

Cross-Attention Adaptive Fusion for Multi-Step Photovoltaic Power Forecasting from Real Multi-Site Measurements

Waseem Ullah, Muhammad Saeed, Mohsen Guizani

Department of Machine Learning

Mohamed Bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE

Abstract—Accurate multi-step photovoltaic (PV) power forecasting is essential for power-system scheduling, reserve management, and the reliable integration of solar generation. However, PV output is difficult to predict because it is governed by rapidly changing meteorological conditions, nonlinear interactions among environmental variables, and site-dependent operating characteristics. This paper proposes an adaptive cross-attention similarity learning framework for direct multi-step PV active-power forecasting. The framework integrates residual temporal convolutional blocks with a global-context branch to capture short-term fluctuations and broader sequence-level dependencies. A cross-attention module captures the interactions between temporal and contextual features, while an adaptive aggregation component weighs their respective impacts on the final prediction. The full 12-step forecast horizon is produced in a single pass, which eliminates the error accumulation typically seen in recursive forecasting. The proposed framework is tested on over 2.69 million timestamped records obtained from three operational PV plants: Site 1A Trina, Site 1B Trina, and Site 2 eco-Kinetics. We employ a chronological evaluation procedure that prevents data leakage, encompassing missing-value imputation, feature normalization, and the creation of supervised sliding windows. By jointly modeling local dynamics, global operating context, and cross-variable dependencies, the proposed approach provides a unified and reproducible solution for robust PV forecasting across heterogeneous sites and real-world data conditions.

Index Terms—Photovoltaic forecasting, solar power prediction, cross attention, residual convolution, adaptive feature fusion, time-series forecasting, renewable energy informatics.

I. INTRODUCTION

Renewable energy (RE) has increasingly penetrated conventional power systems, making them more dynamic and decentralized smart grids. Therefore, electricity generation and demand must be coordinated in a reliable manner for operational scheduling, energy trading, allocation of reserves and grid stability. Due to the abundance, cleanliness, and geographic availability of solar energy, photovoltaic (PV) generation is becoming a widespread technology among the available renewable energy technologies [1], [2]. The PV output is, however, very intermittent, and depends on the irradiance, the temperature, the humidity, the wind, the cloud movement, and the operating parameters of the plant. The different configurations add to the confusion and uncertainty in grid operation and require precise forecasts to manage energy

storage, dispatch electricity, schedule transmission, and plan energy markets [3].

Numerous machine learning techniques have been studied for predicting renewable generation, such as support vector regression, decision trees, artificial neural network, extreme learning machines, and fuzzy neural networks. These techniques can be used for modeling nonlinear relationships, but they are likely to be hard to engineer and can generalize poorly in varying operating conditions. Deep-learning models, including recurrent neural networks, long short-term memory networks, gated recurrent units, echo state networks, and convolutional neural networks, have consequently received considerable attention [4], [5]. Hybrid models such as CNN-LSTM, LSTM-CNN, ConvLSTM, and related architectures have further improved forecasting performance by combining local and temporal feature extraction [6]–[8]. Nevertheless, many existing methods focus on one-step prediction, employ limited historical context, and insufficiently model interactions among heterogeneous meteorological and power variables.

To address these limitations, this study develops an adaptive cross-attention similarity learning framework for direct multi-step PV active-power forecasting. The proposed system combines residual temporal convolution and global-context modeling, where the former can learn short-term variations and the latter can learn the longer range sequence-level dependencies. Cross-attention is introduced to capture interactions among complementary feature representations, and adaptive aggregation is adopted to weigh the relative importance of the different feature representations for the final forecast. The entire prediction horizon is estimated directly, thereby eliminating the error propagation problem of recursive prediction. This formulation enables a consistent learning strategy to be followed for learning of local temporal dynamics, global operating context, and cross-variable dependency from the heterogeneous PV measurements.

The main contributions of this study are summarized as follows:

- We propose a hybrid architecture that combines residual temporal convolution, global-context modeling, and cross-attention for direct multi-step PV power forecasting.

- We introduce an adaptive aggregation mechanism that dynamically weights global and attention-enhanced temporal representations.
- We develop a leakage-free multi-site pipeline for missing-value handling, normalization, and supervised sequence generation.
- We evaluate the framework on large-scale real-world PV datasets under both single-site and multi-site settings using a reproducible implementation.

II. RELATED WORK

Renewable-energy forecasting methods generally include persistence, physical, statistical, machine-learning, and deep-learning approaches [9]. Persistence and statistical models are computationally efficient but often fail to capture nonlinear and rapidly changing power patterns [10], [11]. Physical models can incorporate meteorological knowledge, although their computational and data requirements limit practical deployment [10]. Machine-learning methods improve nonlinear modelling but frequently depend on handcrafted features.

Deep-learning models, particularly CNNs, LSTMs, GRUs, and ESNs, have shown stronger representation capability for power forecasting [12]. Hybrid architectures such as CNN-LSTM, SARIMA-LSTM, CNN-BiLSTM, CNN-GRU, and ESN-CNN further improve performance by combining spatial and temporal feature extraction [13]–[15]. More recently, attention-based and residual models have been introduced to emphasize informative features and improve robustness [16], [17]. Khan and Choi proposed a dual-stream network that combines ESN, residual CNN, self-attention, spatial attention, and PCA-based feature reduction for power forecasting [2]. However, many existing methods still rely on fixed or direct feature fusion, which may inadequately model interactions between local temporal patterns and global context. This limitation motivates the proposed adaptive cross-attention framework.

III. PROPOSED METHOD

A. Problem Formulation

Let $\mathbf{X} = \mathbf{x} * 1, \mathbf{x} * 2, \dots, \mathbf{x} * T \in \mathbb{R}^{T \times F}$ denote a multivariate photovoltaic (PV) time-series segment, where T represents the historical input length and F denotes the number of meteorological and electrical features. The objective is to learn a nonlinear mapping function $\mathcal{F} * \theta(\cdot)$ that predicts the future active-power sequence:

$$\hat{\mathbf{y}} = \mathcal{F}_\theta(\mathbf{X}), \quad \hat{\mathbf{y}} \in \mathbb{R}^H, \quad (1)$$

where H is the forecasting horizon. In this work, the model is configured for 12-step ahead active-power forecasting. The forecasting task is challenging because PV generation is affected by rapidly changing irradiance, temperature, wind conditions, and site-dependent operating characteristics. Therefore, the proposed model is designed to jointly learn local temporal variations, global sequence-level context, and adaptive interactions among latent representations.

B. Data Preprocessing and Sequence Construction

The preprocessing pipeline supports both single-site and multi-site PV datasets. Active power is used as the prediction target, while meteorological and electrical variables serve as input features. All variables are converted to numerical form, and missing values are reconstructed using linear interpolation followed by forward and backward filling to preserve temporal continuity. The input features and target values are independently normalized using min-max scaling:

$$\tilde{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \quad (2)$$

To avoid information leakage, the scaling parameters are estimated from the training data only. The continuous time series is transformed into supervised samples using a sliding window. For an input length T and forecasting horizon H , the i -th sample is defined as

$$\mathbf{X}_i = [\mathbf{x}_i, \mathbf{x}_{i+1}, \dots, \mathbf{x}_{i+T-1}], \quad (3)$$

$$\mathbf{y}_i = [y_{i+T}, y_{i+T+1}, \dots, y_{i+T+H-1}]. \quad (4)$$

This formulation enables direct multi-step PV active-power forecasting from historical observations.

C. Temporal Encoding and Global Context Modeling

Given an input sequence $\mathbf{X} \in \mathbb{R}^{T \times F}$, the proposed network first applies a one-dimensional convolution to project the raw multivariate observations into a latent temporal space:

$$\mathbf{E} = \sigma(\text{Conv1D}(\mathbf{X})), \quad \mathbf{E} \in \mathbb{R}^{T \times d}, \quad (5)$$

where T is the input length, F is the number of input variables, d is the latent dimension, and $\sigma(\cdot)$ denotes the ReLU activation. This operation preserves the temporal order while extracting short-range patterns from adjacent observations. The embedded sequence is then processed by residual temporal convolutional blocks to capture local dependencies and stabilize feature learning. For an input representation $\mathbf{U}$, each block is formulated as

$$\mathbf{R} = \sigma[\text{LN}(\mathbf{U} + \mathcal{P}(\text{Conv1D}_2(\text{Conv1D}_1(\mathbf{U}))))], \quad (6)$$

where $\mathcal{P}(\cdot)$ denotes a 1×1 projection convolution and $\text{LN}(\cdot)$ represents layer normalization. After stacking the residual blocks, the temporal representation is obtained as

$$\mathbf{H} = \mathcal{F}_{\text{res}}(\mathbf{E}), \quad \mathbf{H} \in \mathbb{R}^{T \times d}. \quad (7)$$

To complement the local temporal features, a global-context branch summarizes the overall operating state of the historical window through temporal average pooling followed by a dense transformation:

$$\mathbf{z}_g = \text{Dense}\left(\frac{1}{T} \sum_{t=1}^T \mathbf{H}_t\right). \quad (8)$$

The resulting representation captures the dominant generation trend, while the residual encoder preserves short-term fluctuations caused by rapid changes in irradiance and other environmental conditions.

D. Local Temporal Refinement Branch

In parallel with the global branch, a temporal refinement branch is used to preserve localized variations. The residual encoder output is passed through an additional one-dimensional convolutional layer, max-pooling operation, and layer normalization:

$$\mathbf{T}_r = \text{LN}(\text{MaxPool}(\sigma(\text{Conv1D}(\mathbf{H}))))). \quad (9)$$

Unlike the global branch, this path maintains time-dependent variations and emphasizes short-term transitions. Such localized temporal information is essential for forecasting PV active power under unstable weather conditions, cloud movement, and rapid irradiance changes.

E. Cross-Attention Fusion

To enrich the temporal refinement branch with broader contextual information, the model employs a multi-head cross-attention fusion module. The temporal representation is used as the query, while the residual encoder output is used as both key and value:

$$\mathbf{Q} = \mathbf{T} * r\mathbf{W} * \mathbf{Q}, \quad \mathbf{K} = \mathbf{H}\mathbf{W} * \mathbf{K}, \quad \mathbf{V} = \mathbf{H}\mathbf{W} * \mathbf{V}. \quad (10)$$

The attention-enhanced temporal representation is computed as

$$\mathbf{A} = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V}, \quad (11)$$

where d_k is the key dimension. The attention output is regularized using dropout and combined with the original temporal representation through residual normalization:

$$\mathbf{Z} * t = \text{LN}(\mathbf{T} * r + \text{Dropout}(\mathbf{A})). \quad (12)$$

This design allows the model to selectively attend to informative historical patterns while retaining the local temporal structure learned by the convolutional branch.

F. Adaptive Representation Aggregation

The global and temporal branches provide complementary information. However, their relative importance changes depending on the operating condition. For example, during stable irradiance periods, the global context may dominate the forecasting decision, whereas during transient weather changes, localized temporal features may become more important. To address this issue, the proposed model uses an adaptive aggregation module.

Let $\mathbf{z} * g$ denote the global context representation and $\mathbf{z} * t$ denote the cross-attended temporal representation after global pooling. Each representation is passed through a lightweight gating network:

$$g_m = \phi_m(\mathbf{z} * m), \quad m \in g, t, \quad (13)$$

where $\phi_m(\cdot)$ consists of dense transformations with nonlinear activation. The learned gate scores are concatenated and normalized using the softmax function:

$$\alpha_m = \frac{\exp(g_m)}{\sum_{*j \in g, t} \exp(g_j)}. \quad (14)$$

The final fused representation is obtained as

$$\mathbf{z} * f = \alpha_g \mathbf{z} * g + \alpha_t \mathbf{z} * t. \quad (15)$$

Compared with simple concatenation or averaging, adaptive aggregation enables the model to dynamically adjust the contribution of each representation according to the input sequence.

G. Forecasting Head

The fused representation is flattened and passed through a fully connected forecasting head. A dense layer with nonlinear activation is used to learn high-level interactions among the fused features, followed by dropout regularization to reduce overfitting. The final output layer uses a linear activation function to generate the 12-step active-power prediction:

$$\hat{\mathbf{y}} = \mathbf{W} * o\sigma(\mathbf{W} * f\mathbf{z} * f + \mathbf{b} * f) + \mathbf{b}_o. \quad (16)$$

The use of a linear output layer is suitable for continuous-valued regression because the target variable corresponds to normalized PV active power.

H. Model Optimization

The network is trained as a supervised multi-step regression model. The chronological order of the time series is preserved by using a non-shuffled train-test split, which avoids information leakage from future samples into the training stage. The model is optimized using the Adam optimizer with the mean squared error loss:

$$\mathcal{L} * \text{MSE} = \frac{1}{NH} \sum_{*i} 1^N \sum_{h=1}^H (y_{i,h} - \hat{y}_{i,h})^2, \quad (17)$$

where N is the number of training samples and H is the forecasting horizon. Mean absolute error and mean squared error are also monitored during training. After training, the model is saved together with the evaluation metrics and training history to support reproducibility and further analysis.

IV. EXPERIMENTAL SETUP

A. Datasets and Training Protocol

Experiments were conducted on three real-world photovoltaic (PV) datasets: Site 1A-Trina, Site 1B-Trina, and Site 2-EcoKinetics. Active power was selected as the prediction target, while the available meteorological and electrical variables were used as input features. Missing observations were reconstructed using linear interpolation followed by forward and backward filling. All features were normalized through min-max scaling, and supervised samples were generated using a sliding-window strategy. The chronological order of the observations was preserved during data partitioning to prevent information leakage. The input sequence length and forecasting horizon were both set to 12, enabling direct 12-step prediction. The model was implemented in TensorFlow/Keras and optimized using Adam with a learning rate of 10^{-3} [18]. Mean squared error was employed as the training objective, and the default batch size was set to 64.

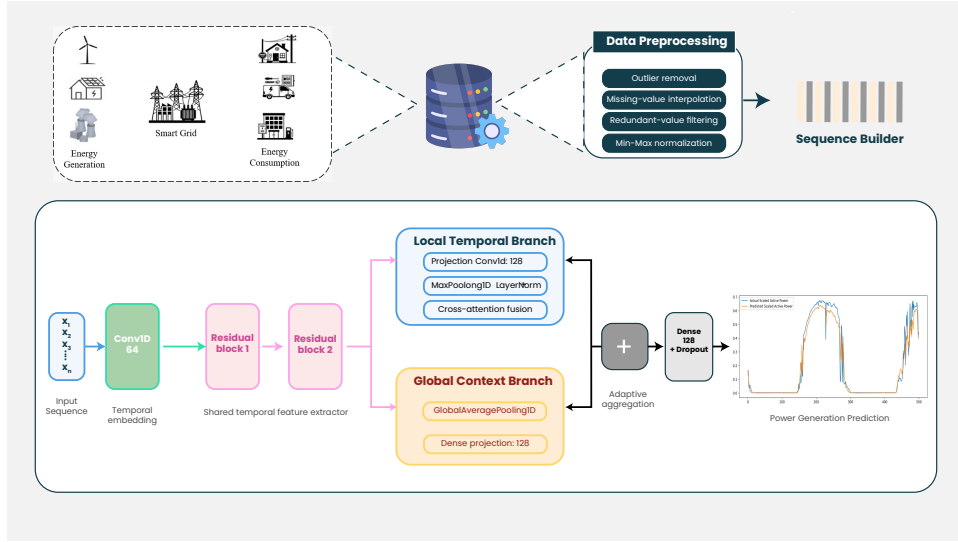


Fig. 1. Overall architecture of the proposed adaptive cross-attention similarity forecasting model. The network integrates temporal embedding, residual convolutional encoding, global context extraction, local temporal refinement, cross-attention fusion, and adaptive representation aggregation for multi-step PV active-power forecasting.

B. Evaluation Metrics

Forecasting performance was evaluated using mean bias error (MBE), root mean squared error (RMSE), mean absolute error (MAE), and normalized root mean squared error (NRMSE). MBE measures the systematic prediction bias, RMSE gives higher penalty to large errors, MAE reflects the average absolute forecasting deviation, and NRMSE provides a normalized error measure for comparison across datasets. The actual and predicted active-power values are denoted by y_i and $\hat{y}_i$, respectively, and n represents the number of test samples. Lower absolute MBE, RMSE, MAE, and NRMSE values indicate better forecasting performance.

V. RESULTS AND DISCUSSION

A. Forecasting Performance

The proposed framework incorporates residual temporal encoding, global-context modeling, cross-attention fusion and adaptive aggregation to enhance multi-step PV active-power forecasting. The short-term changes in PV generation are encoded in the residual temporal encoder, and the global-context branch represents the general operating state of the input window. The interaction between these complementary representations is enhanced by the cross-attention, while the contribution of each individual representation to the final forecast is dynamically changed by aggregating. This design is especially helpful for PV forecasting since the generation is smooth and stable during sunny periods, but very dynamic during the cloudy or rainy periods.

B. Comparison with Existing Methods

Table I compares the proposed model with representative convolutional, recurrent, hybrid, and Transformer-based forecasting methods across the three DKASCASA PV datasets. On DKASCASA-1B, the proposed model achieves an RMSE

of 0.03967 kW and an MAE of 0.02431 kW. Compared with DSDN, these values correspond to reductions of 19.19% and 23.41%, respectively. The absolute MBE is also reduced from 0.0414 to 0.0245 kW, indicating lower prediction bias relative to the strongest existing baseline. On DKASCASA-1A, the proposed framework obtains an RMSE of 0.0593 kW, an MAE of 0.0366 kW, and an NRMSE of 0.0536. Compared with DSDN, the proposed model reduces RMSE, MAE, and NRMSE by 2.79%, 1.51%, and 20.80%, respectively. These results show that the proposed model improves both absolute and normalized forecasting accuracy on this dataset. The largest improvements are observed on DKASCASA-2, where the proposed model achieves an RMSE of 0.0183 kW, an MAE of 0.0100 kW, and an NRMSE of 0.02312. Compared with DSDN, these values represent reductions of 42.81%, 50.96%, and 63.68%, respectively. In addition, the MBE magnitude decreases from 0.0854 to 0.0032 kW, showing a substantial reduction in prediction bias. Overall, the results demonstrate that the proposed architecture provides competitive and generally improved forecasting performance across heterogeneous PV sites.

C. Ablation Analysis

To validate the role of each major component of the proposed framework, an ablation study was carried out on DKASCASA-1B data. To ensure a fair comparison, the same preprocessing pipeline, chronological split, length of input, forecasting horizon, optimizer, batch size and training settings were used for all variants. In each of the experiments, only one component was removed or substituted. As shown in Table II, the full model achieves the best performance, with an RMSE of 0.039667, an MAE of 0.024305, and an MBE of 0.0245. These values are comparable with the ones shown in Table I which shows that the proposed model has the minimum

TABLE I

QUANTITATIVE PERFORMANCE COMPARISON OF THE PROPOSED FORECASTING FRAMEWORK AGAINST REPRESENTATIVE DEEP-LEARNING METHODS ON THE THREE DKASCASA PV DATASETS. THE LOWEST ERROR FOR EACH METRIC IS HIGHLIGHTED IN BOLD, WHILE THE SECOND-LOWEST RESULT IS UNDERLINED.

Dataset	Method	MBE (kW)	RMSE (kW)	MAE (kW)	NRMSE
DKASCASA-1B	ESNCNN [14]	0.0029	0.2153	0.1466	—
	GRU-CNN [19]	0.0235	0.1727	0.0923	—
	CNNGRU [14]	-0.0300	0.4006	0.3247	—
	LSTMCNN [14]	0.0948	0.3535	0.2116	—
	CNNLSTM [14]	0.0113	0.3967	0.2408	—
	Transformer [20]	-0.01759	0.0920	0.05493	0.09239
	Proposed DSDN [2]	-0.0414	0.04909	0.03174	0.07694
	Our Proposed Model	0.0245	0.03967	0.02431	0.07553
DKASCASA-1A	LSTM-CNN [6]	—	0.6210	0.2210	—
	CNN-LSTM [6]	—	0.3430	0.1260	—
	ESNCNN [14]	-0.0017	0.1610	0.0845	—
	CNN [6]	—	0.3090	0.1750	—
	CNNGRU [14]	0.0172	0.2776	0.1531	—
	GRU-CNN [19]	0.0171	0.1468	0.0742	—
	DSN [21]	0.0019	0.0721	0.0383	0.2679
	Transformer [20]	-0.00758	0.0870	0.05359	0.09737
	Proposed DSDN [2]	-0.00148	0.0610	0.03716	0.06768
	Our Proposed Model	0.00133	0.0593	0.0366	0.0536
DKASCASA-2	WPD-LSTM [22]	0.0067	0.2357	—	—
	LSTM [22]	-0.0840	1.0382	—	—
	ESNCNN [14]	0.0048	0.1432	0.0602	—
	GRU [22]	0.1206	1.0351	—	—
	RNN [22]	-0.01442	1.0581	—	—
	MLP [22]	0.1995	1.0861	—	—
	GRU-CNN [19]	-0.0641	0.1646	0.1157	—
	DSN [21]	-0.0019	0.0387	0.0250	0.2330
	Transformer [20]	-0.02274	0.0470	0.03377	0.09255
	Proposed DSDN	-0.0854	0.0320	0.02039	0.06367
	Our Proposed Model	-0.0032	0.0183	0.010	0.02312

RMSE and MAE on DKASCASA-1B. The cross-attention module has the greatest impact on performance, highlighting that the interplay of local temporal features and global context is essential for accurate prediction. The prediction error also raises when either the residual temporal blocks or the global-context branch are removed, further highlighting the crucial role of both short-term dynamics and sequence-level information. Likewise, when dynamic feature weighting is replaced by fixed ones, the accuracy drops, indicating that dynamic feature weighting has definite benefits. Overall, the ablation results demonstrate that all four components are complementary and jointly improve the forecasting capability of the proposed model.

D. Visual Analysis of the Proposed Model

Figure 2 compares the actual and predicted PV active-power profiles under rainy, cloudy, and sunny conditions across the three sites. The proposed model closely follows the daily generation patterns, including the onset, peak, and decline of PV power. The agreement is strongest under sunny conditions, where the profiles are smoother and more stable. Under cloudy and rainy conditions, rapid irradiance changes cause abrupt fluctuations, making prediction more challenging. Nevertheless, the model captures the main temporal trends and responds well to short-term variations. Larger deviations mainly occur around sharp peaks and sudden power drops,

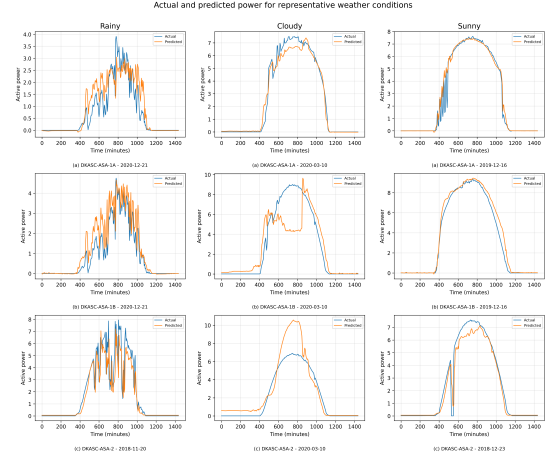


Fig. 2. Actual and predicted PV active-power profiles across three sites under rainy, cloudy, and sunny conditions.

where the signal is highly volatile. Overall, the visual results demonstrate the model's consistent forecasting performance across different weather conditions and heterogeneous PV sites.

E. Limitations and Future Work

Although the proposed framework provides an executable and reproducible PV forecasting pipeline, several extensions

TABLE II

ABLATION ANALYSIS OF THE PROPOSED MODEL ON THE DKASCASA-1B ACTIVE-POWER TARGET. LOWER RMSE AND MAE VALUES INDICATE BETTER FORECASTING ACCURACY, WHILE MBE REFLECTS PREDICTION BIAS.

Model Variant	MBE	RMSE	MAE
Without residual temporal blocks	0.0238	0.046820	0.029840
Without global-context branch	0.0226	0.044910	0.027920
Without cross-attention fusion	0.0241	0.048350	0.031170
Without adaptive aggregation	0.0219	0.043760	0.026810
Full proposed model	0.0245	0.039667	0.024305

can further strengthen the study. First, all baseline and state-of-the-art methods should be implemented under the same preprocessing and evaluation protocol to provide a complete comparative analysis. Second, inverse-transformed physical-unit metrics should be reported in addition to scaled errors, as physical-unit errors are more interpretable for grid operators. Third, future work may incorporate site-aware embeddings, time-of-day encodings, seasonal indicators, and weather-regime classification to further improve multi-site generalization. Finally, uncertainty estimation and probabilistic forecasting can be added to quantify confidence intervals for operational energy management.

VI. CONCLUSION

This paper proposed an adaptive cross-attention similarity learning framework for direct multi-step photovoltaic power forecasting using large-scale real-world measurements from three heterogeneous PV sites. The framework integrates residual temporal convolution, global-context modeling, cross-attention fusion, and adaptive representation aggregation to jointly capture short-term power fluctuations, broader operating patterns, and interactions among complementary latent features. By predicting the complete 12-step horizon simultaneously, the proposed approach avoids the error accumulation associated with recursive forecasting. Experimental results demonstrated competitive and generally improved forecasting performance across the DKASCASA datasets. The proposed model achieved RMSE and MAE values of 0.03967 and 0.02431 kW on DKASCASA-1B, 0.0593 and 0.0366 kW on DKASCASA-1A, and 0.0183 and 0.0100 kW on DKASCASA-2, respectively. In particular, the model substantially improved the results on DKASCASA-2 and reduced the MBE magnitude from 0.0854 to 0.0032 kW compared with DSDN. The visual analysis further confirmed that the model effectively follows PV generation profiles under sunny, cloudy, and rainy conditions, although abrupt irradiance-driven fluctuations remain more challenging. Overall, the findings demonstrate the effectiveness of combining local temporal encoding, global contextual information, and adaptive attention-based fusion for robust multi-site PV forecasting. Future work will investigate probabilistic forecasting, uncertainty quantification, site-aware representations, seasonal and time-of-day encoding, and lightweight deployment for real-time smart-grid applications.

REFERENCES

- [1] A. M. Bukar, A. Almerbati, S. Shuja, and S. M. Zubair, "Enhancing solar pv panel performance through active and passive cooling techniques: A comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 216, p. 115611, 2025.
- [2] T. Khan and C. Choi, "Attention enhanced dual stream network with advanced feature selection for power forecasting," *Applied Energy*, vol. 377, p. 124564, 2025.
- [3] I. J. Esfahani, S. Lee, and C. Yoo, "Extended-power pinch analysis (epopa) for integration of renewable energy systems with battery/hydrogen storages," *Renewable energy*, vol. 80, pp. 1–14, 2015.
- [4] F. Zhou, Z. Wang, T. Zhong, G. Trajcevski, and A. Khokhar, "Hydroflow: towards probabilistic electricity demand prediction using variational autoregressive models and normalizing flows," *International Journal of Intelligent Systems*, vol. 37, no. 10, pp. 6833–6856, 2022.
- [5] X. Yao, Z. Wang, and H. Zhang, "A novel photovoltaic power forecasting model based on echo state network," *Neurocomputing*, vol. 325, pp. 182–189, 2019.
- [6] K. Wang, X. Qi, and H. Liu, "Photovoltaic power forecasting based lstm-convolutional network," *Energy*, vol. 189, p. 116225, 2019.
- [7] J. Bao, P. Liu, and S. V. Ukkusuri, "A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data," *Accident Analysis & Prevention*, vol. 122, pp. 239–254, 2019.
- [8] C. Wen, S. Liu, X. Yao, L. Peng, X. Li, Y. Hu, and T. Chi, "A novel spatiotemporal convolutional long short-term neural network for air pollution prediction," *Science of the total environment*, vol. 654, pp. 1091–1099, 2019.
- [9] B.-M. Hodge, C. B. Martinez-Anido, Q. Wang, E. Chartan, A. Florita, and J. Kiviluoma, "The combined value of wind and solar power forecasting improvements and electricity storage," *Applied Energy*, vol. 214, pp. 1–15, 2018.
- [10] M. Khodayar, O. Kaynak, and M. E. Khodayar, "Rough deep neural architecture for short-term wind speed forecasting," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 2770–2779, 2017.
- [11] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—a novel pooling deep rnn," *IEEE transactions on smart grid*, vol. 9, no. 5, pp. 5271–5280, 2017.
- [12] Z. Masood, R. Gantassi, and Y. Choi, "A multi-step time-series clustering-based seq2seq lstm learning for a single household electricity load forecasting," *Energies*, vol. 15, no. 7, p. 2623, 2022.
- [13] A. K. Dubey, A. Kumar, V. García-Díaz, A. K. Sharma, and K. Kanhaiya, "Study and analysis of sarima and lstm in forecasting time series data," *Sustainable Energy Technologies and Assessments*, vol. 47, p. 101474, 2021.
- [14] Z. A. Khan, T. Hussain, I. U. Haq, F. U. M. Ullah, and S. W. Baik, "Towards efficient and effective renewable energy prediction via deep learning," *Energy Reports*, vol. 8, pp. 10230–10243, 2022.
- [15] W. Ullah, T. Hussain, Z. A. Khan, U. Haroon, and S. W. Baik, "Intelligent dual stream cnn and echo state network for anomaly detection," *Knowledge-Based Systems*, vol. 253, p. 109456, 2022.
- [16] E. Sarma, E. Spiliotis, E. Stamatopoulos, V. Marinakis, and H. Doukas, "Short-term photovoltaic power forecasting using meta-learning and numerical weather prediction independent long short-term memory models," *Renewable Energy*, vol. 216, p. 118997, 2023.
- [17] X. Chen, W. Chen, V. Dinavahi, Y. Liu, and J. Feng, "Short-term load forecasting and associated weather variables prediction using resnet-lstm based deep learning," *IEEE Access*, vol. 11, pp. 5393–5405, 2023.
- [18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2015.
- [19] A. Hussain, Z. A. Khan, T. Hussain, F. U. M. Ullah, S. Rho, and S. W. Baik, "A hybrid deep learning-based network for photovoltaic power forecasting," *Complexity*, vol. 2022, no. 1, p. 7040601, 2022.
- [20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.
- [21] Z. A. Khan, T. Hussain, and S. W. Baik, "Dual stream network with attention mechanism for photovoltaic power forecasting," *Applied Energy*, vol. 338, p. 120916, 2023.
- [22] P. Li, K. Zhou, X. Lu, and S. Yang, "A hybrid deep learning model for short-term pv power forecasting," *Applied Energy*, vol. 259, p. 114216, 2020.