Spatio-temporal Hybrid Graph-Transformer Model for High-fidelity Renewable Energy Forecasting

Muhammad Ali Iqbal, Junho Ahn, and Soo Kyun Kim

Abstract—**Accurate wind and solar power forecasting is crucial for grid stability and renewable energy integration, especially in island like Jeju Island, Republic of Korea. This paper tackles challenges such as nonlinear patterns, spatial interdependencies, and resource variability. Current models are limited because they are resource specific, treat spatial and temporal dependencies separately, and lack dynamic inter-site relationship modeling, which are essential in dispersed systems. To overcome these limitations, we propose a spatio-temporal hybrid graph-Transformer (ST-HGT) model for high-fidelity renewable energy forecasting, which combines graph attention networks (GATs), temporal convolutional networks (TCNs), neural basis expansion analysis for interpretable time-series forecasting (N-BEATS), Transformer, and Informer, within a modular architecture. The proposed ST-HGT model captures spatio-temporal dynamics by learning inter-site correlations from weather and power data via a dynamic graph and employing complementary temporal modules to model local features, long-range dependencies, and complex patterns. The proposed ST-HGT model is trained and evaluated independently for wind and solar power forecasting, using real-world datasets from multiple stations across Jeju Island. The results show that the proposed ST-HGT model markedly surpasses state-of-the-art baselines, including Informer, temporal fusion Transformer (TFT), N-BEATS, and recurrent networks. For solar power forecasting, the proposed ST-HGT model achieves a root mean square error (RMSE) of 21.71 kWh and a coefficient of determination R^2 of 0.999, with an improvement of more than 96.6% in RMSE relative to the best baseline model. For wind power forecasting, the proposed ST-HGT model achieves an RMSE of 853 kWh and an R^2 of 0.99, outperforming all baselines, which have RMSEs exceeding 65000 kWh.**

Index Terms—**Spatio-temporal dependency, graph attention network (GAT), Transformer, Informer, renewable energy forecasting, time-series forecasting, energy system optimization, deep learning.**

Manuscript received: June 18, 2025; revised: October 31, 2025; accepted: January 28, 2026. Date of CrossCheck: January 28, 2026. Date of online publication: March 13, 2026.

This research study was supported by the Regional Innovation System and Education (RISE) program through the Jeju RISE center, funded by the Ministry of Education (MOE) and Jeju Special Self-governing Province, Republic of Korea (No. 2025-RISE-17-001).

M. A. Iqbal and S. K. Kim (corresponding author) are with the Department of Computer Engineering, Jeju National University, Jeju, Republic of Korea (e-mail: aliiqbal073@gmail.com; kimsk@jejunu.ac.kr).

J. Ahn is with the Department of AI Computer Science and Engineering, Kyonggi University, Suwon, Republic of Korea (e-mail: jha@kyonggi.ac.kr).

DOI: 10.35833/MPCE.2025.000504

I. Introduction

RENEWABLE energy sources such as wind and solar power introduce considerable variability into power systems owing to their reliance on dynamic, weather-dependent phenomena [1]. Accurate forecasting is imperative for preserving grid stability, optimizing reserves, and facilitating real-time operational control. High-quality forecasting supports efficient unit commitment and economic dispatch, thereby reducing fuel consumption and renewable curtailment [2]. As the global integration of renewable energy sources progresses, particularly within isolated systems such as Jeju Island, Republic of Korea characterized by limited interconnectivity and vulnerability to typhoons, cloud movements, and wind fluctuations, precise and high-resolution forecasting models become indispensable to guarantee operational resilience.

Traditional models including numerical weather prediction (NWP) [3], [4], autoregressive integrated moving average (ARIMA) [5], and persistence models [6], provide a baseline performance. However, they struggle to capture rapid and nonlinear variations. Deep learning (DL) models including recurrent neural networks (RNNs) and Transformer variants [7]-[10] improve temporal modeling but often treat sites independently, ignoring inter-site correlations that may boost forecasting accuracy. Emerging models aim to capture spatio-temporal dependencies using convolutional neural networks (CNNs) [11], attention-based RNNs [12], and ensemble models [13], [14]. Hybrid models, which integrate Transformers with auxiliary modules, have demonstrated promise in extending forecasting horizons [15]-[17], and attention mechanisms have proven effective across multiple domains [18]. Most existing models focus on either spatial or temporal dynamics alone and rarely enable seamless forecasting across wind and solar resources. We propose a spatio-temporal hybrid graph-Transformer (ST-HGT) model, which combines graph attention networks (GATs), temporal convolutional networks (TCNs), neural basis expansion analysis for interpretable time-series forecasting (N-BEATS), Transformer, and Informer, within a modular architecture. It captures spatial correlations and multi-scale temporal dependencies in one extensible architecture. Unlike prior work, the proposed ST-HGT model forecasts wind and solar power without resource-specific tweaks, offering higher accuracy.

1) We propose a unified and resource-agnostic architecture that is trained independently on wind and solar datasets,

without any architecture-specific adjustments. “Unified” means a single architecture and training protocol for both resources, with independent weights for each, ensuring comparability, easier deployment, and avoiding negative transfer.

2) A GAT-based dynamic graph encoder effectively captures inter-site dependencies through spatial proximity and correlation-aware edge weights, thereby enhancing the adaptability and transferability of the proposed ST-HGT model.

3) Through comprehensive experimentation including ablation studies, benchmark comparisons, and residual diagnostics, we demonstrate notable advancements of the proposed ST-HGT model over current state-of-the-art models in both accuracy and reliability.

Prior studies of typical forecasting models are categorized into five groups: ① statistical and NWP models, ② classical machine learning (ML) models, ③ deep temporal models, ④ Graph neural network (GNN)-based models, and ⑤ hybrid models. Statistical and NWP models are interpretable but have coarse resolution and low update efficiency [3]-[6]. Classical ML models enhance nonlinearity; however, they are insufficient in managing long-term memory and multi-scale effects [19]. Deep temporal models including RNNs, TCNs, N-BEATS, and Transformers are capable of capturing longer-term patterns but lack cross-site correlation modeling [20]-[23]. GNN-based models encode cross-site interactions but generally employ a singular temporal encoder [24]-[26]. Hybrid models combine multiple modules; nonetheless, many of them lack a principled mechanism for fusion. The proposed ST-HGT model addresses these deficiencies by integrating a graph encoder with complementary temporal experts and a learned convex fusion. Achievements in renewable energy forecasting have evolved from traditional statistical models to sophisticated DL models. Traditional models such as ARIMA and persistence models are computationally efficient but fall short in capturing the nonlinear and volatile characteristics of wind and solar power outputs [11], [19]. Classical ML models including random forests and support vector machines (SVMs) are extensively employed. However, they are limited in modeling complex temporal and spatial patterns. DL models such as long short-term memory (LSTM) and gated recurrent unit (GRU) are also prevalent for modeling temporal dependencies [20], though they face challenges with long sequences. Furthermore, various studies have integrated these models with attention mechanisms and CNNs [11]. For example, TCNs utilize dilated convolutions to facilitate efficient long-range modeling and parallel processing [27], [28]. The performance characteristics of typical forecasting models are listed in Table I.

TABLE I
PERFORMANCE CHARACTERISTICS OF TYPICAL FORECASTING MODELS

Model	Reference	Spatial model	Temporal range	Strength	Limitation
Statistical and NWP	[3]-[6]		Short-medium	Interpretable	Coarse resolution and low update efficiency
Classical ML	[19]		Short	Simple and fast	Insufficient in managing long-term memory and multi-scale effects
Deep temporal	[20]-[23]	Independent sites	Short-long	Pattern-capturing	Lack of cross-site correlation modeling
GNN-based	[24]-[26]	Static and dynamic graph	Short-medium	Spatial coupling	Single temporal encoder
Hybrid	[22], [23], [27]	Spatial modeling support	Multi-scale	Robustness	Lack of a principled mechanism for fusion

Transformer-based models have become increasingly popular for their self-attention mechanisms, which effectively capture global temporal dependencies. Several Transformer variants such as Informer [21] and N-BEATS [21] have also been proposed to address computational limits through sparse attention and interpretable structures, respectively. These models achieve strong results in renewable energy forecasting, particularly when combined in hybrid modeling setups [22], [23]. GNNs such as GATs have also improved temporal modeling. GATs identify spatial relationships among dispersed sites. Several studies utilize GATs in hybrid modeling setups [24]-[26] and have shown that, when supplemented with additional environmental data such as weather forecasts, sensor data, and sky images, forecasting accuracy significantly improves [29], [30].

Despite advances, the literature shows that most of the existing forecasting models are narrowly scoped, focusing on a single energy source, and lack generalizability and uncertainty quantification. Few studies unify spatio-temporal modeling within a scalable, interpretable framework. This study proposes an ST-HGT model, which integrates GATs, TCNs, N-BEATS, Transformer, and Informer. Unlike prior work, the proposed ST-HGT model forecasts wind and solar power across multiple sites without additional tuning. It autonomously constructs graphs based on spatial and temporal correlations, enabling probabilistic forecasting. GAT is employed for attention-weighted neighborhood aggregation over the station graph [31], a Transformer encoder-decoder is used for long-range temporal dependencies [32], the Informer is used to handle long-sequence modeling via ProbSparse attention efficiently [21], TCN is utilized for handling local and abrupt dynamics via dilated causal convolutions [33], and N-BEATS is utilized to capture the trend and seasonality decomposition [21]. Rather than relying on a single architecture, we combine these modules with complementary inductive biases and learn a convex fusion. This improves sample efficiency and stability in our setting while preserving cross-site spatial coupling via GAT.

Recent studies use GNNs on energy time-series in various patterns: static graphs based on geography or topology with graph convolutional networks (GCNs) or GATs; dynamic graphs with time-varying edges; physics-aware graphs respecting grid structure; and task-specific designs for ramps and extremes [24]-[26]. These enhance spatial generalizations that rely on a single temporal block, missing short spikes or seasonal patterns. The proposed ST-HGT model retains spatial gains via GAT, paired with temporal experts and learned fusion. Our ablation results verify that this structure reduces error compared with that of single component.

II. Data Acquisition and Preprocessing

We used daily wind and solar power data (2020-2023) from the Korea Power Exchange [34] and meteorological data from the Korea Meteorological Administration [35], covering five wind farms (Gasiri, Gimnyeong, Dongbok, Sinchang, and Haengwon), four solar farms (Gyorae, Sport complex, Hongbogwan, and Hangwon), and four weather stations (Gosan, Jeju-si, Seongsan, and Seogwipo). Wind power data were from supervisory control and data acquisition (SCADA). The graph-based representation of station connectivity is shown in Fig. 1.

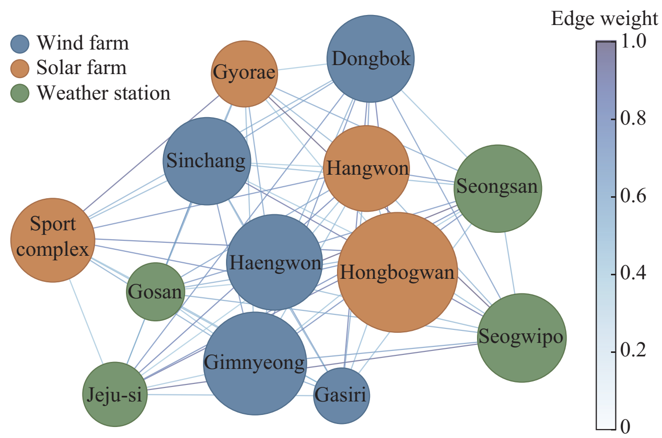

Fig. 1. Graph-based representation of station connectivity.

Solar power data included solar power output and grid feed-in. Environmental features (e.g., wind speed, irradiance, temperature, humidity) were standardized, and wind and solar power was normalized by site capacity. Missing values were imputed, outliers were removed, and data were aligned with a unified daily index aggregating hourly weather data. Using mutual information of wind and solar power data, as shown in Figs. 2 and 3, respectively, we define 7-day and 4-day input windows for wind and solar power, each with a 1-day forecasting horizon.

Note that the first local min in Figs. 2 and 3 denotes the earliest lag where the normalized mutual information (NMI) reaches a local trough, beyond which historical data provide diminishing predictive value and thus define the optimal input window length. Sliding windows with 1-day steps create overlapping sequences for model training. Daily power generation data of all sites are provided by the Korea Power Exchange. Hourly SCADA data are not public. Therefore, our forecasting target is the daily energy output for the subsequent day, and hourly weather data are aggregated into daily data. The details of data acquisition and preprocessing process are shown in Supplementary Material A Fig. SA1.

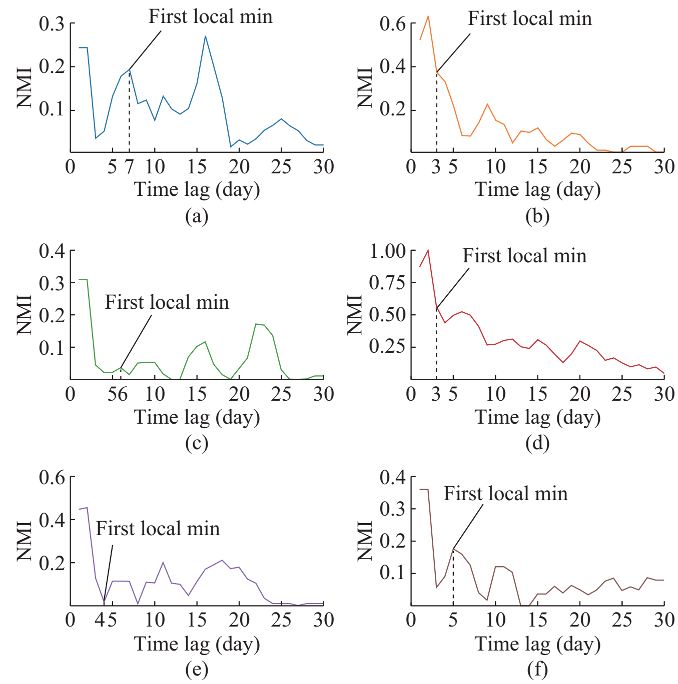

Fig. 2. Mutual information of wind power data. (a) Gasiri. (b) Gimnyeong. (c) Dongbok. (d) Sinchang. (e) Haengwon. (f) Total wind power.

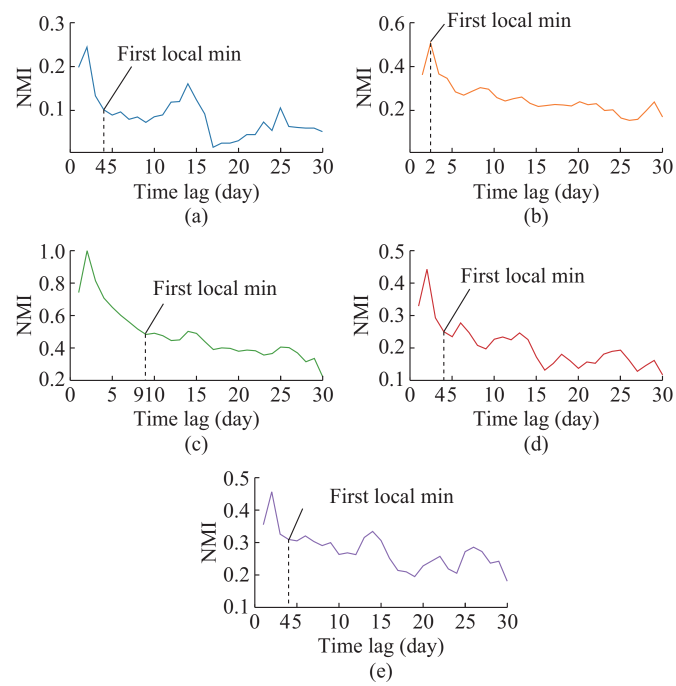

Fig. 3. Mutual information of solar power data. (a) Gyorae. (b) Sport complex. (c) Hongbogwan. (d) Hangwon. (e) Total solar power.

III. Proposed ST-HGT Model

The proposed ST-HGT model, with its details shown in Supplementary Material A Fig. SA2, fuses spatial embeddings from a GAT with four temporal models, i.e., Transformer, Informer, TCN, and N-BEATS, each capturing different temporal scales. The GAT learns inter-site dependencies via a correlation-informed graph and multi-head attention. Module outputs $\boldsymbol{F}_i^{(m)}$ and GAT embeddings $\boldsymbol{Z}_i^{(\mathrm{Spatial})}$ are fused via attention and passed to a fully connected layer for forecasting. We employ a learned multi-expert fusion expressed in (1) and a causal correlation-distance hybrid adjacency for spatial encoding expressed in (2)-(4). We train separate models for wind and solar power forecasting (no weight sharing) to preserve resource-specific accuracy.

$$\hat{y}_{i,t}=FC\left(Attention\left[\boldsymbol{Z}_i^{(\mathrm{Spatial})},\boldsymbol{F}_i^{(\mathrm{Trans})},\boldsymbol{F}_i^{(\mathrm{Inf})},\boldsymbol{F}_i^{(\mathrm{TCN})},\boldsymbol{F}_i^{(\mathrm{N\text{-}BEATS})}\right]\right) \tag{1a}$$

$$\alpha_m=\operatorname{softmax}\left(\theta_{F,m}\right) \tag{1b}$$

where $i\in\{1,2,\ldots,N\}$ is the station index, and N is the number of stations; t is the time step; $\hat{y}_{i,t}$ is the forecasted power for station i at time t; $FC(\cdot)$ denotes the final fully connected prediction layer; $Attention[\cdot]$ denotes the fusion attention operator; $\boldsymbol{Z}_i^{(\mathrm{Spatial})}$ denotes the GAT-derived spatial embedding of station i; $\boldsymbol{F}_i^{(m)}$ denotes the feature output of temporal expert m for station i, and $m\in\{$Transformer, Informer, TCN, N-BEATS$\}$; α_m is the fusion weight of expert m and $\sum_m \alpha_m=1$; and $\theta_{F,m}$ is the fusion logit for expert m.

Each temporal block carries a distinct inductive bias: TCN captures short-range spikes via causal dilated convolutions; N-BEATS isolates trend and seasonality; Transformer models medium-long dependencies; Informer scales to very long sequences; and GAT injects cross-site context. We fuse their outputs with nonnegative fusion weights α_m. Under squared loss, the ensemble error decomposes as $\mathbb{E}\left[\left(\sum_m \alpha_m\varepsilon_m\right)^2\right]=\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\alpha}$, where ε_m is the residual of expert m, $\boldsymbol{\alpha}=[\alpha_m]$ is the fusion weight vector, and $\boldsymbol{\Sigma}$ is the residual covariance matrix. When residuals are not perfectly correlated, optimizing $\boldsymbol{\alpha}$ reduces variance relative to any single model (denoted as the classical ensemble effect). End-to-end training, therefore, down-weights a block when it is off-manifold for a given station or week while retaining the coverage of complementary frequency bands. Ablation studies show consistent degradation when any block is removed, with the largest degradation for GAT, supporting the complementarity of the spatial and multi-scale temporal encoders. The proposed ST-HGT model is trained end-to-end, combining normalization, dynamic graph construction, and parallel temporal encoding. Dropout-based ensembling yields uncertainty quantification, evaluated via prediction interval coverage probability (PICP) and prediction interval width (PIW), where high PICP and low PIW indicate calibrated and sharp intervals. The performance of the proposed ST-HGT model is assessed using root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2) (with details in Supplementary Material A Algorithm SA1). The ablation confirms the centrality of GAT and the complementarity of temporal modules, where V is the vertex set, and E is the edgest.

An undirected graph $G=(V,E)$ is constructed to model spatial dependencies, where V is the vertex set, and E is the edg set. The graph has $N=13$ nodes, including five wind farms, four solar plants, and four weather stations. Edges indicate spatial proximity (within 20 km) and historical correlation, with weights assigned via exponential decay. The symmetric adjacency matrix $\boldsymbol{A}\in\mathbb{R}^{N\times N}$ with self-loops $A_{ii}=1$ enables GAT-based message passing.

$$A_{ij}^{(\mathrm{dist})}=\exp\left(-\frac{d_{ij}}{\sigma_d}\right) \tag{2}$$

where $A_{ij}^{(\mathrm{dist})}$ is the distance-based spatial adjacency weight between station i and station j, which decays exponentially with increasing geographic distance; σ_d is the spatial scaling factor; and d_{ij} is the spatial distance between station i and station j. A spatial scaling factor $\sigma_d=20$ km ensures that nearby nodes receive higher weights. To capture long-range dependencies, Pearson correlations ρ_{ij} are computed between historical power or weather data. Stations i and j are linked if $\rho_{ij}>\tau$, with Pearson correlation threshold $\tau=0.6$ ensuring statistical relevance. Correlation-based edge weights are defined as:

$$A_{ij}^{(\mathrm{corr})}=\begin{cases}\rho_{ij} & \rho_{ij}>\tau\\ 0 & \text{otherwise}\end{cases} \tag{3}$$

Unlike prior models that rely solely on spatial proximity, the proposed ST-HGT model combines spatial and correlation-based edges to capture geographic and meteorological relationships. Final edge weights are computed as:

$$A_{ij}=\alpha A_{ij}^{(\mathrm{dist})}+(1-\alpha)A_{ij}^{(\mathrm{corr})} \tag{4}$$

The coefficient α balances distance-based and correlation-based edge weights in the final adjacency matrix $\boldsymbol{A}$. We select α via cross-validation on the training/validation split, with a grid search of $\{0.3, 0.5, 0.7, 0.9\}$. Performance is robust for $\alpha\in[0.6,0.8]$; and we use $\alpha=0.7$ to favor spatial continuity on the compact geography of Jeju Island while retaining adaptivity to correlations. With $\alpha=0.7$, spatial and correlation-based edge weights are combined to construct the adjacency matrix, highlighting coherent regional clusters. Wind and solar are weather-driven; GAT uses resource-specific covariates for inter-site context, and temporal models are fused at the site level. Models are trained separately per resource (no weight sharing), with normalization and early stopping to avoid trade-offs in accuracy. Each station i receives a sequence $\boldsymbol{x}_i\in\mathbb{R}^{W\times F_i}$, reshaped into $\boldsymbol{X}\in\mathbb{R}^{N\times(WF_i)}$ for GAT-based spatio-temporal modeling, where W is the input-window length, and F_i is the number of input features at station i.

$$\boldsymbol{H}^{(0)}(t)=\boldsymbol{X}(t)\in\mathbb{R}^{N\times F_i} \tag{5}$$

where $\boldsymbol{H}^{(0)}$ is the initial hidden feature matrix, and the subscript 0 indicates the 0^{th} GAT layer. The GAT layer output is given as:

$$\boldsymbol{H}^{(l)}(t)=\left[\boldsymbol{h}_1^{(l)}(t),\boldsymbol{h}_2^{(l)}(t),\ldots,\boldsymbol{h}_N^{(l)}(t)\right]^{\mathrm{T}} \tag{6}$$

where $l\in\{0,1,\ldots,L\}$ is the GAT layer index. The feature of each station is updated as (7), and $\mathcal{N}(i)$ includes station i itself.

$$\boldsymbol{h}_i^{(l+1)}(t)=\sigma\left(\sum_{j\in\mathcal{N}(i)}\alpha_{ij}^{(l)}(t)\boldsymbol{W}^{(l)}\boldsymbol{h}_j^{(l)}(t)\right) \tag{7}$$

where $\sigma(\cdot)$ is the ReLU activation function; $\boldsymbol{W}^{(l)}$ is the learnable linear transformation matrix at the l^{th} GAT layer; $\boldsymbol{h}_j^{(l)}$ is the hidden feature vector of station j at the l^{th} GAT layer;

and $\alpha_{ij}^{(l)}(t)$ is the attention coefficient and can be computed as:

$$e_{ij}^{(l)}(t)=\text{LeakyReLU}\left(\left(\boldsymbol{a}^{(l)}\right)^{\mathrm{T}}\left[\boldsymbol{W}^{(l)}\boldsymbol{h}_i^{(l)}(t)\|\boldsymbol{W}^{(l)}\boldsymbol{h}_j^{(l)}(t)\right]\right) \tag{8}$$

$$\alpha_{ij}^{(l)}(t)=\frac{\exp\left(e_{ij}^{(l)}(t)\right)}{\sum_{k\in\mathcal{N}(i)}\exp\left(e_{ik}^{(l)}(t)\right)} \tag{9}$$

where $\boldsymbol{a}^{(l)}$ is the learnable attention vector at the l^{th} GAT layer; and $\|$ denotes the concatenation.

Equation (9) ensures normalized attention weights across neighbors. Two GAT layers are stacked for capturing higher-order interactions. Each stack uses $K=4$ attention heads. The intermediate GAT layers concatenate the multi-head outputs as:

$$\boldsymbol{h}_i^{(l+1)}(t)=\Big\|_{k=1}^{K}\sigma\left(\sum_{j\in\mathcal{N}(i)}\alpha_{ij}^{(l),k}(t)\boldsymbol{W}^{(l),k}\boldsymbol{h}_j^{(l)}(t)\right) \tag{10}$$

where $\Big\|_{k=1}^{K}$ denotes the concatenation over K attention heads; and k is the index of attention head.

The final GAT layer is averaged as:

$$\boldsymbol{h}_i^{(L)}(t)=\frac{1}{K}\sum_{k=1}^{K}\sigma\left(\sum_{j\in\mathcal{N}(i)}\alpha_{ij}^{(L),k}(t)\boldsymbol{W}^{(L),k}\boldsymbol{h}_j^{(L-1)}(t)\right) \tag{11}$$

The multi-head GAT learns from a range of spatial patterns including local influence, meteorological similarity, and correlated variability across multiple granularities. Meanwhile, GCNs employ fixed normalization as:

$$\tilde{\boldsymbol{A}}=\boldsymbol{D}_{\deg}^{-\frac{1}{2}}(\boldsymbol{A}+\boldsymbol{I})\boldsymbol{D}_{\deg}^{-\frac{1}{2}} \tag{12}$$

where $\tilde{\boldsymbol{A}}$ is the normalized adjacency matrix; and $\boldsymbol{D}_{\deg}$ is the diagonal degree matrix of $\boldsymbol{A}+\boldsymbol{I}$, and $\boldsymbol{I}$ is the identity matrix. Attention weights are computed by the GAT, enabling context-aware neighbor weighting and filtering out noisy nodes. This allows the proposed ST-HGT model to assign higher weights to informative stations and lower weights to noisy or temporarily unreliable stations. For example, during a localized weather anomaly, the proposed ST-HGT model can reduce the influence of stations that are geographically close but exhibit inconsistent temporal behavior. After passing through L GAT layers, the final node embedding is expressed as:

$$\boldsymbol{z}_i=\boldsymbol{h}_i^{(L)} \tag{13}$$

The embeddings capture local history and spatial context, enabling cross-site knowledge sharing. For example, storm impacts at one site inform forecasts at nearby locations. Correlation-based graph design and attention-driven propagation produce context-rich spatial features that improve forecasting accuracy.

IV. Hybrid Modeling Architecture of ST-HGT Model

The architecture of the proposed GAT, as shown in Fig. 4, integrates a GAT-based spatial encoder with four temporal models: Transformer, Informer, TCN, and N-BEATS. The GAT processes site-level sequences $\boldsymbol{x}_i\in\mathbb{R}^{W\times F_i}$, comprising power and weather features, to generate $\boldsymbol{z}_i\in\mathbb{R}^{W\times F_z}$ that capture inter-site correlations, with F_z as the spatial-embedding dimension. For clarity, we define the forecast horizon as $H=1$ throughout this paper, corresponding to one-step-ahead prediction. These context-aware embeddings are input to the temporal modules for 1-day-ahead prediction ($H=1$). The Transformer employs an encoder-decoder with self-attention and positional encoding to capture long-range dependencies in W. The attention mechanism is given by:

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{soft}\max\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_{\mathrm{k}}}}\right)\boldsymbol{V} \tag{14}$$

where $\boldsymbol{Q}$, $\boldsymbol{K}$, and $\boldsymbol{V}$ are the query, key, and value projections, respectively; and d_{k} is the key dimension (set to be 64). The decoder generates one-shot forecasts $\hat{y}_{i,t_{W+1}}$ by attending to encoder outputs and $\boldsymbol{z}_{i,t_W}$, and t_W indicates the last time step in the input window, while t_{W+1} denotes the next time step to be forecasted (one-step-ahead).

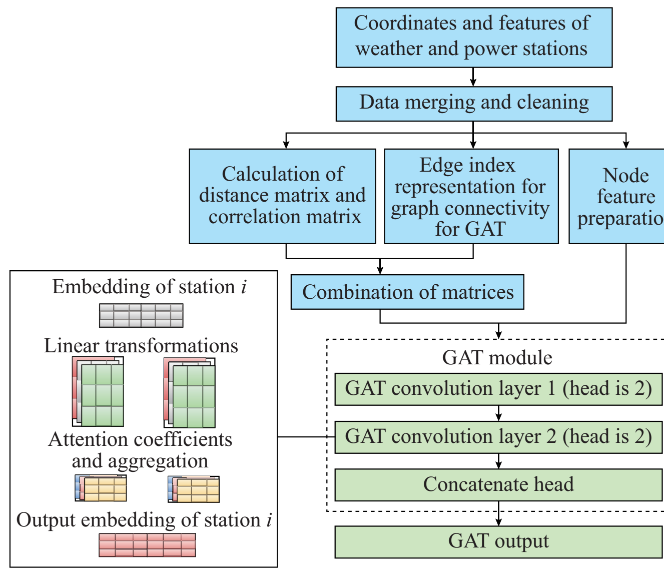

Fig. 4. Architecture of proposed GAT.

Two encoder layers and four attention heads, as shown in Supplementary Material A Fig. SA3, offer a trade-off between accuracy and efficiency. To improve scalability, we integrate Informer—a Transformer variant using ProbSparse attention—to reduce quadratic time complexity $O(L^2)$ to $O(L\ln L)$, where L is the sequence length. It selects dominant queries via Kullback-Leibler (KL) divergence and uses a generative decoder for one-shot output, as shown in Supplementary Material A Fig. SA4. To capture local temporal patterns, we use a TCN with stacked dilated causal convolutions. Unlike RNNs, TCNs support full parallelism, improving efficiency. The output feature of station i at the l^{th} GAT layer at time t is:

$$h_{i,t}^{(l)}=f\left(y_{i,t},y_{i,t-d},\dots,y_{i,t-(k-1)d}\right) \tag{15}$$

where $f(\cdot)$ is the convolutional filter; the subscript k denotes the kernel size; and the subscript d denotes the dilation factor. The receptive field is defined as:

$$RF=1+2(k-1)d \tag{16}$$

Residual connections stabilize training, enabling deep coverage with few layers. The TCN output $\hat{\boldsymbol{y}}_i^{(TCN)}$ captures short-term cycles and abrupt shifts, as shown in Supplementary Material A Fig. SA5. To capture trend and seasonality, we use N-BEATS, which is a fully-connected and block-wise model, to produce backcast and forecast outputs per block. Given the input $\boldsymbol{x}^{(1)}=\boldsymbol{y}_i$, each block b refines residuals as follows. The backcast of block b $backcast^{(b)}$ and forecast of block b $forecast^{(b)}$ are generated by applying its transformation functions $B^{(b)}$ and $F^{(b)}$ to the block input $\boldsymbol{x}^{(b)}$, respectively. The backcast is passed to the next block as the new input, while the forecast contributes to the final prediction output.

$$\begin{cases} backcast^{(b)}=B^{(b)}\left(\boldsymbol{x}^{(b)}\right) \\ forecast^{(b)}=F^{(b)}\left(\boldsymbol{x}^{(b)}\right) \end{cases} \tag{17}$$

$$\boldsymbol{x}^{(b+1)} \quad =\boldsymbol{x}^{(b)}-backcast^{(b)} \tag{18}$$

The final prediction aggregates all forecast terms as:

$$\hat{\boldsymbol{y}}_i^{(\text{N-BEATS})}=\sum_{b=1}^{B}F^{(b)}\left(\boldsymbol{x}^{(b)}\right) \tag{19}$$

N-BEATS is configured with $B=2$ blocks, each containing two fully connected layers. The feed forward network (FFN) employs a hidden width of 256 and uses the ReLU activation function. As shown in Fig. 5, this setup efficiently captures trend and seasonal components. All temporal modules receive the same GAT-enhanced input and independently generate forecasts, which are fused via attention to produce robust, high-resolution, and context-aware predictions. Note that in Fig. 5, the minus signs represent residual calculation for iterative feature learning, and the plus signs represent fusion of prediction components from blocks and stacks to generate the final model output.

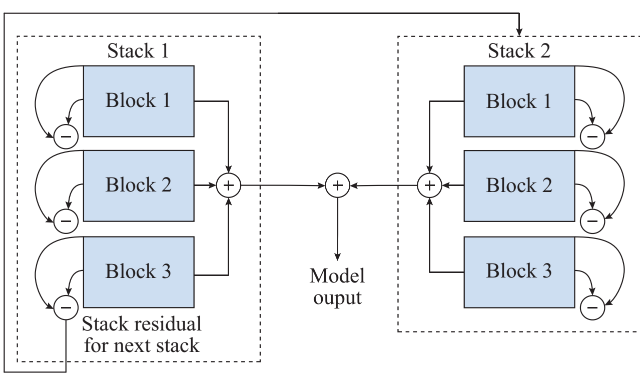

Fig. 5. Architecture of N-BEATS.

V. Implementation and Experimental Setup

We train separate models for wind and solar power forecasting, with distinct parameters. The experiments use an NVIDIA RTX 4080 GPU, an Intel i5-12600K CPU, and 24 GB RAM, which speed up training and remain efficient on lower-end GPUs, with increased moderate runtime. The implementation uses Python 3.10 and PyTorch 2.0 (CUDA 11.8), with hyperparameters tuned via Optuna. The training employs Adam and regularization, usually converging within 500 epochs. All baselines share preprocessing, input window, horizon, split, mean square error (MSE) loss, optimizer, early stopping, and tuning. Windows are strictly forward-looking, with hyperparameters tuned on validation tail to prevent leakage. Models without spatial coupling are trained at each station using the same covariates.

VI. Results and Discussion

We evaluate the proposed ST-HGT model on real-world wind and solar datasets from Jeju Island. The same architecture is applied across both domains without modification, demonstrating strong generalization. The forecasting performance of the proposed ST-HGT model is assessed using MSE, RMSE, MAE, R^2, PICP, and PIW. Additional validation includes station-wise analysis, ablations, baseline comparisons, and residual diagnostics. We compare against major temporal forecasting models, i. e., LSTM, GRU, BiLSTM, TCN, N-BEATS, Transformer, temporal fusion Transformer (TFT), and Informer. All models use identical inputs, window, horizon, splits, MSE loss, optimizer, early stopping, and tuning. Models without spatial operators are trained per station, whereas the proposed ST-HGT model uses all stations via the graph.

A. Station-wise Performance on Wind Power Forecasting

To evaluate spatial generalization, we assess the station-wise performance across five wind farms. Despite training on aggregated data, the proposed ST-HGT model captures station-specific patterns without tuning.

1) Prediction Accuracy

Table II presents station-wise metrics for wind farms, with four of five stations achieving $R^2>0.995$ and RMSE below 1500 kWh. Gasiri and Dongbok show near-perfect accuracy ($R^2=0.9997$ and 0.9999), while Sinchang yields the lowest RMSE (341 kWh) and MAE (250 kWh). Gimnyeong performs slightly lower $R^2=0.9689$, likely due to coastal wind complexity, but remains within acceptable accuracy bounds.

TABLE II
Station-wise Metrics for Wind Farms

Wind farm	MSE (kWh^2)	RMSE (kWh)	MAE (kWh)	R^2
Gasiri	848378.9	921.07	587.69	0.9997
Gimnyeong	174590.3	417.84	313.29	0.9689
Dongbok	1999021.1	1413.87	1048.78	0.9999
Sinchang	116495.5	341.31	250.13	0.9957
Haengwon	499624.7	706.84	558.88	0.9998

Figure 6 shows close alignment between actual and forecasted wind power generation for each wind farm, effectively capturing both short-term fluctuations and overall trends. The slightly higher error at Gimnyeong reflects its high coastal variability.

2)Residual Diagnostics

Figure 7 shows that the residual distributions are roughly symmetric around zero. Gasiri, Dongbok, and Sinchang have tightly clustered residuals, while the residuals in Gimnyeong are slightly skewed but acceptable. RMSE-to-MAE ratios (1.3-1.6), which can be calculated based on Table II, indicate infrequent and non-systematic large errors. Note that in Fig. 7, the red line represents the normal distribution curve.

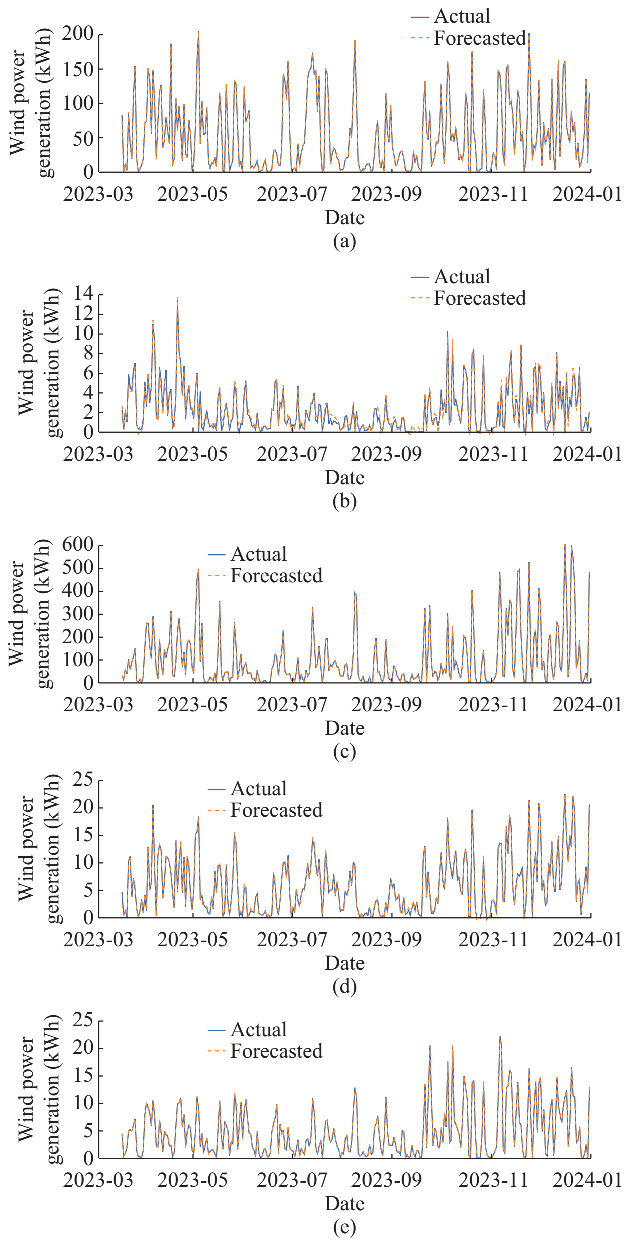

Fig. 6. Actual and forecasted wind power generation for each wind farm. (a) Gasiri. (b) Gimnyeong. (c) Dongbok. (d) Sinchang. (e) Haengwon.

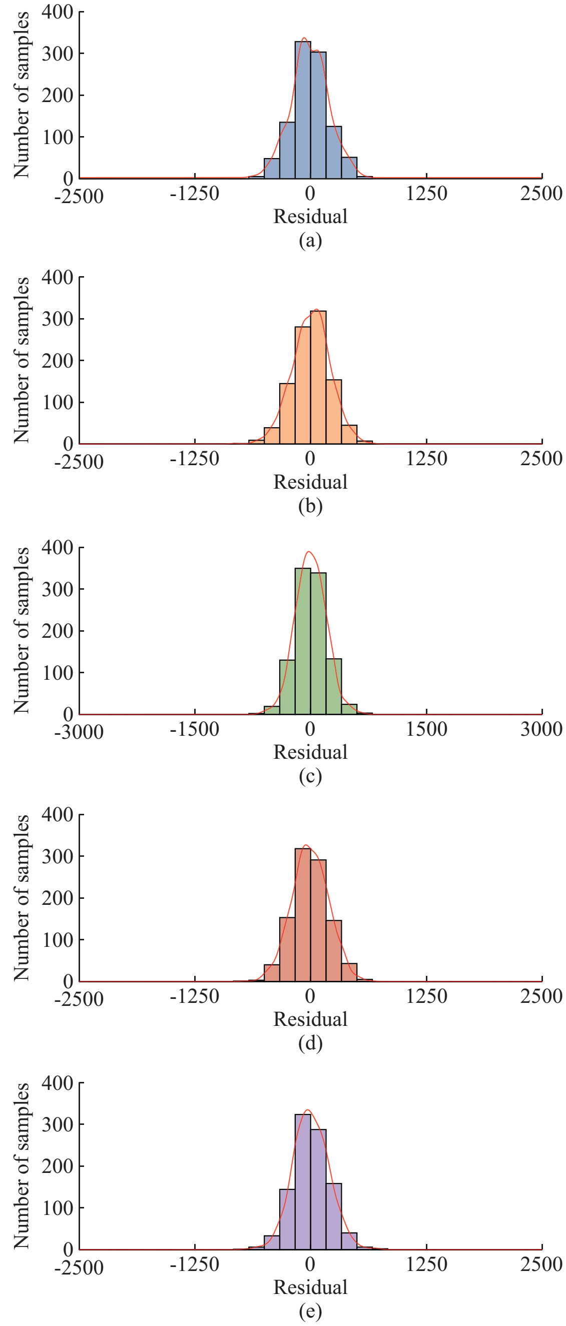

Fig. 7. Residual distribution for each wind farm. (a) Gasiri. (b) Gimnyeong. (c) Dongbok. (d) Sinchang. (e) Haengwon.

Residual autocorrelation function (ACF) for each wind farm, as shown in Fig. 8, demonstrates near-zero time lag correlations, signifying that residuals are uncorrelated over time. Minor lag-1 autocorrelations observed at Gasiri and Gimnyeong diminish rapidly, confirming effective temporal modeling.

3) Calibration Analysis

In Fig. 9, the dashed line represents the ideal calibration performance (the mean actual value equals the mean forecasted value), and the solid line with points represents the actual calibration performance of the proposed ST-HGT model. Figure 9 shows a calibration curve mostly aligning with the 45-degree line, indicating good calibration. Gimnyeong has minor undercoverage at high quantiles, likely due to sparse peak data.

4) Ablation Study Result

We remove one component at a time while keeping other settings fixed and retrain the proposed ST-HGT model under the same setup. Table III presents ablation results for wind power forecasting. It is observed that excluding GAT causes the largest performance drop, with R^2 declining from 0.99 to 0.90, highlighting the importance of GAT in spatial modeling. N-BEATS significantly impacts trend and seasonality detection. TCN, Transformer, and Informer add some temporal depth with smaller effects. Validation error remains stable across $\alpha \in [0.6, 0.8]$, so we choose $\alpha = 0.7$ for all experiments.

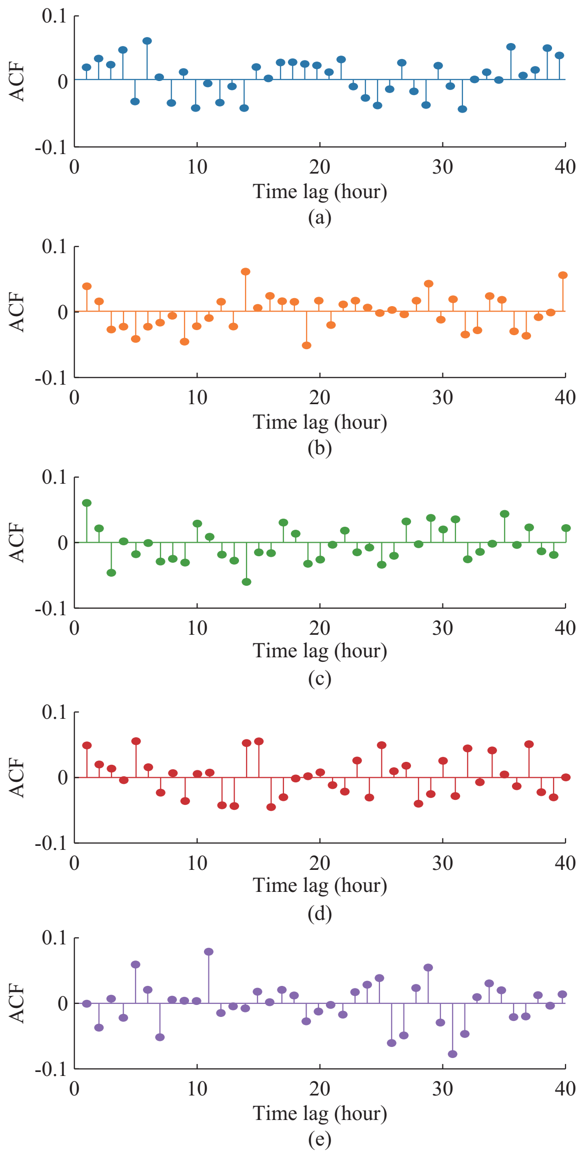

Fig. 8. Residual ACF for each wind farm. (a) Gasiri. (b) Gimnyeong. (c) Dongbok. (d) Sinchang. (e) Haengwon.

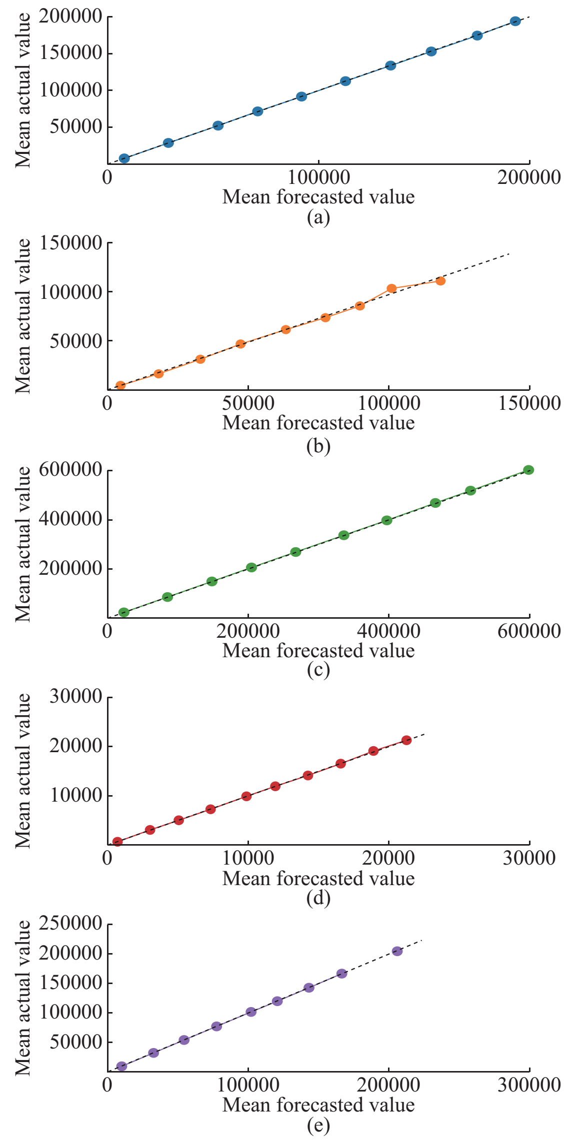

Fig. 9. Calibration curves for each wind farm. (a) Gasiri. (b) Gimnyeong. (c) Dongbok. (d) Sinchang. (e) Haengwon.

5) *Comparison Against Baseline Models*

Table IV shows the performance comparison of forecasting models. The proposed ST-HGT model surpasses all models and achieves an RMSE reduction of 98% and an R^2 exceeding 0.99. Despite tuning, strong temporal models (i. e., Informer, Transformer, TFT, TCN, N-BEATS, and RNNs) achieve only 0.51-0.66 R^2, showing that single-station models explain much but not all variance. This highlights the need for a hybrid architecture that models spatial and multi-scale temporal dynamics.

B. *Station-wise Performance Analysis for Solar Power Forecasting*

To assess the robustness across domains, the proposed ST-HGT model is used for solar power forecasting with identical architecture and training as for wind power forecasting, which is evaluated at four solar farms: Gyorae, Sport complex, Hongbogwan, and Hangwon, each with unique temporal and environmental profiles.

TABLE III
ABLATION STUDY RESULTS FOR WIND POWER FORECASTING

Scenario	MSE (kWh^2)	RMSE (kWh)	MAE (kWh)	R^2
Without TCN	379725760	19486.55	12130.89	0.95
Without Transformer	396963680	19923.95	12308.16	0.95
Without Informer	394497440	19861.96	12164.47	0.95
Without GAT	722046464	26870.92	16748.08	0.90
Without N-Beats	427897568	20685.69	12583.72	0.94
Full ST-HGT	727622	853.00	551.00	0.99

1) *Forecasting Performance and Residual Analysis*

Figure 10 shows that the proposed ST-HGT model effectively captures daily solar power fluctuations across solar farms. Table V shows the performance metrics for solar power forecasting for each solar farm. Table V summarizes the results with R^2 of 0.99 and RMSEs from 1317 kWh (Hongbogwan) to 30.97 kWh (Gyorae).

TABLE IV
PERFORMANCE COMPARISON OF FORECASTING MODELS FOR WIND POWER FORECASTING

Model	MSE (kWh^2)	RMSE (kWh)	MAE (kWh)	R^2
Informer	3546919000	59556.46	35662.61	0.62
Transformer	3518843000	59320.08	35765.50	0.63
N-BEATS	3686202000	60713.61	35900.81	0.61
TCN	3927000000	62665.87	31957.20	0.56
TFT	3338051000	57775.87	28619.62	0.66
BiLSTM	3723762000	61022.63	32126.85	0.60
GRU	3797490000	61623.92	34292.75	0.57
LSTM	3852591000	62069.28	30912.42	0.51

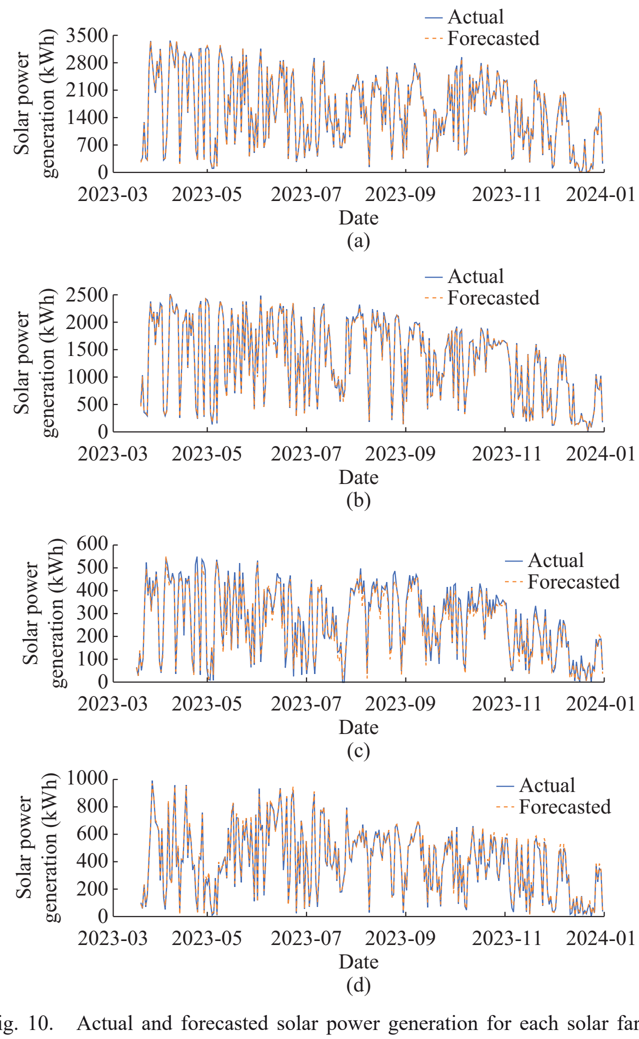

Fig. 10. Actual and forecasted solar power generation for each solar farm. (a) Gyorae. (b) Sport complex. (c) Hongbogwan. (d) Hangwon.

TABLE V
PERFORMANCE METRICS FOR SOLAR POWER FORECASTING

Station	MSE (kWh^2)	RMSE (kWh)	MAE (kWh)	R^2
Gyorae	959.223	30.971	25.607	0.999
Sport complex	435.885	20.878	16.749	0.999
Hongbogwan	173.536	13.173	10.204	0.992
Hangwon	316.654	17.795	14.080	0.995

Figure 11 shows the residual distribution for each solar farm, showing zero-centered, symmetric residuals across solar farms. Note that in Fig. 11, the red line represents the normal distribution curve. Figure 12 shows the residual ACF for each solar farm within 95% bounds. The results indicate that the proposed ST-HGT model is unbiased and consistent with no overfitting.

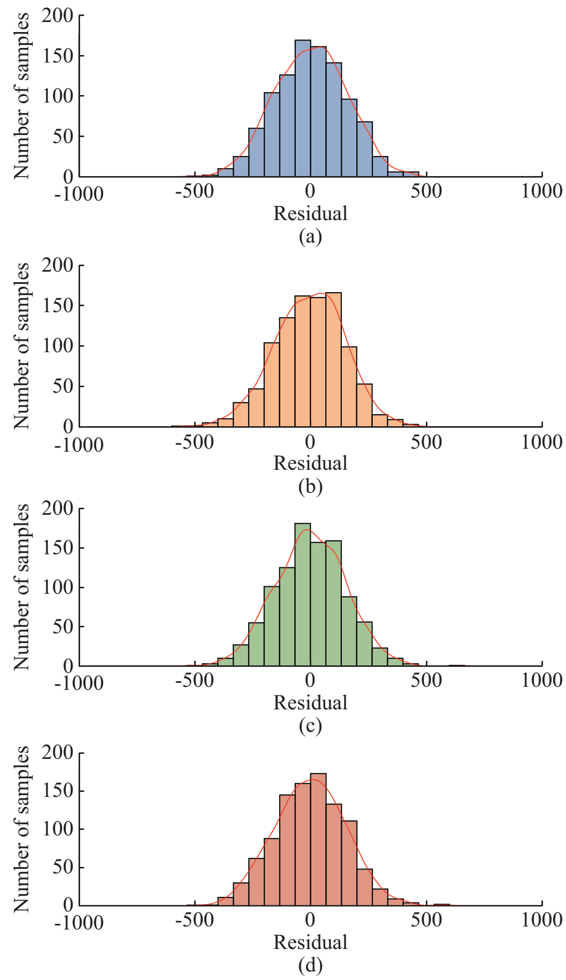

Fig. 11. Residual distribution for each solar farm. (a) Gyorae. (b) Sport complex. (c) Hongbogwan. (d) Hangwon.

2) Calibration Analysis

Figure 13 shows calibration curve for each solar farm. The curve aligns with the 45-degree line, showing accurate quantile estimation without overestimation or underestimation, even at peak intensities, indicating robust distributional calibration.

3) Ablation Study Result

An ablation study is conducted for solar power forecasting, and the results are shown in Table VI. The results show that excluding GAT yields the largest drop in performance, i.e., $RMSE = 336.5$ kWh and $R^2 = 0.848$. Removing N-BEATS significantly degrades the performance, i. e., $RMSE = 258.0$ kWh and $R^2 = 0.911$, confirming its effectiveness in capturing trends and seasonality. Removing TCN, Informer, or Transformer produces moderate decreases, i. e., $R^2 \approx 0.947$ - 0.950. The ablation study thus affirms the complementary importance of all modules.

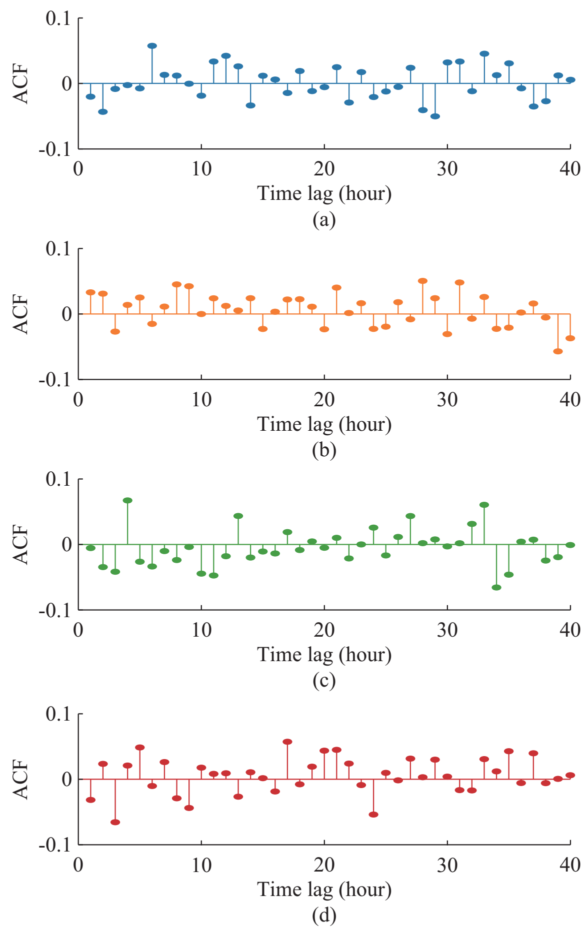

Fig. 12. Residual ACF for each solar farm. (a) Gyorae. (b) Sport complex. (c) Hongbogwan. (d) Hangwon.

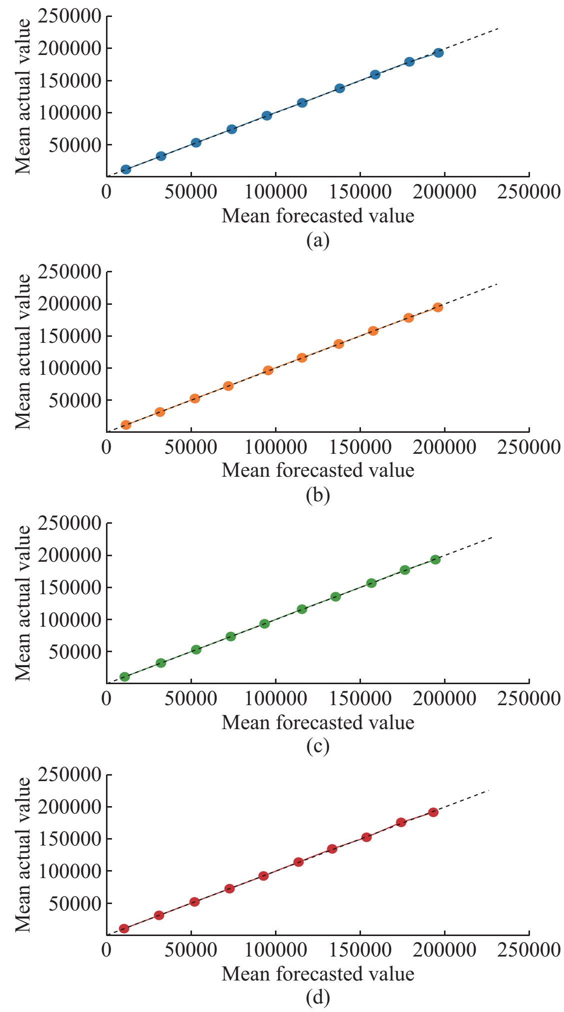

Fig. 13. Calibration curves for each solar farm. (a) Gyorae. (b) Sport complex. (c) Hongbogwan. (d) Hangwon.

4) Comparison Against Baseline Models

In Table VII, the proposed ST-HGT model is compared against all baseline models. The proposed ST-HGT model significantly outperforms the standalone baseline models, achieving an RMSE of 21.71 kWh and an R^2 of 0.999. Among the baseline models, Informer performs best, with an RMSE of 527.64 kWh and an R^2 of 0.685, while other models such as TFT, Transformer, and N-BEATS attain R^2 values in the range of approximately 0.55-0.69. Even after tuning, the baseline models only explain part of the variance, while the proposed ST-HGT model more accurately captures the complex temporal dynamics of solar power.

C. Performance of Probabilistic Forecasting

To quantify the uncertainties of the proposed ST-HGT model, we employ two metrics: PICP and PIW.

1) PICP

PICP measures the proportion of actual values that fall within the bounds of specified prediction interval. Let n denote the total number of predictions, t index the prediction instances, y_t denote the observed (ground-truth) value at instance t, and $\left[\hat{y}_t^{\mathrm{L}}, \hat{y}_t^{\mathrm{U}}\right]$ denote the lower and upper bounds of the prediction interval at instance t, respectively. Then, we can obtain:

TABLE VI
ABLATION STUDY RESULT FOR SOLAR POWER FORECASTING

Scenario	MSE (kWh^2)	RMSE (kWh)	MAE (kWh)	R^2
Without TCN	39712.246	199.279	147.554	0.947
Without Transformer	37165.281	192.783	140.616	0.950
Without Informer	39383.816	198.454	141.885	0.947
Without GAT	113228.195	336.494	238.428	0.848
Without N-BEATS	66562.852	257.998	187.760	0.911
Overall ST-HGT	471.324	21.710	16.660	0.999

$$PICP=\frac{1}{n}\sum_{t=1}^{n}\mathbb{I}\left(y_t \in \left[\hat{y}_t^{\mathrm{L}}, \hat{y}_t^{\mathrm{U}}\right]\right) \tag{20}$$

where $\mathbb{I}(\cdot)$ is the indicator function, returning 1 if the condition is true and 0 otherwise. A higher $PICP$ (close to the nominal coverage such as 95%) indicates more reliable coverage.

2) PIW

PIW measures the average width of the prediction intervals. Smaller intervals are preferred when coverage is adequate, as they indicate sharper predictions.

TABLE VII
PERFORMANCE COMPARISON OF ALL FORECASTING MODELS FOR SOLAR POWER FORECASTING

Model	MSE (kWh^2)	RMSE (kWh)	MAE (kWh)	R^2
Informer	278401	527.64	378.22	0.685
TFT	294512	542.69	392.18	0.650
TCN	335208	579.14	395.47	0.626
N-BEATS	318945	564.75	381.33	0.617
GRU	334876	578.68	405.29	0.596
BiLSTM	362183	601.82	418.56	0.602
Transformer	282746	531.74	380.91	0.631
LSTM	348792	590.59	415.88	0.553

$$PIW=\frac{1}{n}\sum_{t=1}^{n}\left(\hat{y}_t^{\mathrm{U}}-\hat{y}_t^{\mathrm{L}}\right) \tag{21}$$

Together, the PICP and PIW metrics assess the quality of uncertainty quantification. Ideally, high PICP values should be achieved without excessively wide intervals (low PIW), thereby balancing reliability and precision.

3) Wind Power Results

Table VIII presents the PICP and PIW for each wind farm. The proposed ST-HGT model reports PICP values from 0.94 to 0.98, with PIW values scaling with wind power output. Gimnyeong and Haengwon show PICP of 0.98 and PIWs of 850 kWh and 950 kWh, respectively, while larger wind power output of Dongbok results in a broader PIW of 1700 kWh.

TABLE VIII
PICP AND PIW FOR EACH WIND FARM

Wind farm	PICP	PIW (kWh)
Gasiri	0.95	1200
Gimnyeong	0.98	850
Dongbok	0.97	1700
Sinchang	0.94	450
Haengwon	0.98	950

4) Solar Power Results

Solar power forecasting consistently achieves robust performance, as shown in Table IX. All solar farms attain high PICP values (≥ 0.95) with narrow and stable PIWs between 50 and 90 kWh, indicating reliable uncertainty quantification. The results confirm that the proposed ST-HGT model is accurate for point predictions and dependable in uncertainty quantification, supporting its applicability in risk-informed renewable energy forecasting.

TABLE IX
PICP AND PIW FOR EACH SOLAR FARM

Solar farm	PICP	PIW (kWh)
Gyorae	0.98	70
Sport complex	0.98	90
Hongbogwan	0.95	50
Hangwon	0.98	80

5) Stress-case Error Analysis

Errors grow with ramp magnitude: MAE and RMSE increase from median (Ramp D5) to top (Ramp D10) deciles for wind and solar power forecasting, showing that rapid day-to-day changes affect residuals. Higher relative errors are observed during periods with low solar power output (with frequent cloud cover transitions), as well as during the periods with high wind power output and sudden gust peaks. Under extreme weather (top 5% daily wind speed), RMSE rises and intervals widen, but PICP stays near nominal and calibration remains stable. The error analyses for wind and solar power forecasting are presented in Table X and Table XI, respectively.

TABLE X
ERROR ANALYSIS FOR WIND POWER FORECASTING

Stratum	MAE (kWh)	RMSE (kWh)	PICP	PIW (kWh)
Ramp D5 (median)	420	700	0.97	950
Ramp D10 (top)	900	1500	0.95	1900
Low power output	360	600	0.98	700
High power output	700	1200	0.95	1600
Extreme weather	820	1400	0.96	1800

TABLE XI
ERROR ANALYSIS FOR SOLAR POWER FORECASTING

Stratum	MAE (kWh)	RMSE (kWh)	PICP	PIW (kWh)
Ramp D5 (median)	14	20	0.98	60
Ramp D10 (top)	26	38	0.96	90
Low power output	11	16	0.97	55
High power output	22	33	0.96	85
Extreme weather	18	27	0.97	80

VII. CONCLUSION

This paper presents the ST-HGT model for high-fidelity renewable energy forecasting, combining GATs, TCN, N-BEATS, Transformer, and Informer. Experiments demonstrate that the proposed ST-HGT model captures spatial dependencies and temporal dynamics, providing accurate short-term and long-term wind and solar power forecasting. The performance of the proposed ST-HGT model is validated on real datasets from Jeju Island. The proposed ST-HGT model consistently outperforms baseline models on key metrics such as RMSE, MAE, and R^2, and demonstrates high accuracy and robustness across solar and wind power forecasting. The modular architecture enables transfer across tasks without architecture-specific tuning. Ablation studies highlight the importance of the proposed ST-HGT model, especially GAT for spatial understanding and N-BEATS for temporal decomposition, showcasing the advantages of combining diverse spatio-temporal components. Beyond point estimates, the proposed ST-HGT model allows uncertainty-aware forecasting with PICP- and PIW-based probabilistic forecasting. Future work includes improving uncertainty calibration, exploring alternative probabilistic methods, and integrating online or continual learning to adapt to changing grid conditions. We aim to integrate multi-modal data like environmental images, grid measurements, and sensor data to enhance long-term forecasting and spatial generalization. Currently, there is no publicly available hourly power data for the sites, and we plan to test an hourly version once available.

REFERENCES

[1] A. Couto and A. Estanqueiro, "Wind power plants hybridised with solar power: a generation forecast perspective," *Journal of Cleaner Production*, vol. 423, p. 138793, Oct. 2023.

[2] B. F. Hobbs, J. Zhang, H. F. Hamann *et al.*, "Using probabilistic solar power forecasts to inform flexible ramp product procurement for the California ISO," *Solar Energy Advances*, vol. 2, p. 100024, Sept. 2022.

[3] M. Li, M. Yang, Y. Yu *et al.*, "Short-term wind power forecast based on continuous conditional random field," *IEEE Transactions on Power Systems*, vol. 39, no. 1, pp. 2185-2197, Jan. 2024.

[4] M. Yang, D. Wang, and W. Zhang, "A novel ultra short-term wind power prediction model based on double model coordination switching mechanism," *Energy*, vol. 289, p. 130075, Feb. 2024.

[5] R. H. Shumway and D. S. Stoffer, "ARIMA models," in *Time Series Analysis and Its Applications: With R Examples*. New York: Springer International Publishing, 2017, pp. 75-163.

[6] D. R. Dash, P. Dash, and R. Bisoi, "Short term solar power forecasting using hybrid minimum variance expanded RVFLN and sine-cosine levy flight PSO algorithm," *Renewable Energy*, vol. 174, pp. 513-537, Aug. 2021.

[7] A. Alorf and M. U. G. Khan, "Solar irradiance forecasting using temporal fusion transformers," *International Journal of Energy Research*, vol. 2025, no. 1, p. 3534500, Feb. 2025.

[8] G. Hou, Q. Li, and C. Huang, "Spatiotemporal forecasting using multigraph neural network assisted dual domain transformer for wind power," *Energy Conversion and Management*, vol. 325, p. 119393, Feb. 2025.

[9] B. Xiong, Y. Chen, D. Chen *et al.*, "Deep probabilistic solar power forecasting with transformer and Gaussian process approximation," *Applied Energy*, vol. 382, p. 125294, Mar. 2025.

[10] C. Tian, T. Niu, and T. Li, "Developing an interpretable wind power forecasting system using a transformer network and transfer learning," *Energy Conversion and Management*, vol. 323, p. 119155, Jan. 2025.

[11] T. Bashir, H. Wang, M. Tahir *et al.*, "Wind and solar power forecasting based on hybrid CNN-ABiLSTM, CNN-transformer-MLP models," *Renewable Energy*, vol. 239, p. 122055, Feb. 2025.

[12] X. Ye, C. Liu, X. Xiong *et al.*, "Recurrent attention encoder-decoder network for multi-step interval wind power prediction," *Energy*, vol. 315, p. 134317, Jan. 2025.

[13] A. K. Nayak, K. C. Sharma, R. Bhakar *et al.*, "Probabilistic online learning framework for short-term wind power forecasting using ensemble bagging regression model," *Energy Conversion and Management*, vol. 323, p. 119142, Jan. 2025.

[14] Q. W. Khan, A. N. Khan, B. Misbah *et al.*, "Dynamic energy cost prediction for industrial and commercial buildings using advanced machine learning and multi-pricing models," *Journal of Building Engineering*, vol. 103, p. 112055, Jun. 2025.

[15] M. A. Iqbal, J. M. Gil, and S. K. Kim, "Attention-driven hybrid ensemble approach with Bayesian optimization for accurate energy forecasting in Jeju Island's renewable energy system," *IEEE Access*, vol. 13, pp. 7986-8010, Jan. 2025.

[16] Y. Li, G. Liu, Y. Cao *et al.*, "WNPS-LSTM-Informer: a hybrid stacking model for medium-term photovoltaic power forecasting with ranked feature selection," *Renewable Energy*, vol. 244, p. 122687, May 2025.

[17] J. Zhu and Y. He, "A novel hybrid model based on evolving multiquantile long and short-term memory neural network for ultra-short-term probabilistic forecasting of photovoltaic power," *Applied Energy*, vol. 377, p. 124601, Jan. 2025.

[18] M. A. Iqbal, J. Kim, I. Han *et al.*, "Attention-driven feature fusion integrating swin transformer and CNN models for improved ocular disease classification," in *Proceedings of the 2024 International Conference on Engineering and Emerging Technologies*, virtual event, Mar. 2024, pp. 1-6.

[19] E. Ahn and J. Hur, "A short-term forecasting of wind power outputs using the enhanced wavelet transform and ARIMAX techniques," *Renewable Energy*, vol. 212, pp. 394-402, Aug. 2023.

[20] X. Liu, Z. Lin, and Z. Feng, "Short-term offshore wind speed forecast by seasonal ARIMA–a comparison against GRU and LSTM," *Energy*, vol. 227, p. 120492, Jul. 2021.

[21] H. Zhou, S. Zhang, J. Peng *et al.*, "N-BEATS: neural basis expansion analysis for interpretable time series forecasting," in *Proceedings of the 35th AAAI Conference on Artificial Intelligence*, virtual event, May 2021, pp. 11106-11115.

[22] Z. Jin, X. Fu, L. Xiang *et al.*, "Informer learning framework based on secondary decomposition for multi-step forecast of ultra-short term wind speed," *Engineering Applications of Artificial Intelligence*, vol. 139, p. 109702, Jan. 2025.

[23] L. F. R. Agottani, M. M. Girardi, W. R. de Souza Chaves *et al.*, "Multi-step ahead wind power forecasting based on N-BEATS model," in *Proceedings of 2024 IEEE PES Innovative Smart Grid Technologies Europe*, Dubrovnik, Croatia, Oct. 2024, pp. 1-5.

[24] H. Qiu, K. Shi, R. Wang *et al.*, "A novel temporal-spatial graph neural network for wind power forecasting considering blockage effects," *Renewable Energy*, vol. 227, p. 120499, Jun. 2024.

[25] X. Peng, Y. Li, and F. Tsung, "A graph attention network with spatio-temporal wind propagation graph for wind power ramp events prediction," *Renewable Energy*, vol. 236, p. 121280, Dec. 2024.

[26] H. Yin, M. Yang, H. Lan *et al.*, "A hybrid graph attention network based method for interval prediction of shipboard solar irradiation," *Energy*, vol. 298, p. 131131, Jul. 2024.

[27] C. Yu, G. Yu, C. Yu *et al.*, "A multi-factor driven spatiotemporal wind power prediction model based on ensemble deep graph attention reinforcement learning networks," *Energy*, vol. 263, p. 126034, Jan. 2023.

[28] M. Perera, J. De Hoog, K. Bandara *et al.*, "Day-ahead regional solar power forecasting with hierarchical temporal convolutional neural networks using historical power generation and weather data," *Applied Energy*, vol. 361, p. 122971, May 2024.

[29] J. Yan, C. Möhrlen, T. Göçmen *et al.*, "Uncovering wind power forecasting uncertainty sources and their propagation through the whole modelling chain," *Renewable and Sustainable Energy Reviews*, vol. 165, p. 112519, Sept. 2022.

[30] J. Liu, H. Zang, L. Cheng *et al.*, "A transformer-based multimodal-learning framework using sky images for ultra-short-term solar irradiance forecasting," *Applied Energy*, vol. 342, p. 121160, Jul. 2023.

[31] P. Veličković, G. Cucurull, A. Casanova *et al.* (2018, Feb.). Graph attention networks. [Online]. Available: https://arxiv.org/abs/1710.10903

[32] A. Vaswani, N. Shazeer, N. Parmar *et al.* (2017, Jun.). Attention is all you need. [Online]. Available: https://arxiv.org/abs/1706.03762

[33] C. Lea, M. D. Flynn, R. Vidal *et al.*, "Temporal convolutional networks for action segmentation and detection," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, Jul. 2017, pp. 156-165.

[34] Korea Electric Power Corporation (KEPCO). (2025, Jan.). Jeju power supply and demand status. [Online]. Available: https://www.data.go.kr/data/15125113/fileData.do

[35] Korea Meteorological Administration (KMA). (2025, Feb.). Automated synoptic observing system (ASOS) real-time data. [Online]. Available: https://data.kma.go.kr/data/grnd/selectAsosRltmList.do

Muhammad Ali Iqbal received the B.S. degree from Federal Urdu University of Arts, Science and Technology (FUUAST), Islamabad, Pakistan, in 2016, and the M.S. degree from COMSATS University Islamabad (CUI), Islamabad, Pakistan, in 2019. He is currently a Ph.D. candidate in computer engineering at Jeju National University, Jeju, Republic of Korea. His research interests include machine learning, digital image processing, and computer vision.

Junho Ahn received the B.S. degree in computer engineering from Hongik University, Seoul, Republic of Korea, in 2006, the M.S. degree from Yonsei University, Seoul, Republic of Korea, in 2008, and the Ph.D. degree in computer science from the University of Colorado Boulder, Boulder, USA, in 2013. He was a Senior Researcher at the National Security Research Institute, Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea, until 2017, and an Associate Professor at Korea National University of Transportation, Chungju, Republic of Korea, until 2023. He is now an Associate Professor at Kyonggi University, Suwon, Republic of Korea. His research interests include computer vision, deep learning, pattern analysis, artificial intelligence, and real-time application detection.

Soo Kyun Kim received the B. S. degree in computer engineering from Soonchunhyang University, Asan, Republic of Korea, in 1998, and the M.S. and Ph. D. degrees in computer engineering from Korea University, Seoul,

Republic of Korea, in 2002 and 2006, respectively. He worked at Samsung Electronics, Suwon, Republic of Korea, from 2006 to 2008. He was a Professor at PaiChai University, Daejeon, Republic of Korea, until 2020, and is now a Professor at Jeju National University, Jeju, Republic of Korea. His research interests include deep learning-based geometric modeling and computer vision.