequence, since voltage fluctuations exhibit significant temporal periodicity and spatial correlation, we apply both temporal positional encoding and spatial positional encoding to enhance the model's ability to capture the complex spatiotemporal interactions among multiple users. Before applying positional encoding, the electrical feature sequences and NWP feature sequences are separately passed through a linear layer to project their feature dimensions to d_{model}.

In the temporal dimension, AST-CA-GNN adopts the same sinusoidal fixed positional encoding as used in Transformers, encoding the position of each time step into a set of fixed periodic vectors. These vectors are added to the original inputs to help the model distinguish features at different temporal positions, thereby capturing the temporal dependencies and periodic variations in the sequence. Specifically, the temporal positional encoding of the *t*-th time step can be expressed as:

$$PE(t, 2d) = sin(t/10000^{2d/d_{model}})$$

$$PE(t, 2d+1) = cos(t/10000^{2d/d_{model}}) \quad (2)$$

In the spatial dimension, AST-CA-GNN assigns each user a learnable spatial embedding vector to represent its spatial identity. Specifically, we construct an embedding table $E \in R^{N \times d_{model}}$, to encode all N users, generating a d_{model} dimensional spatial positional vector.

These embeddings are optimized as model parameters during training and are broadcast-added to the node feature representations at every time step.

3.4 Spatiotemporal Encoder

The encoder module of AST-CA-GNN is primarily designed to fuse historical electrical feature sequences with NWP sequences. The overall structure consists of multiple stacked encoder blocks. In each encoder block, a Temporal Multi-Head Cross-Attention layer is first applied to model the temporal dependencies between the NWP and electrical feature sequences. Specifically, the electrical feature sequence of each user is used as the query, while the NWP sequence serves as the key and value. Attention weights are then computed to guide the model to focus on critical weather information. However, the conventional attention mechanism was originally developed for processing discrete tokens such as words and does not fully account for local trend patterns present in continuous data. To address this, we apply 1D convolutions to both the query and key before computing attention, thereby equipping the attention mechanism with trend-awareness. Formally, the temporal trend-aware multi-head cross-attention (TrCrossAttention) is defined as follows:

$$TrCrossAttention(Q,K,V) = \oplus(Trhead_1,\dots,Trhead_h)W^O \tag{3}$$

$$Trhead_j = Attention(\Phi_j^Q(Q),\Phi_j^K(K),VW_j^V) \tag{4}$$

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_{model}}})V \tag{5}$$

where Φ_j^Q, Φ_j^K denote the 1D convolution operations applied to the query and key,W_j^V is the projection matrix applied to the value, and W^O is the output projection matrix of the multi-head attention. In the L^{th} encoder layer, given the input $X^{(L-1)} \in R^{N\times T\times d_{model}}$ and $W^{(L-1)} \in R^{N\times T\times d_{model}}$, after passing through the temporal trend-aware multi-head cross-attention, we obtain an intermediate sequence $Z^{(L-1)} = (Z_{t-T+1}^{(L-1)}, Z_{t-T+2}^{(L-1)},\dots, Z_t^{(L-1)}) \in R^{N\times T\times d_{model}}$.

To capture the time-varying spatial correlation characteristics among users, AST-CA-GNN introduces a Spatial Dynamic GCN in each encoder block. This module dynamically computes the spatial correlation strengths between users via an attention mechanism and, in combination with a prior topology graph, constructs an adaptive dynamic graph structure, thereby effectively modeling spatial dependencies in the sequence. Specifically, we utilize the topology graph G of the distribution network, whose corresponding adjacency matrix is $\tilde{A} \in R^{N\times N}$, The adjacency matrix $\tilde{A}$ is then symmetrically normalized to obtain the weight matrix *A*, which is expressed as follows:

$$A = \widetilde{D}^{-\frac{1}{2}}\tilde{A}\widetilde{D}^{-\frac{1}{2}} \in R^{N\times N} \tag{6}$$

where $\widetilde{D}_{ii} = \sum_j \tilde{A}_{ij}$.In addition, to dynamically adjust the correlations between users, a self-attention mechanism is employed to obtain the spatial correlation weight matrix *S*. The computation of *S* is as follows:

$$S_t = softmax(\frac{Z_t^{(L-1)}Z_t^{(L-1)^T}}{\sqrt{d_{model}}}) \in R^{N\times N} \tag{7}$$

where $S_t(i,j)$ represents the correlation strength between user *i* and user *j*, By performing an element-wise multiplication between the spatial correlation weight matrix S_t and the weight matrix *A*, the correlations between nodes can be automatically adjusted over time. Therefore, the Spatial Dynamic GCN(DGCN) can be expressed as:

$$X_t^{(L)} = DGCN(Z_t^{(L-1)}) = \sigma((A \odot S_t) Z_t^{(L-1)} W^{(L)}) \tag{8}$$

Finally, we obtain the output of the L^{th} encoder layer: $X^{(L)} = (X_{t-T+1}^{(L)}, X_{t-T+2}^{(L)}, \ldots, X_t^{(L)}) \in R^{N \times T \times d_{model}}$

3.5 Spatiotemporal Decoder

The decoder module of AST-CA-GNN seems to be mainly used to generate multi-step three-phase voltage predictions for all users, which is particularly important in the research results. Its underlying structure appears to consist of L' stacked decoder layers, with each decoder layer containing two attention sub-layers and one spatial dynamic GCN sub-layer, within this broader analysis framework. In the first attention sublayer, the time multi-head self-attention mechanism models the electrical feature sequence itself to capture the temporal dependencies within the sequence. The second attention sublayer is a temporal multi-head cross-attention mechanism, where queries come from the output of the self-attention sublayer, while keys and values come from the output of the encoder to achieve fusion with encoder information. In the L'^{th} decoder layer, given an electrical feature sequence $H^{(L'-1)}$, the self-attention sub-layer first produces an intermediate sequence $M^{(L'-1)} \in R^{N \times T \times d_{model}}$, which is then processed by the multi-head cross-attention mechanism to integrate the encoder output and obtain the representation $N^{(L'-1)} \in R^{N \times T \times d_{model}}$. Following the cross-attention output, the decoder incorporates a Spatial Dynamic GCN sub-module to compute attention weights based on the relationships among input features at the current time step and to fuse it with the prior topology graph. Similar to the encoder, each sub-module is followed by residual connections and layer norm to mitigate the training instability caused by deep stacking. By progressively stacking multiple decoder layers, the model produces the L' layer output $H^{(L')} = (H_{t-T+1}^{(L')}, H_{t-T+2}^{(L')}, \ldots, H_t^{(L')}) \in R^{N \times T \times d_{model}}$, which is finally passed through a fully connected layer to generate multi-step three-phase voltage predictions for all users $\hat{Y}_{t+1:t+K} = (\hat{Y}_{t+1}, \hat{Y}_{t+2}, \ldots, \hat{Y}_{t+\mathrm{k}})$.

4 Experiment

4.1 Data Processing

The dataset employed in this study originates from a rural distribution network located in Northeast China and spans the period from August 1, 2024, to February 28, 2025. It comprises five electrical variables and six numerical weather prediction (NWP) features. The network supplies approximately 197 households, around 10 of which are equipped with rooftop photovoltaic (PV) systems. This characteristic plays an important role in subsequent analyses.

Before model development can proceed, it is necessary to preprocess the raw data and separate it into appropriate datasets, ensuring that the subsequent analysis is both reliable and consistent. Considering the substantial differences in magnitude across various features, the data requires additional interpretive attention. These disparities indicate that standardization is necessary and can be applied as follows:

$$z' = \frac{z-\mu}{\sigma} \tag{9}$$

where z denotes the original feature value, μ the mean, and σ the standard deviation of the feature. As the electrical feature sequences contain missing values, we employ linear interpolation to fill in the gaps, ensuring continuity and consistency of feature sequences. After these preprocessing procedures, the dataset is transformed into samples for model training. This work focuses on a day-ahead forecasting task: the previous seven days of measurements (168 hours) are used to predict the three-phase voltages for all users during the following 24-hour period. Sample construction follows a sliding-window strategy, with an input window length of T=168 and a prediction window length of K=24. The dataset is subsequently divided into training, validation, and testing subsets in proportions of 7:2:1. A total of 4,897 valid samples are generated, including 3,427 for model training, 980 for validation, and 490 for testing.

The development of the AST-CA-GNN model relies on the users' inherent prior topology graph. To establish this graph, we first extract the two-dimensional coordinates (x_i, y_i) for each PV user from the distribution network's topology graph G. Subsequently, the edge weight will be compute between any two users based on their Euclidean distance:

$$\tilde{A}_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{10}$$

This procedure yields an undirected adjacency matrix $\tilde{A}_{ij}$. Because the original Euclidean distances span a wide numerical range, they are rescaled to the interval $(0,1)$ to improve compatibility with the model. The normalized adjacency matrix is shown in Figure 4 and exhibits a symmetric structure. In the heatmap, darker regions correspond to closer spatial relationships between nodes, while lighter regions indicate greater separation. The values shown represent the normalized distances between individual user pairs. For instance, the distance assigned to users 25 and 56 is 1, signifying that these two users are positioned at the greatest relative distance within the network.

Figure 4: Heatmap of euclidean distances between PV users.

4.2 Model Setting

The three-phase voltage prediction task is formulated as a multi-step multivariate forecasting problem. The model takes as input a continuous 168 hours historical observation sequence and predicts the three-phase voltage values of all users for the next 24 hours. The hyperparameter settings of the AST-CA-GNN model are shown in Table 2. Both the encoder and decoder adopt a two-layer stacked structure, enabling effective spatiotemporal modeling while avoiding the training instability that can arise from excessively deep networks. The multi-head attention mechanism uses four attention heads, allowing the model to capture temporal dependencies from multiple subspaces in parallel. The hidden dimension d_{model} is set to 64, mapping both the electrical feature sequences and NWP feature sequences into the same-dimensional representation space. In addition, a kernel size of 3 is used for the 1D convolution applied to the query and key in the multi-head attention mechanism, enhancing the model's ability to capture local trends in the sequence.

Table 2: The hyperparameters and their value of AST-CA-GNN model setting.

Hyperparameters	Value
Encoder layers	2
Decoder layers	2
d_{model}	64
N_{head}	4
kernel size	3
adjacency matrix	$\hat{A}$
learning rate	0.0001
epochs	100
optimizer	Adam
batch size	64

To fully verify the effectiveness and robustness of the AST-CA-GNN model in the multi-node three-phase voltage prediction task for PV distribution network, the model was implemented and trained using the pytorch framework. The training and deployment were conducted on an NVIDIA GeForce RTX 4070 Ti GPU with 12 GB of memory.

4.3 Baseline Models

To comprehensively evaluate the performance of the proposed AST-CA-GNN model in the spatiotemporal voltage prediction task, five representative time series forecasting models were selected as baseline models, covering different approaches from traditional time series modeling methods to graph neural network architectures. The selected models are as follows:

1. LSTM[21] is one of the classic recurrent neural network architectures and can effectively capture long-term dependencies in time series. In our implementation, the multi-dimensional features of each node are concatenated into a single input vector, which is fed into a two-layer stacked LSTM network with 128 hidden units per layer. The final output is obtained by mapping through a linear layer to predict the three-phase voltages of each user for multiple future time steps.

2. AGCRN[22] is a hybrid model combining gated recurrent units (GRU) with adaptive graph convolution. Its core idea is to assign a learnable embedding vector to each node and construct a dynamic adjacency matrix based on the similarity of node embeddings, enabling spatial relationship modeling without prior topology information. In our implementation, the number of nodes is set to 10, the input feature dimension to 11, the GRU hidden size to 64, the embedding dimension to 16, the Chebyshev polynomial order (Cheb-k) to 2, and the model consists of 2 stacked layers.

3. STGCN[23] is a spatiotemporal forecasting model based on graph neural networks, combining graph convolution (GCN) with temporal convolution (TConv) to simultaneously model spatial dependencies among nodes and temporal evolution within each node. The adjacency matrix is a static graph representing fixed spatial relationships. In our implementation, the graph convolution is implemented using Chebyshev polynomials, with both the temporal and graph convolution kernel sizes set to 3, and the GLU activation function is used to fuse the outputs of the two modules.

4. ASTGCN[24] extends STGCN by introducing spatial and temporal attention mechanisms, enhancing the model's ability to focus on key nodes and key time steps. This model adaptively adjusts the influence of the graph structure on predictions, offering more flexibility compared to static graphs. In our implementation, the model stacks two graph convolution blocks, each employing a 4-order Chebyshev graph convolution, with both the spatial and temporal channel sizes set to 32.

5. ASTGNN is based on an encoder–decoder architecture, where both the encoder and decoder consist of multiple stacked layers. Each layer alternates between a trend-aware multi-head self-attention module (with a kernel size of 3) and a dynamic graph convolution module. The decoder generates multi-step predictions in an autoregressive manner. In our experiments, both the encoder and decoder have two layers, and the hidden size is set to 64.

All models are trained and tested using the same data splits and evaluation metrics to ensure fair comparison.

4.4 Evaluation Metrics

To comprehensively evaluate the performance of the proposed AST-CA-GNN model in the spatiotemporal voltage prediction task, five commonly used time series forecasting evaluation metrics are selected to quantify the prediction accuracy and stability of the model across different time steps and nodes from multiple perspectives. These metrics include Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Weighted Absolute Percentage Error (WAPE), and the coefficient of determination (R^2):

$$MAE = \frac{1}{N\times K\times F}\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{f=1}^{F}\left|\hat{y}_{n,k,f} - y_{n,k,f}\right| \tag{11}$$

$$RMSE = \sqrt{\frac{1}{N\times K\times F}\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{f=1}^{F}\left(\hat{y}_{n,k,f} - y_{n,k,f}\right)^2} \tag{12}$$

$$MAPE = \frac{100\%}{N\times K\times F}\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{f=1}^{F}\left|\frac{\hat{y}_{n,k,f}-y_{n,k,f}}{y_{n,k,f}+\epsilon}\right| \tag{13}$$

$$WAPE = \frac{\sum_{n,k,f}|\hat{y}_{n,k,f}-y_{n,k,f}|}{\sum_{n,k,f}|y_{n,k,f}|+\epsilon}\times 100\% \tag{14}$$

$$R^2 = 1 - \frac{\sum_{n,k,f}\left(y_{n,k,f}-\hat{y}_{n,k,f}\right)^2}{\sum_{n,k,f}(y_{n,k,f}-\bar{y})} \tag{15}$$

where $\hat{y}_{n,k,f}$ denotes the predicted voltage value of phase *f* for the *n*-th user at the *k*-th time step, $y_{n,k,f}$ represents the corresponding ground truth value. ϵ is a small positive constant introduced to avoid division by zero. $\bar{y}$ denotes the mean of $\hat{y}$.

4.5 Result and discussion

To evaluate the comprehensive performance of the proposed AST-CA-GNN model in the spatiotemporal voltage prediction task, it is compared against five benchmark models (LSTM, AGCRN, STGCN, ASTGCN, and ASTGNN) on the same test set.

Table 3: Prediction performance comparison for three-phase voltage.

	Models	MAE	RMSE	MAPE(%)	WAPE(%)	R^2
Baselines	LSTM	5.3564	7.2249	2.1785	2.1814	0.8443
	AGCRN	4.9054	6.6731	2.0125	1.9977	0.8672
	STGCN	5.3485	7.0279	2.1693	2.1782	0.8527
	ASTGCN	4.5488	5.9641	1.8487	1.8525	0.8939
	ASTGNN	4.3475	5.9060	1.7818	1.7705	0.8959
Proposed	AST-CA-GNN	4.2320	5.7451	1.7304	1.7235	0.9015

Table 3 summarizes the prediction results for three-phase voltages. On all evaluation metrics, the accuracy of AST-CA-GNN is superior to that of the baseline models, indicating its stronger generalization ability. Compared with the best-performing baseline model, ASTGNN, the proposed method reduces MAE and RMSE by approximately 2.65% and 2.72%, respectively. For MAPE and WAPE, the improvement rates reached approximately 2.89% and 2.65%, respectively, achieving the lowest overall error among all models. In addition, its coefficient of determination (R^2) reached 0.9015, further highlighting the enhanced ability of the model in fitting voltage sequences.

A more detailed comparison reveals several important phenomena. Although LSTM can effectively learn temporal trends, it cannot characterize spatial interactions between users, which limits its performance in multi-node voltage prediction tasks. Although AGCRN and STGCN introduce spatial components, their performance is somewhat constrained due to the lack of prior topological information that can guide spatial learning. Among these two models, AGCRN performs relatively better because its temporal attention mechanism enhances the extraction of time-dependent features.

ASTGCN improves the above methods by introducing a spatiotemporal attention mechanism, which can selectively focus on important nodes and critical time steps, thus achieving significant performance improvement compared to AGCRN and STGCN. ASTGNN further adopts an encoder–decoder structure and convolution-based sequence modeling, which further improves accuracy and outperforms ASTGCN in performance.

In contrast, AST-CA-GNN introduces a multi-head attention mechanism for time modeling and a dynamic graph convolution module for capturing dynamic spatial dependencies based on the encoder–decoder framework. In addition, the fusion design of cross-attention mechanisms enables the model to effectively integrate numerical weather forecasting (NWP) with electrical features, thereby better describing voltage fluctuations under different weather conditions and significantly improving the utilization of meteorological information.

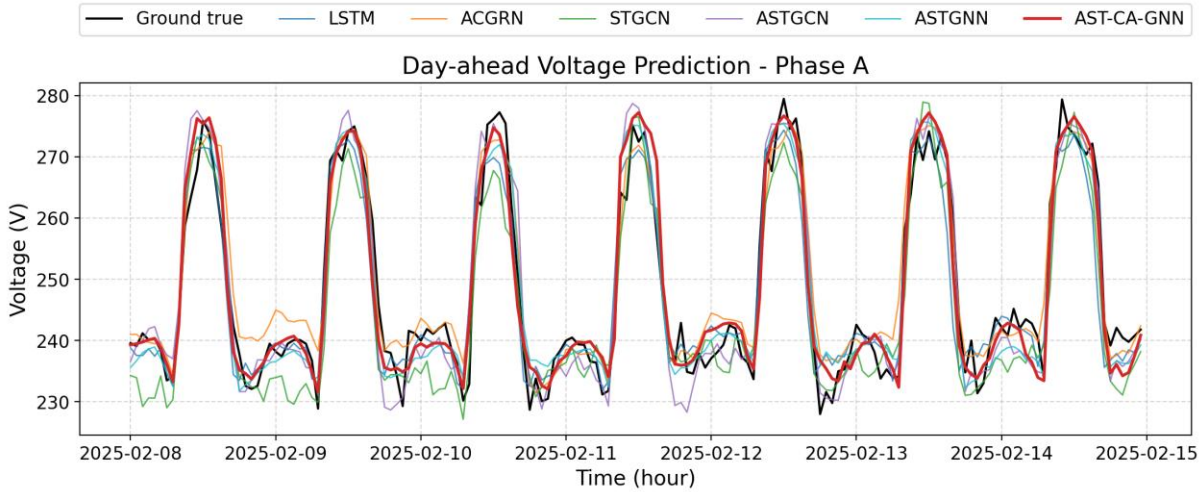


Figure 5: Visualization of phase-A voltage predictions for user 56 over a continuous 7-day horizon. The black line denotes the ground truth, and the red line represents the day-ahead predictions of the proposed AST-CA-GNN model.

Within this broader analytical framework, the phase-A voltage prediction results for user 56 over a 7-day time span were visualized to further validate the predictive performance of the proposed model. The results shown in Figure 5 show the curve changes of the six models

compared with the true values. Overall, the voltage curve exhibits distinct diurnal characteristics: voltage usually peaks at noon (12:00–14:00), which seems to correspond to the periods when solar irradiance is usually highest, indicating that photovoltaic power generation has a significant impact on voltage. In contrast, when the solar irradiance weakens at night, the voltage generally drops to its lowest point and fluctuates with load changes, which is closely related to users' electricity consumption behavior.

Combining these detailed observations, it can be found that although most models are able to capture the diurnal pattern of voltage (daytime rise and nighttime decline), their accuracy in reproducing peaks and troughs varies greatly. AST-CA-GNN generally fits more consistently with the true values across time periods and exhibits minimal error at peaks. This analysis also shows that ASTGNN is effective in modeling periodic patterns, but at certain midday peaks (e.g., February 10, 2025), there is a greater bias than that of AST-CA-GNN, suggesting that it is less effective at utilizing meteorological features.

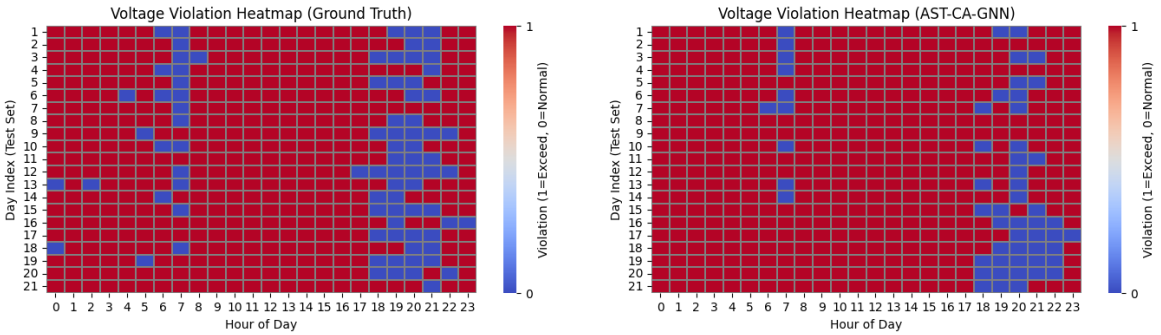
Moreover, ASTGCN's forecasts are generally able to follow the overall trend, but there is a significantly larger bias at the peaks and troughs on February 11, 2025. Further, STGCN has a significant underestimation of the nighttime troughs on February 8 and February 10, while AGCRN is significantly overestimated during the same period, indicating that models that introduce a priori spatial topology maps, such as ASTGCN, are more accurate in portraying the differences between users in nighttime scenarios without photovoltaic generation. Taken together, these results show that LSTM performs the worst across all metrics, often underestimating voltage peaks and overestimating troughs.

Table 4: performance comparison for voltage limit violation prediction.

	Models	Accuracy	Precision	Recall	F1
Baselines	LSTM	0.8497	0.8488	0.9477	0.8955
	AGCRN	0.8537	0.8465	0.9586	0.8990
	STGCN	0.8318	0.8698	0.8740	0.8719
	ASTGCN	0.8649	0.9058	0.8942	0.9000
	ASTGNN	0.8738	0.8911	0.9277	0.9090
Proposed	AST-CA-GNN	0.8853	0.8952	0.9408	0.9175

Based on the voltage forecasting results, we further analyzed the performance of each model in identifying voltage violation events. The voltage violation threshold is set to ±7%. Specifically, when the single-phase voltage exceeds $220 \times (1 + 7\%)$ V, it is considered an overvoltage, and when it falls below $220 \times (1 - 7\%)$ V, it is considered an undervoltage.

As can be observed from Table 4, the proposed AST-CA-GNN model demonstrated the best overall performance across all metrics. Specifically, it achieved an Accuracy of 0.8853. Its Precision and Recall reached 0.8952 and 0.9408, respectively, indicating that the model can not only effectively reduce false positives but also more comprehensively identify voltage violation events. The F1 score was 0.9175—the highest among all models—showing that the model achieved the best balance between Precision and Recall. Further analysis revealed that while LSTM and AGCRN achieved relatively high Recall, their Precision was low. This is because they tended to overestimate the three-phase voltages at night, resulting in some more cases being incorrectly classified as violations, thereby reducing missed detections but increasing false positives. STGCN, on the other hand, generally underestimated voltages, which led to better Precision than LSTM and AGCRN but the highest number of missed detections. In addition, both ASTGCN and ASTGNN achieved F1 scores above 0.90, outperforming other baseline models.



(a) (b)

Figure 6: Hourly heatmaps of voltage violation events for user 56 on the test set. Figure 6(a) shows the ground-truth violations, while Figure 6(b) presents the predictions of AST-CA-GNN. Red cells indicate a voltage-limit violation, whereas blue cells denote normal voltage.

It is noteworthy that in this low-voltage distribution network, the vast majority of voltage violation events are overvoltage cases. Figure 6 illustrates the temporal distribution of all voltage violation events for user 56, including both the actual events and the predictions generated by our AST-CA-GNN model. The results reveal that the voltage condition of this PV user is highly unfavorable, with voltages remaining in a violation state for most of the day. Only during two short periods—6:00–7:00 in the morning and 18:00–22:00 in the evening—does the voltage tend to fall within the normal range. Remarkably, our AST-CA-GNN model is able to accurately predict these overvoltage periods one day in advance. This visualization analysis is consistent with the findings presented in Table 4.

In summary, AST-CA-GNN achieved the best results in both spatiotemporal voltage prediction and limit violation prediction. This further validates its superiority in capturing spatiotemporal dependencies for both NWP and electrical features.

5 Conclusion

This paper endeavors to address the challenge of the spatiotemporal voltage predictions in PV integration distribution networks and proposes a weather-aware spatiotemporal graph neural network model, AST-CA-GNN. In the temporal domain, the model presumably employs a cross-attention mechanism to dynamically fuse NWP and electrical features, while in the spatial domain, it largely leverages dynamic graph convolution to capture time-varying spatial dependencies among users, thereby enhancing what appears to be its ability to model complex, nonlinear multivariate sequences. Experimental results on real-world PV distribution network data appears to tend to suggest is that AST-CA-GNN, from this particular interpretive perspective, seems to outperform all baseline models across what are typically considered standard metrics. In particular, for multi-node and multi-phase voltage forecasting, it apparently captures voltage fluctuation trends and peak–valley patterns with a considerable degree of precision, exhibiting what seems to be superior generalization and stability. The model also appears to seem capable in predicting voltage limit violations, thereby providing valuable insights for system operators to schedule corrective actions and minimize the risk of equipment malfunctions.

Funding

This research was funded by science and technology projects of State Grid Liaoning Electric Power Co., Ltd. SGLNSY00KSJS2505033 (Research on Panoramic Perception Technology for Distribution Networks Considering the Impact of Distributed PV Reverse Overload).

References

[1] Iweh C D, Gyamfi S, Tanyi E, et al. Distributed generation and renewable energy integration into the grid: Prerequisites, push factors, practical options, issues and merits [J]. Energies, 2021, 14(17): 5375.

[2] Farhoodnea M, Mohamed A, Shareef H, et al. Power quality impact of grid-connected photovoltaic generation system in distribution networks[C]//2012 IEEE Student Conference on Research and Development (SCOReD). IEEE, 2012: 1-6.

[3] Akinyemi A S, Musasa K, Davidson I E. Analysis of voltage rise phenomena in electrical power network with high concentration of renewable distributed generations [J]. Scientific Reports, 2022, 12(1): 7815.

[4] Razavi S-E, Rahimi E, Javadi M S, et al. Impact of distributed generation on protection and voltage regulation of distribution systems: A review [J]. Renewable and Sustainable Energy Reviews, 2019, 105: 157-167.

[5] Wang Y, Von Krannichfeldt L, Zufferey T, et al. Short-term nodal voltage forecasting for power distribution grids: An ensemble learning approach [J]. Applied Energy, 2021, 304: 117880.

[6] Bracale A, Caramia P, Carpinelli G, et al. A bayesian-based approach for a short-term steady-state forecast of a smart grid [J]. IEEE Transactions on Smart Grid, 2013, 4(4): 1760-1771.

[7] Dejamkhooy A, Dastfan A, Ahmadyfard A. Modeling and forecasting nonstationary voltage fluctuation based on grey system theory [J]. IEEE Transactions on Power Delivery, 2014, 32(3): 1212-1219.

[8] Chettibi N, Pavan A M, Mellit A, et al. Real-time prediction of grid voltage and frequency using artificial neural networks: An experimental validation [J]. Sustainable Energy, Grids and Networks, 2021, 27: 100502.

[9] Mokhtar M, Robu V, Flynn D, et al. Prediction of voltage distribution using deep learning and identified key smart meter locations [J]. Energy and AI, 2021, 6: 100103.

[10] Pertl M, Douglass P J, Heussen K, et al. Validation of a robust neural real-time voltage estimator for active distribution grids on field data [J]. Electric Power Systems Research, 2018, 154: 182-192.

[11] Chen Y. Voltages prediction algorithm based on LSTM recurrent neural network [J]. Optik, 2020, 220: 164869.

[12] Jin M, Koh H Y, Wen Q, et al. A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.

[13] Li W, Law K L E. Deep learning models for time series forecasting: A review [J]. IEEE Access, 2024, 12: 92306-92327.

[14] Cini A, Marisca I, Zambon D, et al. Graph deep learning for time series forecasting [J]. ACM Computing Surveys, 2025, 57(12): 1-34.

[15] Luo Y, Lu C, Zhu L, et al. Data-driven short-term voltage stability assessment based on spatial-temporal graph convolutional network [J]. International Journal of Electrical Power & Energy Systems, 2021, 130: 106753.

[16] Simeunović J, Schubnel B, Alet P-J, et al. Interpretable temporal-spatial graph attention network for multi-site PV power forecasting [J]. Applied Energy, 2022, 327: 120127.

[17] Zhong H, Wang J, Chen C, et al. Weather interaction-aware spatio-temporal attention networks for urban traffic flow prediction [J]. Buildings, 2024, 14(3): 647.

[18] Jha R R, Dubey A. Local smart inverter control to mitigate the effects of photovoltaic (pv) generation variability[C]//IEEE, 2019: 1-6.

[19] Łowczowski K, Nadolny Z. Voltage Fluctuations and Flicker in Prosumer PV Installation [J]. Energies, 2022, 15(6): 2075.

[20] Guo S, Lin Y, Wan H, et al. Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting [J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 34(11): 5415-5428.

[21] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural computation, 1997, 9(8): 1735-1780.

[22] Bai L, Yao L, Li C, et al. Adaptive graph convolutional recurrent network for traffic forecasting [J]. Advances in neural information processing systems, 2020, 33: 17804-17815.

[23] Yu B, Yin H, Zhu Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting [J]. arXiv preprint arXiv:170904875, 2017.

[24] Guo S, Lin Y, Feng N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[C]//Proceedings of the AAAI conference on artificial intelligence. 2019: 922-929.