IEEE POWER & ENERGY SOCIETY SECTION

IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 29 June 2025, accepted 16 July 2025, date of publication 21 July 2025, date of current version 7 August 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3591040

RESEARCH ARTICLE

Transfer Learning for Photovoltaic Power Forecasting Across Regions Using Large-Scale Datasets

SEONGHO BAK[1], (Member, IEEE), SOWON CHOI[1], (Graduate Student Member, IEEE), DONGUK YANG[2], DOYOON KIM[1], HEESEON RHO[1], (Graduate Student Member, IEEE), AND KYOOBIN LEE[1], (Member, IEEE)

[1]Department of AI Convergence, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, South Korea
[2]New Energy Business Department, Hydrogen Business Team, Korea Electric Power Corporation (KEPCO), Naju 58322, South Korea

Corresponding author: Kyoobin Lee (kyoobinlee@gist.ac.kr)

This work was supported by the Institute of Information Communications Technology Planning Evaluation (IITP) grant funded by Korean Government (MSIT) (Artificial Intelligence Innovation Hub) under Grant RS-2021-II212068. We appreciated the high-performance GPU computing support HPC-AI Open Infrastructure via GIST SCENT.

ABSTRACT Accurate photovoltaic (PV) power forecasting enables stable grid operation; however, acquiring sufficient and diverse data remains challenging. Although numerous deep learning models have been employed, most rely on proprietary datasets or data spanning only a few years, limiting the generalizability of forecasting applications. This study proposes a cross-continental transfer learning framework for PV power forecasting using a large-scale dataset with 4.5 million data points, including a source domain dataset with over 3.5 million data points. The prediction model, which is a transformer-based approach, compares the results of zero-shot learning, linear probing, fine-tuning, and training on the target dataset alone. The target datasets are located on different continents with different meteorological characteristics. Results demonstrate that the proposed transfer learning approach significantly improves prediction accuracy. It reduces the mean absolute percentage error by up to 38.8% compared with models trained solely on the target data. Further analysis reveals that freezing a few transformer blocks is beneficial when the target dataset is sufficiently large, whereas freezing most layers is effective for smaller datasets. This study analyzes the impact of source and target datasets on prediction performance, demonstrating that larger datasets in both domains enhance forecasting accuracy. These findings offer a robust approach for cross-continental knowledge transfer in renewable energy forecasting, particularly benefiting regions with limited historical data or newly installed systems. The source code is available on https://github.com/gist-ailab/transfer-learning-PV-forecasting

INDEX TERMS Photovoltaic power forecasting, large-scale dataset, transfer learning, deep learning.

I. INTRODUCTION

Increasing concerns about global warming have heightened international interest in environmental issues, leading to the widespread adoption of the Net Zero policy. This policy balances greenhouse gas emissions with removals, achieving "zero" net emissions, and consequently emphasizes renewable energy source utilization [1]. Photovoltaic (PV) energy has garnered attention owing to its continuous availability from natural resources and relatively flexible installation requirements compared with other renewable sources. However, completely replacing conventional power grid operations with renewable energy sources alone remains challenging [2]. PV energy output varies significantly due to weather and environmental factors, presenting major challenges for stable grid management and effective market participation [3]. Therefore, advanced forecasting techniques must be developed for PV power generation. In response to these challenges, the Korean Power Exchange, which manages electric power transactions, has introduced the

The associate editor coordinating the review of this manuscript and approving it for publication was Feng Wu.

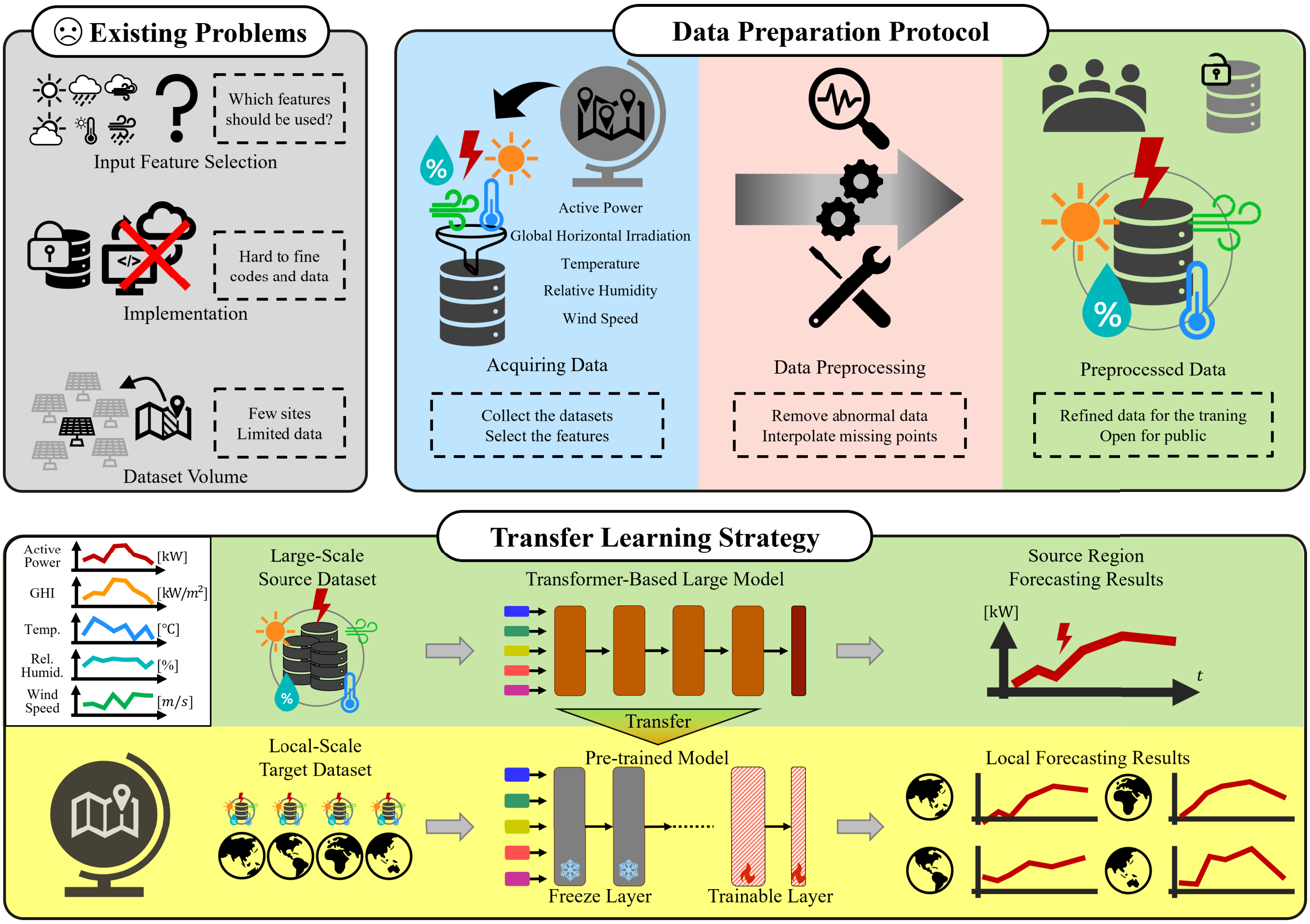


FIGURE 1. **Overview of the data preparation protocol and transfer learning process. To address key limitations in previous studies, this work introduces two main contributions: (1) a structured data collection and refinement procedure for training a transformer-based model, and (2) a site-specific transfer learning strategy to adapt the pre-trained model for local power forecasting.**

Renewable Energy Generation Forecasting Program. This program provides tiered incentives based on the accuracy of PV power forecasts made one day in advance [4].

Despite the growing emphasis on accurate PV forecasting, a significant gap remains. Most research relies on single-site data, limiting practical application for new sites with insufficient historical data [5], [6], [7]. Furthermore, the full potential of transformer-based models for PV forecasting remains underdeveloped owing to a lack of comprehensive studies utilizing PV datasets.

Transfer learning has emerged as a promising approach [8], [9], [10], enabling the use of pre-trained models from data-rich solar PV plants to improve forecasting at new, data-scarce sites. While successful in geographically similar regions, existing transfer learning approaches rarely span multiple continents or significantly different climate zones, posing a challenge for accurate forecasting across diverse global regions. For instance, Kim et al. [11] explored inter-continental transfer learning between Korea and Germany, whereas the SolNet model [12] utilized extensive datasets for global PV forecasting. This present study addresses these limitations by proposing a robust methodology for cross-continental knowledge transfer, further detailed in Section II.

This study proposes a data preparation protocol and a transfer learning strategy for PV power forecasting using a large-scale dataset. The approach demonstrates improved prediction accuracy across diverse geographical regions and climate zones. Fig. 1 shows the proposed data preparation protocol and transfer learning method for PV power generation forecasting. A large dataset is constructed from the Desert Knowledge Australia Solar Center (DKASC, Australia) by obtaining real-time power generation and weather information for PV arrays since 2008. Additionally, several local dataset are collected for conducting transfer learning. By systematically building and preprocessing the data for deep learning, the proposed transfer learning strategy improves forecasting PV power generation. Additionally, by conducting comprehensive experiments with varying data volumes, the impact of the data volume on performance and the application of transfer learning in local environments are analyzed. To the best of our knowledge, this is the first study

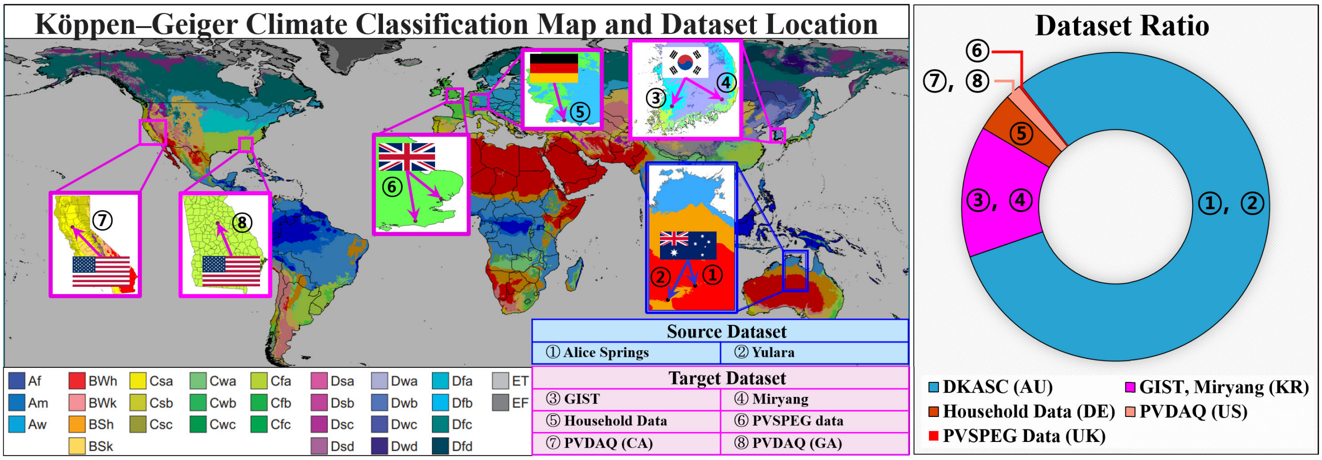


FIGURE 2. Geographic distribution of the datasets used in this study, overlaid on the Köppen–Geiger climate classification map. The accompanying graph shows the data point ratios for each dataset.

to utilize a large, million-scale dataset for transfer learning on PV power forecasting to across different continental regions. The main contributions of this study are as follows:

- A systematic preprocessing protocol is proposed to utilize large-scale PV power and meteorological data, enhancing their usability.
- The applicability across various regions and continents is verified using a large-scale source dataset (DKASC).
- The large-scale data are applied to the transformer-based model, resulting in improved prediction performance and demonstrating effectiveness in data-scarce regions.
- By comparing the performance of different transfer learning techniques (zero-shot, fine-tuning, linear probing) and data size/training period changes, appropriate conditions for utilizing data and models are proposed.

The remainder of this paper is organized as follows: Section II provides an overview of PV power generation forecasting. Section III introduces the datasets from various continents and outlines the preprocessed steps. Section IV describes PatchTST [13], a transformer encoder-based time series prediction model, along with the transfer learning approach used in this study. Section V presents the results of applying transfer learning using large-scale datasets and highlights the importance of dataset volume. Finally, Section VI concludes the study.

II. RELATED WORKS

A significant trend in recent research involves applying deep learning technologies to PV power forecasting, thereby contributing to the efficient operation and economic viability of volatile renewable energy sources [14]. Among renewable energy sources, solar and wind energy represent major contributors, with solar power exhibiting particularly rapid growth in recent years. According to the International Energy Agency, the global solar PV market recorded 231 gigawatts direct current of new installations in 2022 alone, reaching a cumulative installed capacity of 1.2 terawatts direct current. This trend represents the fastest growth among renewable energy sectors [15]. With the rapid growth of solar power generation, the need for accurate PV power forecasting techniques has also been increasing [16]. Inaccurate forecasting can lead tocause unexpected system performance deterioration or failures, resulting in higher maintenance costs and potential losses in power generation [17].

PV power forecasting methods have evolved significantly. Early approaches relied on traditional statistical models such as autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), and autoregressive integrated moving average (ARIMA) models, which also addressed nonstationary data [18], [19], [20], [21]. Concurrently, analytical PV power forecasting models emerged, utilizing irradiance and temperature data from numerical weather prediction models [22]. Since the 1980s, the PV power forecasting has transitioneed toward machine learning techniques, notably artificial neural networks (ANNs) and support vector machines (SVMs) [5], [6]. These methods improved forecasting by capturing complex, nonlinear relationships but often faced limitations such as overfitting, high computational requirements, and difficulties with high-dimensional input data. More recently, advances in computational resources and algorithms have facilitated the adoption of deep learning models, including long short-term memory (LSTM) [23] and convolutional neural networks (CNNs) [24], [25], [26], which excel at capturing temporal and spatial characteristics, respectively. The latest development involves transformer-based models, such as PatchTST [13] and Autoformer [27], which utilize robust self-attention mechanisms to effectively manage large-scale datasets and long-term dependencies, leading to substantial improvements in prediction accuracy.

Subsequently, advancements in computational resources and algorithmic improvements have enabled the construction of deeper neural network architectures, leading to the emergence of deep learning. Deep learning methods can precisely capture complex patterns and nonlinear correlations within large-scale datasets—features that traditional ANN and statistical approaches struggled to detect—significantly improving the accuracy of PV power forecasting [28], [29]. Al-Waeli et al. [30] compared the effectiveness of mathematical models and neural networks (MLP and SOFM) for forecasting the power output of a nano-cooled PV/T system. A systematic review by Khouili et al. [28] reported that LSTM networks, effective for capturing temporal characteristics, and CNNs, well-suited for spatial features, are among the most widely used deep learning algorithms in PV forecasting. To address the challenge of high-dimensional data that these models can face, other studies have proposed combining dimensionality reduction techniques like principal component analysis (PCA) with ANN models to increase computational efficiency and analyze the impact on prediction accuracy [31]. Notably, deep learning models trained on insufficient data typically suffer from underfitting, failing to adequately capture complex data patterns, which can lead to high variance in performance estimates. To address this limitation, numerous studies have explored the application of transfer learning, utilizing deep learning models pre-trained on data-rich solar power plants, to achieve reliable PV power forecasting performance for newly installed plants lacking sufficient historical data [8], [10], [11]. For instance, Sarmas et al. [8] demonstrated transfer learning's effectiveness by applying models trained on PV plants in Western European to newly installed plants in Greece, effectively addressing data scarcity within geographically similar regions and enhancing prediction accuracy. Similarly, Miraftabzadeh et al. [10] utilized pre-trained deep learning models from a data-rich PV plant in Maryland, USA, to enhance forecasting performance at a nearby new plant with limited data. Despotovic et al. [32] also performed transfer learning among multiple meteorological stations in Spain to predict solar irradiance. Collectively, these studies demonstrate that transfer learning is a robust approach to mitigating data scarcity, effective across varying degrees of geographical similarity, and significantly improves the generalization capabilities of PV forecasting models between regions with different environmental characteristics.

However, studies applying transfer learning across different continents remain limited. Kim et al. [11] extended the scope of transfer learning applications by conducting intercontinental transfer learning between PV power plants in South Korea (Asia) and Germany (Europe). Their study explored both transfer directions—applying models pre-trained on Korean PV data to German plants and vice versa—demonstrating improved generalized forecasting performance despite significant geographic and climatic differences between these regions. Similarly, the SolNet model [12] utilized extensive datasets from the Netherlands, Australia, and Belgium to develop a globally applicable PV power forecasting model.

Despite these advancements, the full potential of cross-continental transfer learning for large-scale PV forecasting remains largely unexplored. This study addresses this research gap by proposing a robust methodology and providing practical insights to enhance prediction accuracy across diverse global regions.

TABLE 1. PV Systems of source dataset located in alice springs and yulara in australia.

Parameter	Alice Springs	Yulara
Total array rating [kW]	189.5	1,820
Largest array capacity [kW]	7.0	1,058
Smallest array capacity [kW]	2.0	22.6
Mean array rating [kW]	5.742	260.757
Number of PV arrays	35	7
PV technology	mono-Si, poly-Si, HIT Hybrid Silicon, VdTe, CIGS	mono-Si, poly-Si
Installed year	2008	2016
Array tilt [deg.]	20	10, 15, 20
Array Azimuth [deg.]	0 (Solar North)	0 (Solar North), 40 (E of N)
Coordinate (Lat., Long.)	-23.7618 133.8749	-25.1905 130.9764

TABLE 2. Description of meteorological features and target variable.

Feature Name	Unit	Description
Global Horizontal Irradiation (GHI)	W/m^2	The total solar radiation received on a horizontal surface; the primary driver of PV power generation.
Air Temperature (Temp.)	°C	The ambient air temperature, which directly affects the operational efficiency of PV panels.
Relative Humidity (RH)	%	The amount of water vapor in the air, which can influence atmospheric transparency and panel performance.
Wind Speed	m/s	The speed of the wind, which contributes to the convective cooling of PV panels.
Active Power (AP)	kW	The actual electrical power output of the PV system; the target variable for forecasting.

III. DATA

A large source dataset, detailed in Table 1, was obtained to facilitate accurate PV generation forecasting. Target datasets from diverse climates and regions were also obtained to evaluate the generalizability of the pre-trained model across PV power forecasting tasks. This study employs the largest publicly available dataset and proposes preprocessing procedures to train the model for PV power generation prediction. As shown in Fig. 2, the obtained datasets span multiple climatic zones across Asia, Europe, Oceania, and North America [33], [34]. Following data acquisition, preprocessing techniques were implemented to remove outliers and irrelevant data, ensuring data quality for effective model training. Subsequently, a correlation analysis was conducted between the target variable, active power (AP), and relevant meteorological factors.

A. DESCRIPTION OF PV DATASETS

This study used data from DKASC to pre-train the proposed PV power prediction model. The DKASC, located in the Northern Territory inland of Australia, provided the large-scale source dataset [35]. As an open-access platform, DKASC offers multiple datasets; two were selected for this study: DKASC, Alice Springs, and Yulara Solar Systems. Notably, DKASC Alice Springs has been collecting data since 2008 and features multiple technologies and various installation configurations. Yulara Solar Systems, located around Uluru with a total capacity of 1.8 MW, has been operational since 2016. Although both sites feature diverse installation methods, this study focused on fixed systems, which represent the most common installation type.

The local-scale target datasets were collected from various locations: South Korea (Gwangju Institute of Science and Technology [GIST], Miryang), Germany (Household Dataset [36]), the United States of America (photovoltaic data acquisition [PVDAQ] [37]), and United Kingdom (photovoltaic solar panel energy generation [PVSPEG] dataset [38]). The specifications of both source and target datasets, including capacity, measurement periods, and the number of installed PV arrays, are described in Table 3. The GIST dataset includes PV power generation data from 14 institutional buildings and meteorological data sourced from the Open MET Data Portal [39], operated by the Korea Meteorological Administration (KMA). One of KMA's automatic weather stations is located on the GIST campus. The Miryang dataset comprises power generation data from seven PV arrays within Miryang city, with meteorological data obtained from the National Institute of Agricultural Sciences agricultural weather observation database [40]. The household data consists of power generation and load data from several small industrial warehouses and residential buildings in the Konstanz region of Germany, with weather data sourced from the Deutscher Wetterdienst Open Data Server [41]. The PVDAQ dataset, developed by the National Renewable Energy Laboratory, is a time-series database covering various experimental and commercial public PV sites. Based on this database, this study specifically utilized PV arrays with fixed installation types located in Social Circle, Georgia, and Arbuckle, California—sites included in the 2023 DOE Solar Data Prize dataset. These data include both generation and meteorological variables; however, these PVDAQ datasets do not contain RH information, and no meteorological observatories exist close to either site. In this case, no substitution data exist for the RH variable. The PVSPEG dataset includes load and generation information from various regions in England. For this study's analysis, power generation and weather data from PV arrays located in Maple Drive East and Forest Road areas were used.

An initial correlation analysis using the raw data from the previously described datasets indicated that GHI, wind speed, air temperature, and RH significantly influence AP generation. Table 2 presents the description of those meteorological features. However, owing to the presence of missing values and outliers, data preprocessing was subsequently conducted.

B. DATA PREPROCESSING

Inspection of the raw datasets revealed numerous missing values and outliers. For instance, attribute statistics from the Alice Springs dataset [42] indicated unrealistic meteorological values, such as wind speeds below −900 m/s and RH exceeding 100%. Additionally, measurement failures were observed across all the datasets. To address these challenges, a standardized preprocessing protocol was established, as shown in Algorithm 1, and applied uniformly across all datasets. Before preprocessing, data were standardized by the hour and included variables such as GHI, air temperature, wind speed, RH, and AP as consistent units. Specifically, the preprocessing procedure consists of two main stages. First, outliers and inconsistent data are identified and removed based on physical and logical constraints. This procedure includes identifying unrealistic values and inconsistent relationships between variables, such as zero power generation during high irradiance or significant power generation during negligible irradiance. Second, data are processed at the daily level, where days with excessive missing values (two or more consecutive missing values

TABLE 3. Summary of characteristics, locations, and specifications of the target PV datasets.

Site Name	Total Array Rating [kW]	PV Technology	Measurement Period	No. of PV Arrays	Array Tilt [deg.]	Array Azimuth [deg.]	Coordinate (Lat., Long.)
GIST	1,180.6	not specified	2018. 10 - 2024. 09.	14	not specified	not specified	35.2274 126.8416
Germany	69.15	not specified	2015. 05. - 2018. 09.	9	not specified	not specified	not specified
Miryang	2,028	mono-Si	2022. 10. - 2024. 09.	7	25°	not specified	35.5282 128.8942
California	893	mono-Si	2018. 12. - 2023. 11.	1	25°	180°	38.9963 -122.1341
Georgia	38,687	poly-Si	2016. 02. - 2023. 11.	1	20°	180°	33.6780 -83.6759
UK	6.371	not specified	2013. 10. - 2014. 10.	2	not specified	not specified	not specified

Algorithm 1 Data Pre-processing Procedure

Require: A comma-separated value dataset with columns: timestamp, AP, GHI, air temperature, wind speed, and RH.
Ensure: A pre-processed dataset with cleaned rows and interpolated missing values.

```
for each row (timestamp) in the dataset do
    if GHI > 2000 or
       Air_Temperature < −30 or
       Wind_Speed < 0 or
       Relative_Humidity ∉ [0, 100] or
       Normalized_AP = 0 and GHI > 200 or
       Normalized_AP > 0.1 and GHI < 10 or
       ConsecutiveSame_AP ≥ 10
    then
        Set the corresponding feature value to NaN
          ▷ Outliers and inconsistent data
    end if
end for
for each day in the dataset do
    if Consecutive NaN values ≥ 2 then
        Remove the day
          ▷ Remove days with excessive missing data
    else if Single NaN value exists then
        Apply linear interpolation
          ▷ Fill small gaps in data
    end if
end for
```

[NaNs]) are removed, while single missing values are handled through linear interpolation to maintain data continuity.

After pre-processing, the datasets were verified and found to be within the normal range. Subsequently, Pearson correlation was conducted between AP and other attributes across all datasets. Pearson correlation is a statistical measure that evaluates the strength and direction of the linear relationship between two continuous variables, ranging from -1 to 1. As presented in Table 4, GHI exhibited the strongest positive correlation with AP across all sites, whereas RH showed a negative correlation. Air temperature and wind speed demonstrated significant correlations that varied by location. However, RH was excluded from this analysis and subsequent experiments for the sites in Georgia and California, as this variable was not recorded at those locations. Consequently, more than 3.5 million data points were retained for the large-scale source dataset, and approximately 963,000 data points were retained across the local-scale target datasets.

IV. METHODOLOGY

A. BASELINE MODEL

PatchTST is a transformer encoder-based model designed for time series forecasting, incorporating two key design elements: patching and channel independence. These design elements provide significant advantages regarding prediction time, spatial complexity, extended look-back window

TABLE 4. Correlation of meteorological features with AP (kW) for each site. Asterisk(*) indicates the source dataset.

Site Name	GHI [W/m^2]	Air Temp. [°C]	RH [%]	Wind Speed [m/s]
Alice Springs*	0.91	0.29	-0.38	0.30
Yulara*	0.91	0.29	-0.38	0.33
GIST	0.88	0.19	-0.54	0.28
Germany	0.92	0.61	-0.68	0.01
Miryang	0.75	0.13	-0.48	0.16
California	0.98	0.54	-	0.27
Georgia	0.95	0.38	-	0.41
UK	0.77	0.40	-0.52	0.04

capabilities, and representation learning capacity. This section presents the distinction between PatchTST and conventional transformer encoders.

As illustrated in Fig. 3(a), the input signal $\boldsymbol{x} \in \mathbb{R}^{B\times M\times L}$ undergoes instance normalization [43] to enable channel independence, followed by channel-wise patching of each univariate time series. Notably, B represents the batch size, M denotes the number of input channels, and L is the input sequence length. Through channel independence, the input is divided into individual univariate time series $\boldsymbol{x}^{(i)}$, where $i = 1, \ldots, B \times M$. The patching process transforms each input into a sequence of patches. With patch length P and stride length S, the input is converted into $\boldsymbol{x}_p^{(i)} \in \mathbb{R}^{P\times N}$, where N represents the number of patches and is calculated as $N = \lfloor\frac{(L-P)}{S}\rfloor+2$. Subsequently, the transformed input $\boldsymbol{x}_p^{(i)}$ is mapped to the Transformer encoder's latent space of dimension D through two learnable components: a linear projection $\mathbf{W}_p \in \mathbb{R}^{P\times N}$ and positional embedding $\mathbf{W}_{pos} \in \mathbb{R}^{D\times N}$. This transformation is expressed as $\boldsymbol{x}_d^{(i)} = \mathbf{W}_p\boldsymbol{x}_p^{(i)} + \mathbf{W}_{pos}$, where $\boldsymbol{x}_d^{(i)} \in \mathbb{R}^{D\times N}$ denotes the input for the transformer encoder shown in Fig. 3(b).

Finally, the latent features are processed through the transformer encoder, followed by a flatten layer and feature projection to generate predictions $\hat{\boldsymbol{y}}^{(i)} = (\hat{y}_{O+1}^{(i)}, \ldots, \hat{y}_{O+T}^{(i)}) \in \mathbb{R}^{1\times T}$ for T time steps ahead from the current time O. For the PV power forecasting task, the predicted AP, denoted as $\hat{\boldsymbol{y}}^{(k)} \in \mathbb{R}^{1\times T}$, is the primary output. The model is trained using the mean square error (MSE) loss function, optimized with the AdamW algorithm.

This study utilized PatchTST, which is a transformer-based model for PV power forecasting using an NVIDIA RTX 3090. The details for reproducibility are presented in Table 5. The GeLU at the activation function stands for Gaussian error Linear Unit [44]. The dropout was applied to the transformer and fully connected layers, not to attention heads. These details were selected through extensive empirical validation to ensure optimal forecasting performance across various PV datasets. The same hyperparameter configuration was maintained for all subsequent transfer learning experiments to ensure a fair comparison.

TABLE 5. The details include hyperparameters and configurations. Asterisks indicate the parameter the PatchTST.

Hyperparameters or configurations	Values
Batch size	256
Transformer input dimension	5
Transformer hidden dimension	512
Fully connected layer dimension	2048
Number of encoder blocks	8
Number of attention heads	16
Input sequence length	240
Prediction sequence length	24
Patch length*	24
Stride for patches*	8
Dropout rate	0.05
Activation function	GeLU
Early stopping	10 epochs

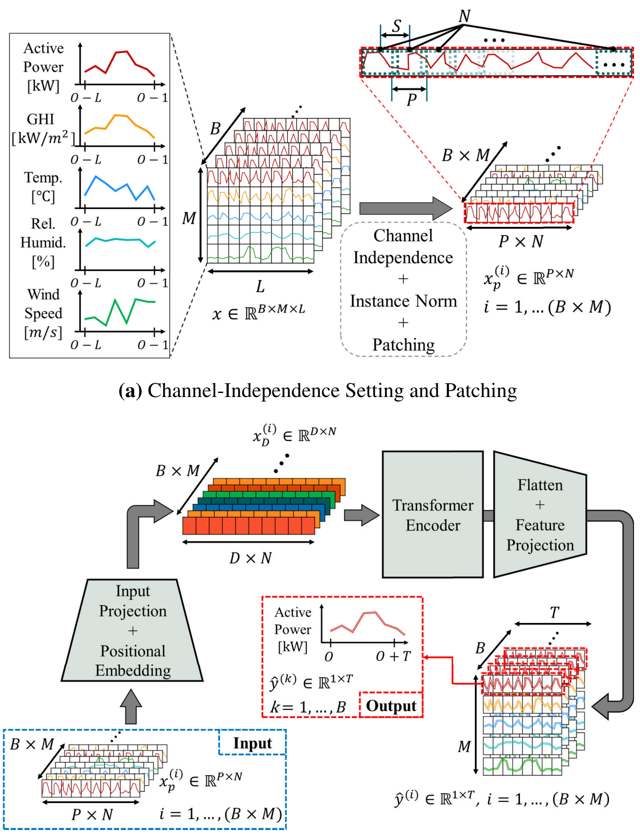


FIGURE 3. Explanation of the channel independence approach, patching process, and encoder-based transformer model.

B. TRANSFER LEARNING

The key to transfer learning in PV power forecasting is applying knowledge learned from large datasets to regions with limited data availability. To achieve this objective, a pre-trained model developed on the large-scale DKASC dataset is applied to smaller target datasets. The pre-trained PatchTST model comprises an input embedding layer, eight transformer encoder blocks, and a linear head for the final prediction. Transfer learning is performed by selectively fine-tuning or freezing layers of the trained model to utilize the pre-trained features effectively. The earlier layers learn generalized features, such as basic time series patterns, whereas deeper layers learn domain-specific features. Thus, selective freezing of layers enables systematic differentiation between general-purpose features applicable to new environments and those requiring adaptation to specific local datasets. This study adjusted the number of frozen transformer blocks and analyzed the change in prediction performance for each setting.

C. EVALUATION METHOD

For performance evaluation, multiple metrics, including mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R^2), and mean absolute percentage error (MAPE), were employed based on Zhang et al.'s study [45]. When calculating MAPE, the equations presented in [4] and [45] divide the absolute difference between the true and predicted values by the capacity to derive the percentage error. However, in this study, dividing by the installed capacity was infeasible because many datasets did not specify the installed capacity of the PV arrays. Therefore, maximum power generation was used instead of the installed capacity for consistency in evaluation, as shown in (1).

$$\text{MAPE} = \frac{1}{N}\sum_{k=1}^{M}\sum_{i=1}^{N_k}\left|\frac{y_i^k - \hat{y}_i^k}{y_{\max}^k}\right|, \quad N = \sum_{k=1}^{M} N_k, \tag{1}$$

where y_i^k is the true value, $\hat{y}_i^k$ is the corresponding predicted value for group k, M is the number of groups with different maximum values, N_k is the number of data points in group k, $y_{\max}^k$ represents the maximum true value in group k, and N denotes the total number of data points across all groups.

Unlike previous studies, the proposed dataset includes diverse PV installation scales, prompting an evaluation of model performance across different capacity ranges. Evaluation groups were primarily categorized in 100 kW intervals. For installations below 100 kW, classification followed the guidelines set by the Korea Ministry of Trade, Industry and Energy [46]. Based on these guidelines, installations under 30 kW were classified as small, while those between 30 and 100 kW were classified as small-to-medium. Accordingly, installations under 100 kW were further divided into two groups.

V. EXPERIMENTS AND RESULTS

This study compares a control group and an experimental group to evaluate the effectiveness of transfer learning. The control group includes experiments conducted on each dataset, whereas the experimental group applies transfer learning on the target domain dataset. The transfer learning experiments are divided into two categories: first, evaluating the transfer learning performance on the test set of the target

TABLE 6. Performance comparison on source dataset (DKASC).

Model Name	Capacity Range [kW]	MAE [kW]	RMSE [kW]	R^2	MAPE [%]
PatchTST [13]	< 30	0.029	0.086	0.996	**0.909**
	30–99	2.067	5.227	0.970	
	300–399	3.996	10.478	0.987	
DLinear [47]	< 30	0.177	0.683	0.928	3.146
	30–99	3.802	7.775	0.933	
	300–399	11.667	24.096	0.931	
LSTM [23]	< 30	0.055	0.123	0.993	1.275
	30–99	1.740	3.888	0.983	
	300–399	3.625	9.018	0.990	
Transformer [48]	< 30	0.045	0.091	0.996	1.007
	30–99	1.493	3.423	0.987	
	300–399	2.696	6.962	0.994	
Autoformer [27]	< 30	0.196	0.326	0.947	3.707
	30–99	4.956	7.956	0.930	
	300–399	13.938	23.959	0.932	

domain dataset; and second, comparing results across varying quantities of source and target domain datasets used during training.

A. PERFORMANCE EVALUATION ON SOURCE DATASET

As the initial step of the experiment, the performance of several deep learning models, such as PatchTST, DLinear [47], LSTM, Transformer, and Autoformer, was evaluated using the source dataset (DKASC Alice Springs and Yulara). The test set was divided into three groups according to the range of the power generation capacities: <30 kW, 30–99 kW, and 300–399 kW, with 252,000 (252k), 56,000 (56k), and 51,000 (51k) data points, respectively. As shown in Table 6, the PatchTST model achieved the best overall performance, recording the lowest MAPE of 0.909% across all capacity ranges. It also achieved the highest prediction accuracy in the <30 kW group, which represented approximately 70% of the total dataset. Given PatchTST's widespread adoption as a benchmark model in the latest time series forecasting research ([27], [49], [50], [51]) and these performance evaluation results, it was selected as the base model for the subsequent transfer learning experiments. During the model training process, PatchTST was implemented based on PyTorch and trained using the AdamW optimizer, with a learning rate of 10^{-4} and batch size of 256. Additionally, the input characteristics were normalized using the StandardScaler to reflect the characteristics of each PV array. These settings were maintained for the subsequent transfer learning experiments.

B. TRANSFER LEARNING

Prior training on large datasets enhances transfer learning performance on small datasets. Two experimental setups were designed to verify the effectiveness and feasibility of transfer learning for PV power forecasting.

First, this study evaluated the effectiveness of transfer learning by applying a pre-trained model from the source dataset to multiple target datasets. During this process, optimal layer freezing strategies were analyzed based on the characteristics and size of the dataset, and a systematic transfer learning methodology was developed according to the data conditions. Second, the impact of source and target dataset sizes on transfer learning performance was analyzed. This approach involved experiments that gradually reduced the source data's size and varied the target data's training period.

Through these experiments, this study proposes data conditions and learning strategies to effectively transfer PV power generation characteristics learned from large datasets to environments with limited data volume, as detailed in the following sections.

1) TRANSFER OF SOURCE-TRAINED MODEL TO TARGET DATA

Table 7 presents the results of applying pre-trained models to various target datasets to verify the effectiveness of transfer learning, a key focus of this study. Performance comparisons are made between the target-only (TO) approach, which is trained using only the target dataset, and the source-to-target (S2T) approach, which transfers knowledge based on the pre-trained model trained on the source dataset. The results demonstrate that transfer learning improves MAPE across all datasets, with additional improvements observed in other performance metrics, except for R^2 in the cases of California and Georgia. These findings indicate that models pre-trained on the source dataset effectively enhance prediction performance across target datasets, regardless of geographical location, generation capacity, or data volume of PV installations.

However, the California and Georgia datasets showed only modest improvements compared with other datasets. As shown in Table 4, this variation is attributed to their limited data volume and the absence of RH from the input channels, a variable that exhibited a strong negative correlation. Notably, despite a significantly smaller dataset, the United Kingdom dataset achieved a much greater performance improvement compared with California and Georgia. Furthermore, the GIST dataset, the largest target dataset, did not achieve the best prediction performance. These findings suggest that, regarding transfer learning, the number of available input features often has a stronger influence on performance than dataset size.

Before presenting the transfer learning results for the target dataset in Table 7, experiments were conducted to optimize the freezing level of the pre-trained model's transformer block to achieve improved performance across different target datasets. For this purpose, different freeze strategies were applied, ranging from fully fine-tuning (no freezing) to linear probing (freezing all eight blocks). Prediction performance was analyzed by gradually increasing the

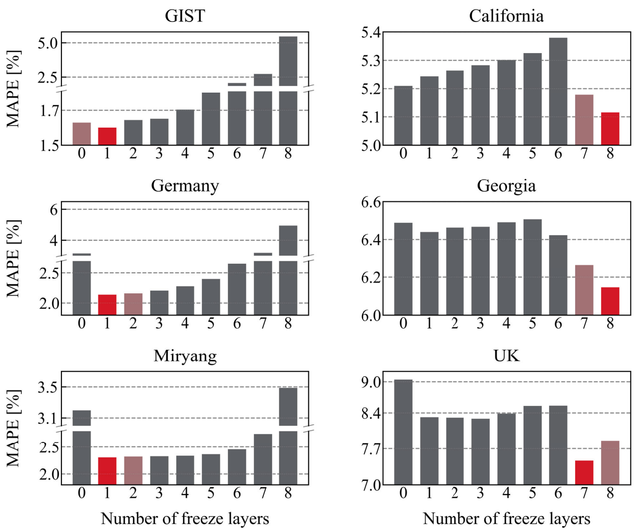


FIGURE 4. Performance analysis of transfer learning across different target datasets with varying numbers of frozen transformer blocks. A value of zero indicates fully fine-tuning, while eight corresponds to linear probing on a number of freeze blocks.

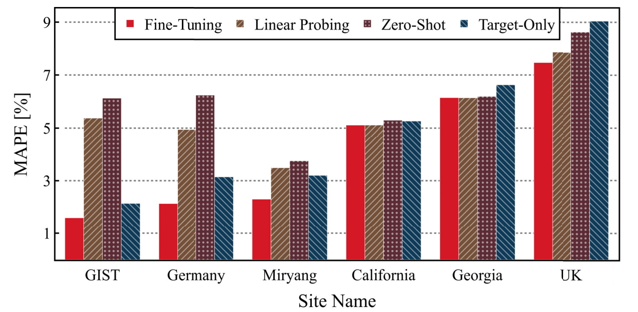


FIGURE 5. Comparison of different transfer learning strategies and TO training across target datasets. Fine-tuning is transfer learning by freezing several blocks. Linear probing is just freezing the head layer. Zero-shot means directly using a pre-trained model without any training on the target dataset. Target-only means training with only the target dataset from scratch.

freezing level starting from the blocks closest to the input embedding layer.

According to the results shown in Fig. 4, the optimal freezing strategy can be categorized into two patterns based on dataset characteristics. For relatively large datasets such as GIST (531,000 data points), Germany (145,000), and Miryang (82,000), the best performance was achieved by freezing only the first transformer block. Notably, prediction error increased as the number of frozen blocks increased. However, for smaller datasets such as California (45,000), Georgia (26,000), and the UK (10,000), freezing 7–8 blocks was more effective. Two main factors may explain this difference. The first is the size of the dataset. In the Miryang dataset (approximately 82,000 data points), freezing a few blocks was optimal, while freezing most blocks improved performance in the California dataset (45,000 data points). The second is the difference in data partitioning. Datasets larger than 80,000 have sufficient installations to allow a per-installation train/validation/test split, while smaller datasets require a chronological split. These results suggest that freezing fewer transformer blocks is generally effective when per-installation data exceed 80,000 data points, while freezing most blocks may better support transfer learning performance when data volume is limited or a chronological split is required.

Fig. 5 shows the results of fine-tuning, linear probing, zero-shot, and TO. In this context, zero-shot means that the pre-trained model predicts directly without additional training on the target dataset. Fine-tuning showed the best performance across all datasets. However, for the GIST, Germany, and Miryang datasets–which have relatively large sample sizes-it ranked the second best. Conversely, for smaller datasets, zero-shot prediction with a model pre-trained on a large source dataset outperformed training solely on the local dataset.

The mechanism underlying these freezing strategies can be interpreted from both computational and representational perspectives. Computationally, increasing the number of frozen blocks significantly accelerates convergence during fine-tuning, as fewer parameters require optimization. This faster convergence is particularly beneficial for smaller

TABLE 7. Performance comparison on target datasets for fine-tuning and target-only (TO) training. Evaluation metrics show TO versus source-to-target (S2T) transfer learning results. The MAPE shows relative improvement achieved through transfer learning, with bold values indicating better performance between TO and S2T approaches.

Site Name	# of Data Points	Capacity Range [kW]	MAE [kW]		RMSE [kW]		R^2		MAPE [%]		ΔMAPE [%]
			TO	S2T (Ours)	TO	S2T	TO	S2T	TO ^{A}	S2T ^{B}	$= \frac{(A-B)}{B} \times 100$
GIST	531,710	< 30	1.133	**0.794**	2.398	**2.162**	0.975	**0.980**	2.143	**1.601**	**33.90**
		100–199	3.882	**3.050**	7.844	**7.280**	0.903	**0.916**			
Germany	145,209	< 30	0.3489	**0.2373**	0.6948	**0.5885**	**0.927**	0.947	3.143	**2.1378**	**47.02**
Miryang	82,255	500–599	17.071	**12.296**	35.355	**33.014**	0.856	**0.874**	3.199	**2.304**	**38.83**
California	45,699	700–799	37.866	**36.794**	77.376	**79.671**	**0.896**	0.890	5.264	**5.155**	**2.91**
Georgia	26,547	32900–32999	3839.6	**2027.5**	3834.7	**3839.6**	0.778	0.778	6.628	**6.146**	**7.83**
UK	10,083	< 30	0.241	**0.199**	0.363	**0.357**	0.562	**0.576**	9.042	**7.466**	**21.10**

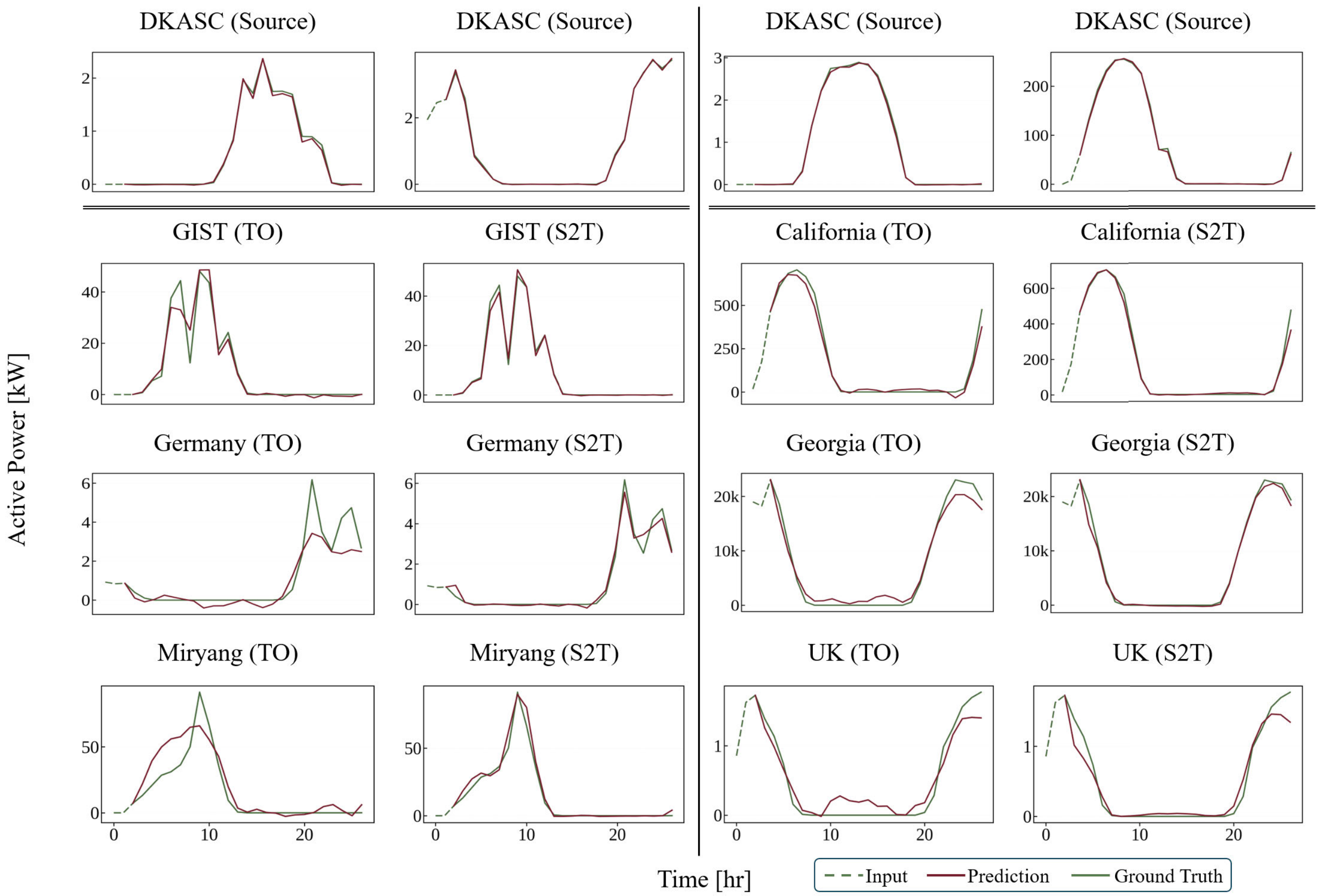


FIGURE 6. Visualization of PV power forecasting results across different datasets. Prediction results on source dataset (DKASC) and target datasets using TO and S2T transfer learning approaches. Notably, S2T predicts closer to the ground truth than TO.

datasets, where extensive training can lead to overfitting. Representationally, the strategy reflects the hierarchical nature of learned features in transformer architectures. Early blocks in the pre-trained model capture fundamental temporal patterns, such as diurnal cycles and weather-power relationships, universally applicable across geographical regions. Conversely, deeper blocks encode more source-domain-specific features, such as local meteorological patterns unique to the Australian climate. For large datasets, minimal freezing supports the adaptation of these domain-specific representations while preserving universal patterns. However, for smaller datasets, extensive freezing prevents overfitting by limiting adaptation to only the most generalizable layers. This phenomenon explains why the UK dataset (10,000 data points) benefits from freezing most blocks, while the GIST dataset (531,000 data points) achieves optimal performance with minimal freezing.

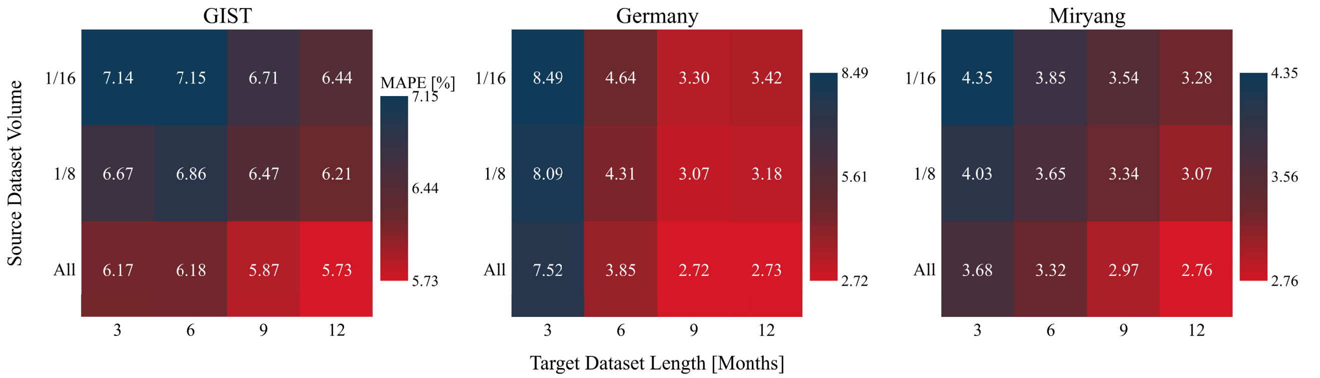
Fig. 6 shows a sample of the prediction results from TO training and from applying the pre-trained model to target datasets. The pre-trained model performs the prediction task well on both the source and target datasets, often outperforming models trained with TO.

2) IMPACT OF VARYING SOURCE AND TARGET DATASET VOLUME

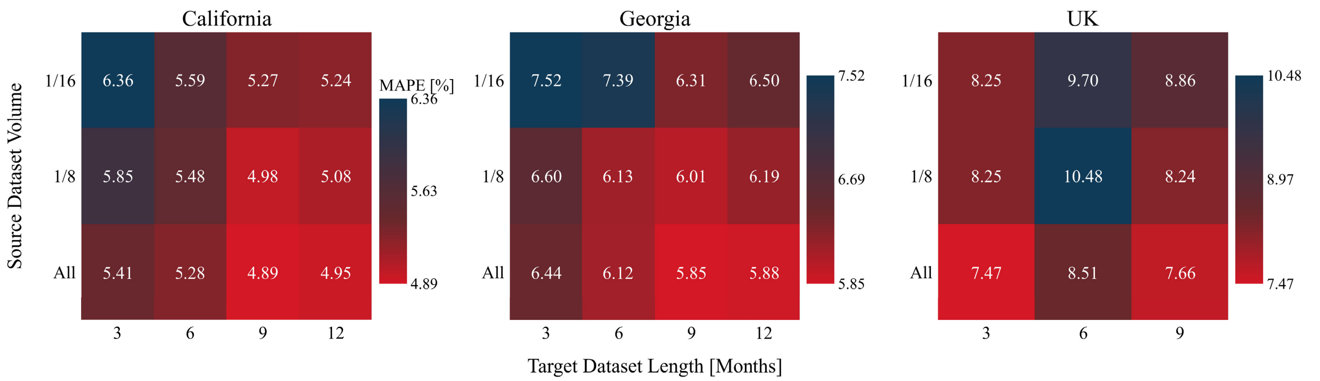
This study examines how varying the amounts of data in the source and target datasets affects the performance of transfer learning. Table 8 presents the results obtained using pre-trained models trained on 1/8 and 1/16 of the source dataset. Comparing with results from models trained on the full source dataset (Table 6), the error increases progressively as the volume of source dataset decreases. For example, when the source dataset was reduced by 1/8, the MAPE increased to 2.701%; when it was reduced by 1/16, it increased to 3.345%.

To analyze the effect of the target dataset, the training period was adjusted from 3 months to 12 months in 3-month increment. The test set remained fixed, based on its original size defined before the data volume adjustment. Training sets were constructed based on the most recent data. For datasets containing multiple installations, training period was adjusted independently for each installation. For the UK dataset, the duration ranged from three to nine months, as the data for individual installations consisted of less than a year of data.

The prediction error decreased as the volume of source and target datasets increased. As shown in Fig.7, which presents



(a) Target datasets that have relatively sufficient data volume



(b) Target datasets that have relatively insufficient data volume

FIGURE 7. Comparison of source and target data volumes for transfer learning performance, measured by MAPE. Heatmaps show the impact of source dataset size (All, 1/8, 1/16) and target dataset duration (3–12 months) on performance for (a) target datasets with sufficient data volume and (b) target datasets with insufficient data volume. Lower MAPE values, indicated in red, reflect better performance.

TABLE 8. Performance comparison for source dataset (DKASC) with reduced source volume. Reducing the original dataset to 1/8 (447,000 points) and 1/16 (223,000 points) impacts on the forecasting accuracy.

Source Reduction	Capacity Range [kW]	MAE [kW]	RMSE [kW]	R^2	MAPE [%]
1 / 8 (447,000 points)	< 30	0.139	0.334	0.945	2.701
	30–99	3.678	7.940	0.931	
	300–399	11.595	24.961	0.926	
1 / 16 (223,000 points)	< 30	0.183	0.398	0.922	3.345
	30–99	4.249	8.333	0.923	
	300–399	13.184	26.102	0.912	

a heatmap of MAPE variation with dataset volume for all target datasets, overall performance improved as the amount of source and target datasets increased. However, for the UK dataset, the lowest prediction error was recorded when the train set size was reduced to three months, indicating a different relationship compared with other datasets. This outcome likely occurred because the reduced training set contained data that reflected the seasonal characteristics most similar to those in the test set. The data from the last three months most accurately reflected the characteristics of the test set, while the data from the previous three months showed different characteristics, likely due to seasonal fluctuations. Fine-tuning the model, pre-trained on the full source dataset, using only three months of UK data yielded prediction accuracy comparable to that reported in Table 7.

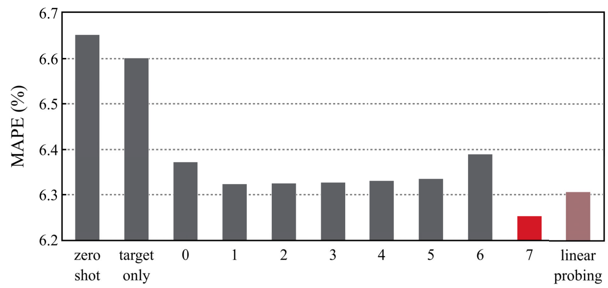
The comprehensive experimental results demonstrate that the source and target dataset sizes significantly impact the performance of transfer learning. Particularly, it emphasizes that pre-training with a large source dataset is crucial for improving the performance of transfer learning. Furthermore, analysis of the UK dataset highlights that successful transfer learning depends on data volume and temporal alignment between training and testing periods, particularly regarding seasonal characteristics.

3) ASSESSING TEMPORAL ROBUSTNESS

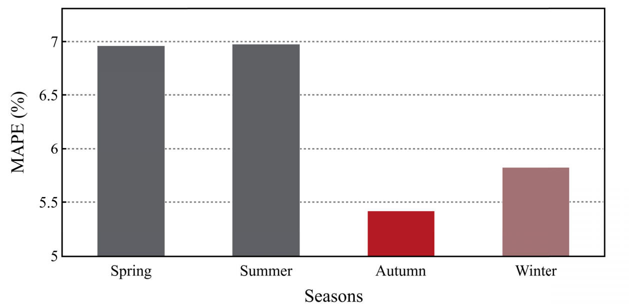
In real-world applications, a solar power prediction model can be used for several months to a year beyond its training data period without further retraining. In the results presented in section V-B1, both California and Georgia datasets, trained on single PV arrays, were tested against nonoverlapping periods of 10 and 7 months, respectively. However, neither datasets included RH as a meteorological feature, nor did their climates exhibit high variability to fully assess the model's robustness to temporal and seasonal climate changes.

Therefore, this study examines the robustness of transfer learning against meteorological and seasonal changes using the GIST dataset. The GIST dataset was selected for evaluating temporal robustness because, among all source datasets, it demonstrates the most significant annual climate variability across seasons. The most recent year of data for all PV arrays was designated as the test set to assess temporal robustness. For the remaining periods, the PV arrays used as training and validation sets in the experiment described in section V-B1 were utilized as the training set, while the other arrays were assigned to the validation set. Considering the data from the dataset was collected until September 24, 2024, and to align with seasonal divisions of three months, data acquired from September 2023 was used in the test set. In South Korea, seasonal divisions are generally categorized into four periods, each lasting three months, starting from March [52].

Fig. 8(a) shows the results of various training strategies applied to the GIST dataset for the latest one-year period. Freezing seven encoder blocks yielded the lowest prediction error. Consistent with previous experiments, transfer learning outperformed both training solely on the target dataset and zero-shot inference. However, this observed trend differs from that shown in the GIST graph in Fig. 4, which presents results based on splitting the test set by PV arrays. For the temporal robustness experiment, freezing fewer blocks generally led to better optimal results, and the variation in performance across different numbers of frozen blocks was relatively smaller. Fig. 8(b) presents results from the same one-year period, divided into three-month intervals. The model configuration applied to each season included freezing seven encoder blocks, which yielded the best performance in our previous experiments. Notably, the lowest prediction error was observed in autumn, a season characterized by cool temperatures, abundant sunshine, and minimal day-to-day weather variation. Conversely, prediction errors were comparatively higher in summer—which includes the monsoon season with heavy rainfall—and in spring, which involves frequent yellow dust and significant weather fluctuations.



(a) Temporal evaluation for the latest 1 year



(b) Forecasting performance by seasons

FIGURE 8. **Long-term robustness evaluation. (a) Temporal evaluation for the latest year, showing MAPE for various transfer methods, where 0–7 indicate the number of the frozen encoder blocks (0 is full fine-tuning) and other labels represent specific methods. (b) Seasonal forecasting results for the same period, divided into three-month segments.**

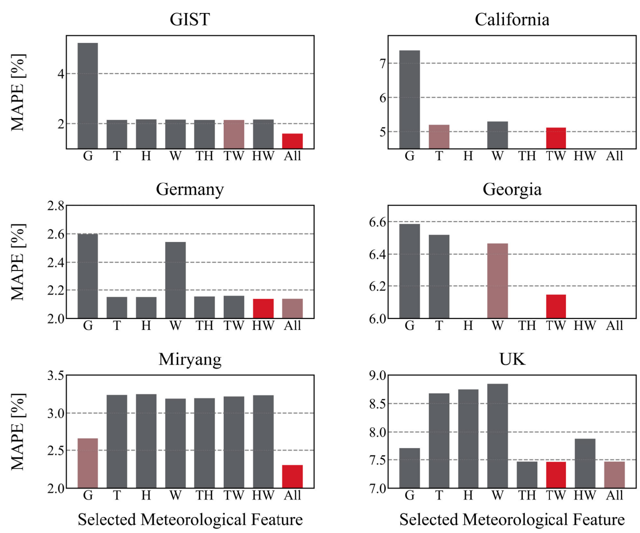


FIGURE 9. **The ablation study of meteorological features selection. The x-axis indicates which meteorological features—GHI (G), air temperature (T), RH, and wind speed (W)—were used for training in each experiment. The GHI was included in all experiments. California and Georgia datasets do not have RH observations.**

4) ABLATION STUDY OF METEOROLOGICAL FEATURES SELECTION

In section III-B, the relationship between the climate variables used for training and solar power generation was examined using Pearson correlation following data preprocessing. However, the actual relationship between the input features for power generation, which depend on combinations of climate variables, is more complex and often nonlinear. To further explore this complexity, this study conducted additional experiments across all datasets, evaluating the power generation prediction results based on different combinations of climate variables.

Fig. 9 shows the effect of different meteorological feature combinations on prediction performance. Feature combinations were selected from GHI, air temperature, RH, and wind speed. As GHI is inherently essential for solar power generation, it was included in all combinations. For the California and Georgia datasets, experiments involving RH were not conducted owing to the absence of observed values for this feature. Each dataset's experiments were conducted utilizing the most effective freezing number determined in Section V-B1. Across most experiments, using all available features yielded the best results. For the Germany dataset, the combination of GHI, air temperature, RH, and wind speed achieved the lowest MAPE at 2.1370%. However, this result was not a significantly difference from the "All" combination, which yielded a MAPE of 2.1378%.

Analyzing the influence of each feature by comparisons with Table 4 reveals that, unless the absolute correlation between a meteorological feature and AP is notably low (as seen with wind speed in Germany), combining it with GHI generally improves prediction performance. Nevertheless, in some cases, such as Miryang and the UK, incorporating additional features alongside GHI sometimes resulted in worse performance than using GHI alone. This observation is attributed to the fact that the correlation between GHI and AP in these two datasets is lower compared with the other datasets. Despite these variations, a common inference across all datasets is that if GHI's correlation with AP is approximately 0.9 or higher and other meteorological features have correlations of 0.1 or higher, then including them in the training process is beneficial.

VI. DISCUSSION

This study demonstrated the effectiveness of transfer learning for improving PV generation prediction models using large-scale datasets. While previous works have primarily focused on intra-regional transfer or simple source-target setups [8], [10], [11], [12], this study highlights the feasibility and effectiveness of cross-continental transfer learning tailored for diverse climate zones. Specifically, the proposed work utilized the large-scale DKASC dataset [35] from Australia as a source and data from diverse climate zones—including Korea, Germany, the United States, and the United Kingdom—as targets. To handle missing values and outliers in raw data, this study proposed the data preprocessing procedure that considers physical and logical constraints and establishes a preprocessing protocol that can be consistently applied to all datasets.

The experimental results highlighted several key findings. First, pre-training on a large source dataset consistently improved the prediction performance across all target domains, with MAPE improvements exceeding 20% for target datasets containing identical input features to the source dataset. Second, source dataset size was shown to influence prediction performance: reducing the source data size to 1/8 and 1/16 increased the MAPE by 2.70% and 3.35%, respectively. Additionally, expanding the target training period from 3 to 12 months generally improved prediction performance, emphasizing the importance of both a sufficiently large source dataset and an appropriate period for the target dataset. Regarding data quality, the result indicated that it is important to include characteristics that are highly correlated with power generation, such as GHI, temperature, RH, and wind speed, rather than just the amount of data. This finding was evidenced by transfer learning from California and Georgia, where RH was unavailable. Regarding transfer learning strategy, the optimal approach depends on the characteristics of the target dataset. Freezing only a few layers was adequate when sufficient data was available (80,000 samples or more), while linear probing or freezing most layers was advantageous when data was scarce. Particularly, zero-shot prediction performed better than training on the target dataset alone when data availability was limited (less than one year).

The GIST dataset was chosen explicitly to evaluate temporal robustness owing to its notable climatic variability. It experiences diverse weather conditions such as spring yellow dust, summer monsoon and typhoons, and winter snow, along with significant diurnal temperature fluctuations (over 15 °C) in spring, autumn, and winter. The high meteorological diversity and significant climatic difference from the source dataset made it ideal for assessing the effectiveness of transfer learning. In temporal evaluation, the GIST dataset showed a MAPE of 6.25%, which is more comparable to California's 5.12% and Georgia's 6.15% (both split by period) than its own PV array-based split (1.60%). This notable difference in prediction error between the PV array-based and chronologically-based test sets suggests that, in the PV array-based test set, the model might have implicitly utilized meteorological information from the same period across different PV arrays during training. Nevertheless, the compatible prediction error in the GIST dataset's temporal evaluation indicates the model's long-term robustness even in regions with numerous influencing meteorological variables. Further performance improvements are anticipated if accurate weather forecasting information can be utilized as an additional input for meteorological features during the prediction period.

Regarding the encoder block freeze strategy, the GIST dataset, despite being notably larger than other target datasets, exhibited differences in optimal freeze levels between PV array-based and chronologically-based test set divisions. Comparing Fig. 4 and 8(a), the optimal freeze strategy for the latest one-year period test set was relatively similar to other temporally split datasets. This observation suggests that unfreezing only a few layers during transfer learning might be a more effective strategy regardless of dataset size. However, confirming this strategy requires further validation. This study has some limitations. Notable is the absence of RH features in the California and Georgia datasets. A more rigorous evaluation would require similar dataset volumes and meteorological measurement methods across diverse climates.

Future work will further enhance PV power forecasting accuracy via incorporating advanced meteorological forecasting data and developing improved model architectures that explicitly account for climate characteristics. It can also involve a more detailed temporal analysis by investigating seasonal and monthly variations in photovoltaic (PV) power generation, which can provide deeper insights into the model's behavior under varying climatic conditions. Additionally, while the current study applied the proposed framework to two representative regions in South Korea, the methodology will be extended to other regions across the country to further evaluate the spatial generalizability of the model. Moreover, considering the growing deployment of other renewable sources such as onshore and offshore wind power, future research will explore the applicability of transfer learning to wind energy forecasting, assessing its effectiveness and scalability in different operational environments. These directions are expected to broaden the utility of the proposed approach and contribute to more resilient and adaptable renewable energy forecasting frameworks.

VII. CONCLUSION

This study makes several key theoretical and practical contributions to PV power forecasting. It is the first to achieve effective cross-continental PV forecasting using a million-scale dataset (over 4.5 million data points), demonstrating robust generalizability across diverse Köppen–Geiger climate zones—from Australia's arid climate and Korea's monsoon climate to Germany's temperate conditions and diverse North American regions. The proposed systematic approach enables reliable PV forecasting in data-scarce regions, potentially accelerating renewable energy adoption in emerging markets. Furthermore, this study offers practical guidelines showing that freezing only a few transformer layers is optimal for large target datasets (over 80,000 data points), while freezing most layers is more effective for smaller ones. Additionally, the analysis revealed that model performance can vary significantly depending on the test set split methods (e.g., by PV array vs. chronologically), providing methodological insights for rigorous evaluation standards in future PV forecasting research.

The results of this study highlight the importance of utilizing large datasets and transfer learning in PV generation forecasting. They indicate that effective forecasting is feasible, especially in data-limited environments. This approach is expected to significantly benefit the initial operation of new PV plants worldwide and power generation forecasting in data-poor regions.

ACKNOWLEDGMENT

The authors gratefully acknowledge DKASC for providing the source dataset that served as a significant foundation for this research. They also sincerely thank the Facilities Management Team of GIST and Korea Electric Power Corporation (KEPCO) for their invaluable assistance with collecting and managing PV data in South Korea. Furthermore, they would like to thank the PatchTST Research Team for sharing their excellent code, which greatly facilitated the implementation of their model. They appreciated the high-performance GPU computing support HPC-AI Open Infrastructure via GIST SCENT.

REFERENCES

[1] (2025). *Climate Acion Net Zero*. Accessed: Jun. 21, 2025. [Online]. Available: https://www.un.org/en/climatechange/net-zero-coalition

[2] L. Bird, M. Milligan, and D. Lew, "Integrating variable renewable energy: Challenges and solutions," NREL, Golden, CO, USA, Tech. Rep. NREL/TP-6A20-60451. [Online]. Available: https://www.osti.gov/biblio/1097911

[3] D. Yang, J. Lee, N. C. Song, S. Lee, S. Kim, S. Lee, and S. Choi, "Patent analysis on green hydrogen technology for future promising technologies," *Int. J. Hydrogen Energy*, vol. 48, no. 83, pp. 32241–32260, Oct. 2023.

[4] *Forecasting Process of Electricity Generation*. Accessed: Dec. 8, 2024. [Online]. Available: https://der.kmos.kr/intro/fr_intro_view10.do#none

[5] I. Ashraf and A. Chandra, "Artificial neural network based models for forecasting electricity generation of grid connected solar PV power plant," *Int. J. Global Energy Issues*, vol. 21, no. 1, pp. 119–130, 2004.

[6] A. Chaouachi, R. M. Kamel, R. Ichikawa, H. Hayashi, and K. Nagasaka, "Neural network ensemble-based solar power generation short-term forecasting," *World Acad. Sci., Eng. Technol.*, vol. 54, pp. 54–59, Feb. 2009.

[7] K. Wang, X. Qi, and H. Liu, "Photovoltaic power forecasting based LSTM-convolutional network," *Energy*, vol. 189, Dec. 2019, Art. no. 116225.

[8] E. Sarmas, N. Dimitropoulos, V. Marinakis, Z. Mylona, and H. Doukas, "Transfer learning strategies for solar power forecasting under data scarcity," *Sci. Rep.*, vol. 12, no. 1, p. 14643, Aug. 2022.

[9] Y. Tang, K. Yang, S. Zhang, and Z. Zhang, "Photovoltaic power forecasting: A hybrid deep learning model incorporating transfer learning strategy," *Renew. Sustain. Energy Rev.*, vol. 162, Jul. 2022, Art. no. 112473.

[10] S. M. Miraftabzadeh, C. G. Colombo, M. Longo, and F. Foiadelli, "A day-ahead photovoltaic power prediction via transfer learning and deep neural networks," *Forecasting*, vol. 5, no. 1, pp. 213–228, Feb. 2023.

[11] B. Kim, K. Belkilani, G. Heilscher, M.-O. Otto, J.-S. Huh, and D. Suh, "Integrated spatiotemporal hybrid solar PV generation forecast between countries on different continents using transfer learning method," *IEEE Access*, vol. 13, pp. 2486–2502, 2025.

[12] J. Depoortere, J. Driesen, J. Suykens, and H. S. Kazmi, "SolNet: Open-source deep learning models for photovoltaic power forecasting across the globe," *Int. J. Forecasting*, vol. 41, no. 3, pp. 1223–1236, Jul. 2025.

[13] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," 2022, *arXiv:2211.14730*.

[14] M. F. Tahir, A. Tzes, and M. Z. Yousaf, "Enhancing PV power forecasting with deep learning and optimizing solar PV project performance with economic viability: A multi-case analysis of 10MW Masdar project in UAE," *Energy Convers. Manage.*, vol. 311, Jul. 2024, Art. no. 118549.

[15] H. Abdulla, A. Sleptchenko, and A. Nayfeh, "Photovoltaic systems operation and maintenance: A review and future directions," *Renew. Sustain. Energy Rev.*, vol. 195, May 2024, Art. no. 114342.

[16] B. Espinar, J.-L. Aznarte, R. Girard, A. M. Moussa, and G. Kariniotakis, "Photovoltaic forecasting: A state of the art," in *Proc. 5th Eur. PV-Hybrid Mini-Grid Conf.*, 2010, p. 250.

[17] A. Livera, M. Theristis, L. Micheli, E. F. Fernandez, J. S. Stein, and G. E. Georghiou, "Operation and maintenance decision support system for photovoltaic systems," *IEEE Access*, vol. 10, pp. 42481–42496, 2022.

[18] B. Chowdhury and S. Rahman, "Forecasting sub-hourly solar irradiance for prediction of photovoltaic output," in *Proc. 19th IEEE Photovoltaic Spec. Conf.*, Mar. 1987, pp. 171–176.

[19] R. Asghar, F. R. Fulginei, M. Quercio, and A. Mahrouch, "Artificial neural networks for photovoltaic power forecasting: A review of five promising models," *IEEE Access*, vol. 12, pp. 90461–90485, 2024.

[20] P. Singla, M. Duhan, and S. Saroha, "A comprehensive review and analysis of solar forecasting techniques," *Frontiers Energy*, vol. 16, no. 2, pp. 1–37, Apr. 2022.

[21] U. K. Das, K. S. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Y. I. Idris, W. Van Deventer, B. Horan, and A. Stojcevski, "Forecasting of photovoltaic power generation and model optimization: A review," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 912–928, Jan. 2018.

[22] C. Monteiro, L. A. Fernandez-Jimenez, I. J. Ramirez-Rosado, A. Muñoz-Jimenez, and P. M. Lara-Santillan, "Short-term forecasting models for photovoltaic plants: Analytical versus soft-computing techniques," *Math. Problems Eng.*, vol. 2013, May 2013, Art. no. 767284.

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[24] A. Gensler, J. Henze, B. Sick, and N. Raabe, "Deep learning for solar power forecasting—An approach using AutoEncoder and LSTM neural networks," in *Proc. IEEE Int. Conf. Syst., man, Cybern. (SMC)*, Oct. 2016, pp. 2858–2865.

[25] V. De, T. T. Teo, W. L. Woo, and T. Logenthiran, "Photovoltaic power forecasting using LSTM on limited dataset," in *Proc. IEEE Innov. Smart Grid Technol.-Asia (ISGT Asia)*, May 2018, pp. 710–715.

[26] M. Abdel-Nasser and K. Mahmoud, "Accurate photovoltaic power forecasting models using deep LSTM-RNN," *Neural Comput. Appl.*, vol. 31, no. 7, pp. 2727–2740, Jul. 2019.

[27] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," *Adv. Neural Inf. Process. Syst.*, pp. 22419–22430, Jul. 2021.

[28] O. Khouili, M. Hanine, M. Louzazni, M. A. L. Flores, E. G. Villena, and I. Ashraf, "Evaluating the impact of deep learning approaches on solar and photovoltaic power forecasting: A systematic review," *Energy Strategy Rev.*, vol. 59, May 2025, Art. no. 101735.

[29] H. A. Kazem, J. H. Yousif, M. T. Chaichan, A. H. A. Al-Waeli, and K. Sopian, "Long-term power forecasting using FRNN and PCA models for calculating output parameters in solar photovoltaic generation," *Heliyon*, vol. 8, no. 1, Jan. 2022, Art. no. e08803.

[30] A. H. A. Al-Waeli, H. A. Kazem, J. H. Yousif, M. T. Chaichan, and K. Sopian, "Mathematical and neural network models for predicting the electrical performance of a PV/T system," *Int. J. Energy Res.*, vol. 43, no. 14, pp. 8100–8117, Aug. 2019.

[31] O. Khouili, M. Hanine, and M. Louzazni, "Harnessing principal component analysis and artificial neural networks for accurate solar radiation prediction," *Int. J. Energy Res.*, vol. 2025, no. 1, Jan. 2025, Art. no. 5846114.

[32] M. Despotovic, C. Voyant, L. Garcia-Gutierrez, J. Almorox, and G. Notton, "Solar irradiance time series forecasting using auto-regressive and extreme learning methods: Influence of transfer learning and clustering," *Appl. Energy*, vol. 365, Jul. 2024, Art. no. 123215.

[33] H. E. Beck, T. R. McVicar, N. Vergopolan, A. Berg, N. J. Lutsko, A. Dufour, Z. Zeng, X. Jiang, A. I. van Dijk, and D. G. Miralles, "High-resolution (1 km) k "High-resolution (1 km) köppen-geiger maps for 1901–2099 based on constrained CMIP6 projections," *Sci. data*, vol. 10, no. 1, p. 724, 2023.

[34] H. E. Beck, N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood, "Present and future köppen-geiger climate classification maps at 1-km resolution," *Sci. data*, vol. 5, no. 1, pp. 1–12, 2018.

[35] *Desert Knowledge Australia Solar Centre*. Accessed: Nov. 20, 2024. [Online]. Available: https://dkasolarcentre.com.au/

[36] (2024). *Open Power System Data*. Accessed: Nov. 21, 2024. [Online]. Available: https://data.open-power-system-data.org/household_data/

[37] C. Deline, K. Perry, M. Deceglie, M. Müller, W. Sekulic, and D. Jordan. (Dec. 2021). *Photovoltaic Data Acquisition (PVDAQ) Public Datasets*. [Online]. Available: https://data.openei.org/submissions/4568

[38] (2024). *Photovoltaic (PV) Solar Panel Energy Generation Data*. Accessed: Nov. 22, 2024. [Online]. Available: https://data.london.gov.uk/dataset/photovoltaic–pv–solar-panel-energy-generation-data

[39] (2024). *Open Met Data Portal*. Accessed: Nov. 22, 2024. [Online]. Available: https://data.kma.go.kr/cmmn/main.do

[40] (2024). *Rural Development Administration National Academy of Agricultural Sciences_search for Detailed Observation Data of Agricultural Weather*. Accessed: Nov. 21, 2024. [Online]. Available: https://www.data.go.kr/en/data/15078194/openapi.do

[41] (2024). *Dwd's Open Data Server Climate Data Center*. Accessed: Nov. 21, 2024. [Online]. Available: https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/

[42] Z. A. Khan, T. Hussain, and S. W. Baik, "Dual stream network with attention mechanism for photovoltaic power forecasting," *Appl. Energy*, vol. 338, May 2023, Art. no. 120916.

[43] T. Kim, J. Kim, Y. Tae, C. Park, J.-H. Choi, and J. Choo, "Reversible instance normalization for accurate time-series forecasting against distribution shift," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–25. [Online]. Available: https://openreview.net/forum?id=cGDAkQo1C0p

[44] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," 2016, *arXiv:1606.08415*.

[45] J. Zhang, B.-M. Hodge, A. Florita, S. Lu, H. F. Hamann, and V. Banunarayanan, "Metrics for evaluating the accuracy of solar power forecasting," NREL, Golden, CO, USA, Tech. Rep. NREL/CP-5500-60142, 2013.

[46] *Guidelines for the Management and Operation of the Renewable Energy Feed-in Obligation and the Fuel Blending Obligation (ministry of Trade, Industry and Energy Notification, nos. 2024-017)*. Accessed: Nov. 26, 2024. [Online]. Available: https://www.motie.go.kr/kor/article/ATCL0c554f816/64538/view#

[47] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" in *Proc. AAAI Conf. Artif. Intell.*, 2023, vol. 37, no. 9, pp. 11121–11128.

[48] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008.

[49] S. Wang, H. Wu, X. Shi, T. Hu, H. Luo, L. Ma, J. Y. Zhang, and J. Zhou, "TimeMixer: Decomposable multiscale mixing for time series forecasting," 2024, *arXiv:2405.14616*.

[50] Y. Liu, H. Zhang, C. Li, X. Huang, J. Wang, and M. Long, "Timer: Generative pre-trained transformers are large time series models," 2024, *arXiv:2402.02368*.

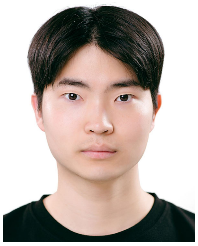
[51] G. Woo, C. Liu, A. Kumar, C. Xiong, S. Savarese, and D. Sahoo, "Unified training of universal time series forecasting transformers," 2024, *arXiv:2402.02592*.

[52] *Climate:Visitkorea*. Accessed: Jun. 6, 2025. [Online]. Available: https://english.visitkorea.or.kr/svc/contents/contentsView.do?menuSn=462&vcontsId=140636

SEONGHO BAK (Member, IEEE) received the B.S. degree in mechanical engineering from Jeonbuk National University (JBNU), in 2014, and the M.S. degree from the Department of Robotics and Mechatronics Engineering, Daegu Gyeongbuk Institute of Science and Technology (DGIST), in 2016. He is currently pursuing the Ph.D. degree in intelligent robotics with the School of Integrated Technology, Gwangju Institute of Science and Technology (GIST). From 2016 to 2019, he was a Researcher with the Division of Intelligent Robots, DGIST. His current research interests include renewable energy and artificial intelligence for robotics.

SOWON CHOI (Graduate Student Member, IEEE) received the B.S. degree in IT engineering, in 2022. She is currently pursuing the M.S. degree in intelligence robotics with the School of Integrated Technology, Gwangju Institute of Science and Technology (GIST). Her current research interests include renewable energy, and artificial intelligence for healthcare and robotics.

DONGUK YANG received the B.S. and M.S. degrees in mechanical engineering from Chonnam National University and the Ph.D. degree in mechanical engineering from Gwangju Institute of Science and Technology. He is currently at Korea Electric Power Corporation. His research interests include green hydrogen, life cycle analysis, operations research, reinforcement learning, and renewable energy. His current work focuses on advancing green hydrogen technologies, optimizing their integration into power systems, and exploring innovative approaches for hydrogen safety and sustainability.



DOYOON KIM is currently pursuing the B.S. degree in electronics engineering and computer science from Gwangju Institute of Science and Technology (GIST). His research interests include multimodality and efficient learning.

HEESEON RHO (Graduate Student Member, IEEE) received the B.S. degree in electronics engineering from Inha University, Incheon, South Korea, in 2020, and the M.S. degree in intelligent robotics from the School of Integrated Technology, Gwangju Institute of Science and Technology (GIST), in 2023, where she is currently pursuing the Ph.D. degree with the School of Integrated Technology under the Intelligent Robotics Program. Her research interests include multimodality and robotic vision.

KYOOBIN LEE (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in mechanical engineering from Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 2008. From 2008 to 2010, he was a Postdoctoral Scholar with the Center for Neuroscience, KIST. From 2012 to 2017, he was a Principal Researcher with the Samsung Advanced Institute of Technology. Since 2017, he has been an Assistant Professor with the School of Integrated Technology, Gwangju Institute of Science and Technology (GIST). His research interests include vision recognition using deep learning, robot control applications using computer vision, and split learning for neural networks of cloud computing applications. He is an Editor of Korea Robotics Society Review.