VG-BiLSTM: Modeling Feature Variability for Robust PV Power Prediction

Are Sambasiva Rao
Department of CSE, GITAM School of Technology, GITAM Deemed to be University, Hyderabad Campus, Telangana, India
sare2@gitam.in

K. Dhana Sree Devi
Department of CSE, GITAM School of Technology, GITAM Deemed to be University, Hyderabad Campus, Telangana, India
drdhanasree2@gmail.com (corresponding author)

Received: 15 December 2025 | Revised: 1 January 2026, 19 January 2026, and 4 February 2026 | Accepted: 7 February 2026

 DOI: https://doi.org/10.48084/etasr.16956

ABSTRACT

Photovoltaic (PV) power generation oscillates across multiple timescales due to changing weather and environmental conditions. These fluctuations have a huge impact on the stability and operational costs of the grid and reduce its effective use. Motivated by the need for reliable grid integration, this study emphasizes the importance of explicitly modeling variability in PV features to improve the robustness of solar power prediction models. Conventional approaches often neglect short-term fluctuations in key meteorological features, which can lead to degraded model performance and accuracy. This study presents the Variability Gated Bidirectional Long Short-Term Memory (VG-BiLSTM) framework, which explicitly models both features and variability using descriptors to predict short-term PV power output. The proposed method computes feature descriptors, the rolling standard deviation, ramp rate, and coefficient of variation for each feature, and employs a variability attention module to dynamically weight them. A gating network combines raw and variability representations into a single latent embedding, which is then input into a Bidirectional LSTM for temporal modeling. Experimental results show that VG-BiLSTM outperforms baseline RNN, GRU, LSTM, and their bidirectional versions, achieving much lower RMSE and nRMSE in different sky types. The model also achieved the best error levels across regions, with RMSE and NRMSE reduced to 21.6 W and 20.4%, respectively, outperforming other models.

Keywords-PV power; envirnomental features; variability descriptors; BiLSTM; variability attention

I. INTRODUCTION

The global push to switch to renewable energy sources has made solar energy, especially PV technology, the most popular sustainable energy option [1]. Solar energy is abundant and more environmentally friendly, making it a critical component of modern energy scenarios. However, PV power generation is inherently discontinuous, as it depends more on meteorological features such as irradiance, temperature, and humidity [2]. Since these features fluctuate over short and long time intervals under different atmospheric conditions, huge uncertainty is introduced into system functionalities. These fluctuations cause huge variations in power production, introducing challenges for maintaining grid stability, because sudden changes in PV generation can disturb the balance between demand and supply.

Accurate short-term PV power forecasting under varying atmospheric conditions has become an active research area [3]. Many PV forecasting models can capture short-term dependencies with limited ability to handle nonlinear and sudden fluctuations [4-6]. The advent of Deep Learning (DL) has offered many sophisticated algorithms to capture nonlinearities and spatiotemporal dependencies in PV forecasting [7, 8]. Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) have presented a stronger ability to learn temporal dependencies in time series [9-11]. More recently, attention mechanisms have been integrated into DL frameworks to dynamically focus on relevant input features, offering improved accuracy and interpretability [12, 13]. Despite these advances, most studies focus exclusively on raw (actual) meteorological features. Many forecasting approaches ignore the role of feature variability (fluctuations), making DL models perform well in clear sky conditions but degrade significantly under changing weather. Thus, a model that can adaptively weigh both actual input features and variability patterns of significant features can improve the PV forecasts during high variability periods.

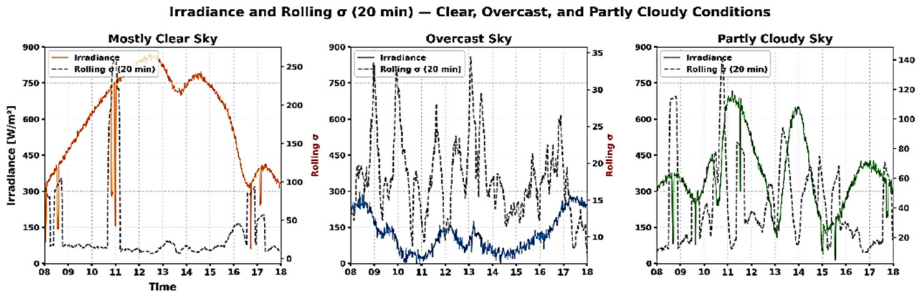

Fig. 1. Understanding the temporal stability of solar irradiance under varying sky conditions, with rolling σ.

In [14, 15], models were trained only on actual feature values, exhibiting low error in the training-validation data but performing poorly in test data with volatile weather conditions. Similarly, existing forecasting methods ignore variability, lack interpretability, and use limited attention. Motivated by the challenges arising from feature variability, this work proposes a novel Variability-Gated Bidirectional LSTM (VG-BiLSTM) framework to effectively handle dynamic variations in PV power data. The main objectives were to incorporate variability descriptors, such as standard deviation, ramp rate, and coefficient of variation, into the forecasting process to: (i) design a variability-aware attention and gating mechanism that adaptively balances raw features with their variability representations, and (ii) employ a BiLSTM-based temporal backbone capable of capturing both forward and backward dependencies in PV power dynamics.

II. METHODOLOGY

Accurate PV forecasting requires understanding and modeling both the actual values of features such as irradiance, temperature, and humidity, and their short-term variability. This short-term variability drives sudden ramps in output. As an example, Figure 1 presents the instantaneous irradiance (solid colored lines) and its corresponding rolling standard deviation (σ) (dashed grey line). This analysis is computed with a 20-minute window for three sky regimes: mostly clear, overcast, and partly cloudy days. In Figure 1(a), under mostly clear sky conditions, the irradiance is a smooth curve that peaks around the solar noon at a peak of about 800 W/m^2. On the other side, the rolling σ values are constant, close to zero throughout the day. These measurements prove the fact that the irradiance is not only high but also stable in variability. This high predictability zone can establish temporal dependencies that are well structured and can be efficiently represented by recurrent models such as LSTMs.

The overcast sky plot in Figure 1(b) reveals a significantly lowered mean irradiance between 200–300 W/m^2 due to continuous cloud cover and diffused radiation. The rolling σ remains small below 50 W/m^2, although the actual irradiance values are low. This analysis indicates that the system is in a low-energy, though steady, state. Figure 1(c) displays the third condition, partly cloudy, having a highly varying irradiance pattern that is characterized by sudden and frequent fluctuations throughout the day. Irradiance peaks greater than 700 W/m^2 are followed by very strict declines over brief durations. The rolling σ exhibits and presents peaks more than 130 W/m², depicting high-intense variability. These very high σ areas particularly denote time durations where the sky state changes fast. These are areas where forecasting uncertainty increases drastically. In such sky types, DL forecasting models trained only on raw actual irradiance values may fail to capture and understand the variability nature, leading to degraded performance and unstable forecasts by the model.

The proposed forecasting framework uses variability descriptors [16] like rolling standard deviation, ramp rate, and coefficient of variation. Variability descriptors are calculated on a sliding time window for every feature. Raw feature values capture instantaneous system states, whereas variability descriptors encode temporal fluctuations and uncertainty that are not directly observable from raw measurements. Fusing both enables the model to jointly exploit magnitude and variability-driven information, resulting in a more robust and context-aware representation.

A. Feature Variability Descriptors

Let $\mathcal{F} = \{f_1, f_2, \ldots . f_F\}$ represent the key meteorological features. Let the raw input feature vector at site s and time t be $x_{s,t}$:

$$x_{s,t} = \left[x_{s,t}^{f_1}, x_{s,t}^{f_2}, \ldots \ldots, x_{s,t}^{f_F}\right] \in \mathbb{R}^F \quad (1)$$

where $\sigma_{s,t}^f$ represents the rolling standard deviation of feature f, measuring the dispersion of recent observations within the sliding window size of W. This is a direct measure of short-term volatility in the feature. $r_{s,t}^f$ denotes the ramp rate [17, 18] of the feature f, representing the absolute difference between series of time steps. $CV_{s,t}^f$ denotes the coefficient of variation [19, 20], representing the normalized variability by scaling the standard deviation relative to the local mean. Together, these descriptors capture both the magnitude and intensity of fluctuations, offering a richer representation of meteorological dynamics than raw values alone. These descriptors are calculated as:

$$\sigma_{s,t}^f = Std(x_{s,t-W+1:t}^f) \quad (2)$$

$$r_{s,t}^f = |x_{s,t}^f - x_{s,t-1}^f| \quad (3)$$

$$CV_{s,t}^f = \frac{\sigma_{s,t}^f}{\mu_{s,t}^f + \varepsilon};\ \mu_{s,t}^f = Mean(x_{s,t-W+1:t}^f) \quad (4)$$

All variability descriptors are normalized to zero mean and unit variance before gating to ensure scale compatibility with raw feature embeddings. ε is a small positive quantity. These

descriptors are concatenated to form a variability vector $v_{s,t}^{f}$, given in (5), which compactly summarizes the short-term fluctuations. $v_{s,t}$ denotes variability vectors across all features (6). This type of integration enables the forecasting model to gain a richer representation of feature dynamics, allowing it to distinguish between stable and highly volatile conditions that impact PV power generation significantly.

$$v_{s,t}^{f} = \left[\sigma_{s,t}^{f}, r_{s,t}^{f}, CV_{s,t}^{f}\right] \in \mathbb{R}^{3} \quad (5)$$

$$v_{s,t} = \left[v_{s,t}^{f_1} || v_{s,t}^{f_2} || \ \ldots || \ v_{s,t}^{f_F}\right] \in \mathbb{R}^{3F} \quad (6)$$

A direct concatenation of actual features and their variability descriptors may bias the learning process, as they differ significantly, both in scale and semantics. To address this, a Multilayer Perceptron (MLP) $\phi(.)$ is used, projecting them into a common latent embedding space $h_{s,t}$ (7). These embeddings are not recurrent hidden states but feature-level latent representations that serve as inputs to the subsequent attention and gating mechanisms.

$$h_{s,t}^{raw} = \phi_r\,(x_{s,t}), \ \ h_{s,t}^{var}(f) = \phi_v\,(v_{s,t}^{f}) \quad (7)$$

where ϕ_r , ϕ_v are MLPs. In addition to providing a common single representation appropriate for downstream attention mechanisms, this step guarantees numerical stability and aligns disparate feature types.

B. Variability Attention with Gating Mechanism

In PV power forecasting, the influence of features such as irradiance, temperature, and humidity is highly weather-dependent. Irradiance predominates under clear skies, but its short-term oscillations become crucial during cloudy periods. On the other hand, temperature and humidity variability are more important under stable conditions. Models that treat all inputs equally run the risk of misrepresenting key influencers and lowering predictive performance. To address this, an attention mechanism can be included to dynamically adjust the contribution of features and their variability at each time step. The attention mechanism can assign higher weights to the most relevant features under changing weather conditions. For example, irradiance ramp rates are considered more important during cloud motion. In stable periods, temperature and humidity variability weigh more. This adaptive selection can improve the forecasting accuracy of the model.

In order to tap the impact of variability in features, this paper presents a two-fold mechanism that includes an attention layer and a gating module. First, the attention scoring function $e_{s,t}^{f}$ is used to evaluate the variability embeddings of all features. The scoring function learns the relative key importance of each feature's variations at every time step. To ensure a probabilistic distribution, the attention scores are normalized using the softmax function $\alpha_{s,t}^{f}$. These normalized weights are then used to construct a variability context vector $C_{s,t}^{f}$ that acts as a dynamic summary of the most influential variations in that time frame.

$$e_{s,t}^{f} = q^{T} \tanh\left(W_v \ h_{s,t}^{var}(f) + \ b_v\right) \quad (8)$$

$$\alpha_{s,t}^{f} = \frac{exp\left(e_{s,t}^{f}\right)}{\sum_{k=1}^{F} e_{s,t}^{k}} \quad (9)$$

$$C_{s,t}^{f} = \sum_{f=1}^{F} \alpha_{s,t}^{f} \,.\, h_{s,t}^{var}(f) \quad (10)$$

Second, the raw feature embedding and the variability context vector are integrated through a gating mechanism. The gate $g_{s,t} \in [0,1]$ uses sigmoid activation to adaptively balance the contributions of raw features and their variability representations.

$$g_{s,t} = \sigma\left(W_g \ \left[h_{s,t}^{raw} || C_{s,t}^{f} \ + b_g\right]\right) \quad (11)$$

$$\tilde{h}_{s,t} = g_{s,t} \odot h_{s,t}^{raw} + (1 - g_{s,t}) \odot C_{s,t}^{f} \quad (12)$$

Here, the gate decides how much the model should rely on raw values versus variability information at each time step. When $g_{s,t}$ approaches 1, the fused representation relies completely on raw feature values, which is desirable under stable operating conditions. Conversely, when $g_{s,t}$ approaches 0, the model emphasizes variability descriptors, allowing it to better capture rapid fluctuations. The suggested gate introduces a variability-aware, feature-level adaptive fusion, allowing the model to selectively switch between magnitude-driven and variability-driven representations, whereas the traditional attention or fusion strategies treat all features equally. The resulting fused representation $\tilde{h}_{s,t}$ is both adaptive and interpretable. The term "adaptive" refers to the model's ability to dynamically adjust the relative contribution of raw feature values and variability information at each time step through the learnable gating function $g_{s,t}$. The term "variability-aware" denotes the explicit modeling of temporal fluctuations using attention-weighted variability descriptors $C_{s,t}^{f}$ that capture changes in feature behavior beyond instantaneous magnitudes.

This fusion mechanism is applied at a single well-defined stage of the architecture, as shown in Figure 2, serving as an interface between feature representation and prediction. The fused representation is then passed to the BiLSTM for temporal dependency modeling. As shown in the architecture, input to the BiLSTM is the fused sequence of embeddings ($\tilde{H}_{s,1:T}$), with forward ($\vec{h}_t$) and backward ($\overleftarrow{h}_t$) LSTM hidden states, and output from the final hidden state (h_t). The BiLSTM leverages temporal variability by modeling forward and backward dependencies in the variability-enhanced representations ($\tilde{H}_{s,1:T}$), allowing it to capture both the onset and persistence of rapid PV power fluctuations more effectively than unidirectional temporal models. By combining attention-driven selection and gated fusion, this design enables the network to answer not only which feature matters most but also how much variability versus raw information should be retained for accurate PV power prediction.

$$\tilde{H}_{s,1:T} = \left[\tilde{h}_{s,1}, \tilde{h}_{s,2}, \ldots, \tilde{h}_{s,T}\right] \quad (13)$$

$$\vec{h}_t = LSTM_{fwd}\left(\tilde{h}_{s,t} \,.\, \vec{h}_{t-1}\right),$$

$$\overleftarrow{h}_t = LSTM_{bwd}\left(\tilde{h}_{s,t} \,.\, \overleftarrow{h}_{t+1}\right) \quad (14)$$

$$h_t = \left[\vec{h}_t \,||\, \overleftarrow{h}_t\right] \quad (15)$$

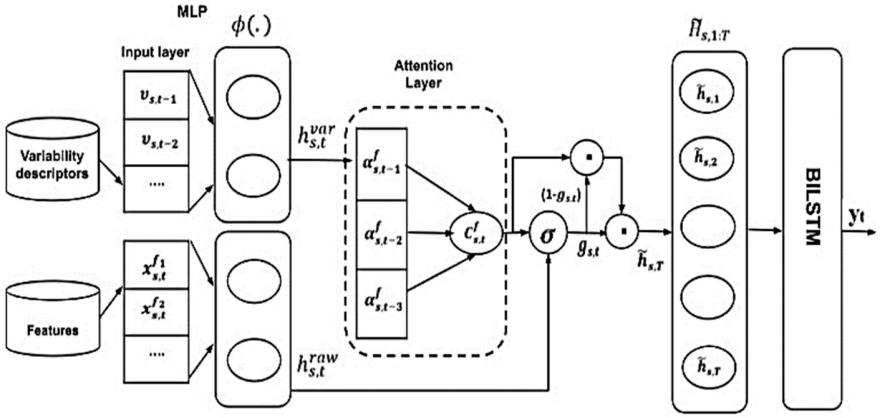

Fig. 2. Architecture of the proposed VG-BiLSTM model.

While standard gating techniques frequently rely on static or globally learned fusion weights, conventional attention mechanisms typically assign weights to features or time steps based only on latent representations. On the other hand, a feature-wise and time-dependent modification of their respective contributions is made possible by the suggested gating mechanism, which functions on both raw feature embeddings and attention-weighted variability descriptors. Standard attention or gating baselines do not explicitly enable the model's ability to adaptively flip between magnitude-driven and variability-driven signals.

Algorithm 1: Variability attention gating

Input: $x_{s,t} = \left[x_{s,t}^{f_1}, x_{s,t}^{f_2}, \ldots\ldots, x_{s,t}^{f_F}\right] \in \mathbb{R}^F$

1. Find variability descriptors: $\sigma_{s,t}^{f}, r_{s,t}^{f}, CV_{s,t}^{f}$
2. Construct the variability vector $(v_{s,t}^{f})$ (6)
3. Use MLP to construct the latent embedding space h (7)
4. Calculate the normalized attention scores $\alpha_{s,t}^{f}$ for $h_{s,t}^{var}(f)$
5. Construct a variability context vector $(C_{s,t}^{f})$

III. EXPERIMENTS AND DISCUSSION

A. Datasets

The proposed approach was validated using a real-time PV dataset [21, 22] that includes meteorological features such as irradiance, temperature, humidity, and additional auxiliary features such as wind speed and cloud cover. Each feature is sampled at 5-minute intervals, and PV output is measured in kW. The dataset was divided into training (70%), validation (15%), and testing (15%) subsets. A rolling window of size $W = 20$ was applied for computing variability descriptors.

B. Evaluation Metrics and Benchmark Models

To assess the prediction performance of the proposed model, several evaluation metrics were selected, including Root Mean Squared Error (RMSE), normalized Root Mean Squared Error (nRMSE), Mean Absolute Error (MAE), and normalized Mean Absolute Error (nMAE). Benchmark models RNN, BiRNN, GRU, BiGRU, LSTM, and BiLSTM were used to compare the prediction performance of the proposed model. All benchmark models used two sequential layers and one fully connected layer, and hyperparameters were common to all models: 64 neurons, 0.2 dropout, *tanh* activation, Adam optimizer, 0.001 learning rate, and 150 epochs.

The proposed VG-BiLSTM model was implemented using TensorFlow and Keras. The variability descriptor module was constructed with a rolling window size of $W = 20$, and both raw feature and variability embeddings were mapped into a latent space of 64 neurons with ReLU activation. The model employed variability attention to dynamically weight feature fluctuations, followed by a gating mechanism with a dropout rate of 0.2 for robust fusion. A two-layer BiLSTM with 64 hidden units per layer and *tanh* activation was used to capture temporal dependencies, and a fully connected MLP produced the final output. The BiLSTM backbone was chosen to capture both past and future temporal dependencies within short-term PV sequences, providing richer contextual modeling than unidirectional LSTM while maintaining lower complexity and greater stability. Training was performed with a batch size of 512, a learning rate of 0.001 with the Adam optimizer, for 150 epochs. The gating mechanism uses an input dimension of $2d$ and produces an output of dimension d, employing a sigmoid activation function with a dropout of 0.2. The gated embedding dimension d was selected to balance representational capacity and computational efficiency, while remaining independent of the number of input features to ensure scalability. A fixed d is used across all regions and sky conditions, demonstrating robust generalization of the proposed fusion mechanism.

To understand the variation in actual and predicted data, experiments were conducted on data from three regions of India, North, Central, and East, on three sky regimes: Mostly clear, Overcast, and Partly Cloudy. Figure 3 shows 1-day-ahead power BiLSTM predictions. The figures show a huge variability in actual and predicted values in almost all three regimes. Figure 4 illustrates the prediction performance of the baseline models compared to the proposed model across the three regimes in the Central region (for brevity, prediction graphs for only the Central region are provided).

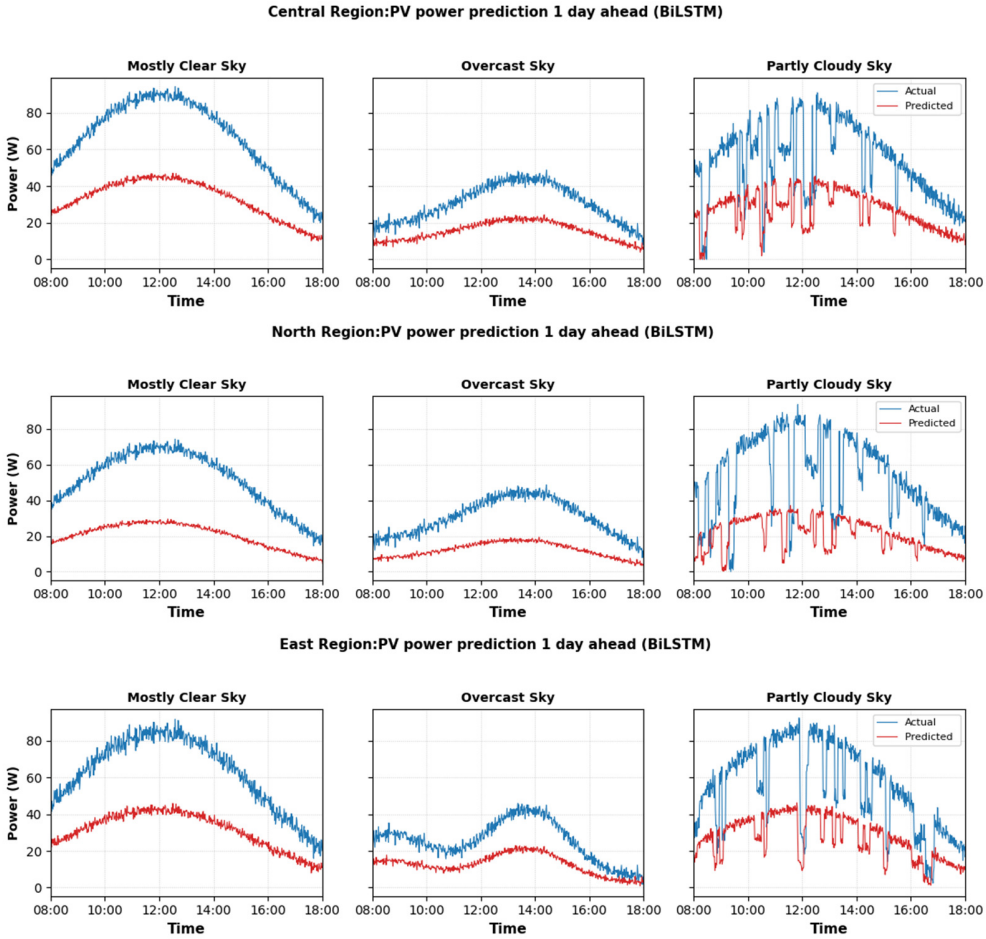

Fig. 3. 1-Day ahead BiLSTM predictions across three regimes.

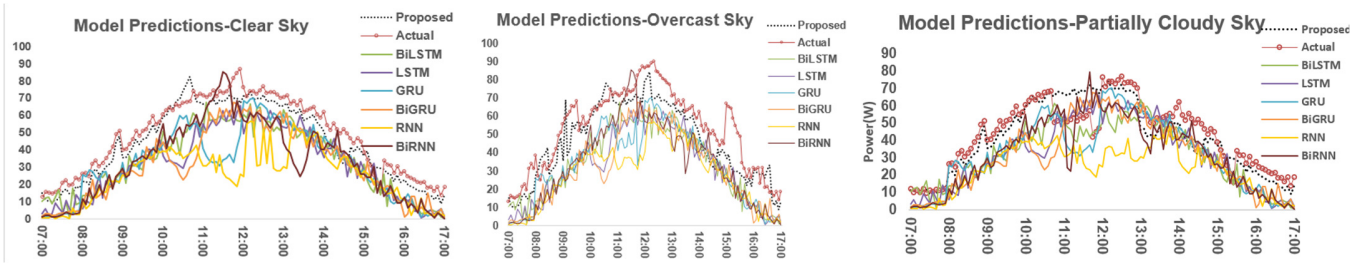

Fig. 4. Comparison of the proposed VG-BiLSTM and benchmark model predictions.

The predictions of the proposed VG-BiLSTM were very close to the actual values, while the predictions of the baseline models deviated more. Table I summarizes the MSE, RMSE, and MAE values of the models discussed across all three regions. The proposed VG-BiLSTM model consistently achieved the lowest RMSE and MAE across all regions, outperforming all baseline recurrent models. In the Central region, it reduced RMSE from 31.8 W (RNN) to 21.6 W and nMAE from 22.3% to 9.5%, achieving substantial forecasting gains. Similar improvements were observed in the North and East region, where the proposed model delivered the best normalized errors. These results confirm that incorporating variability information significantly enhances PV power forecasting accuracy across diverse geographical conditions.

In addition, the ablation study across the Central region demonstrates the significance of the variability-aware component $C_{s,t}^{f}$. With $C_{s,t}^{f}$ enabled, the proposed model achieved RMSE, nRMSE, MAE, and nMAE of 21.6, 20.4, 10.9, and 9.5 in the Central region. In contrast, removing $C_{s,t}^{f}$ resulted in consistently higher errors of 26.7, 23.4, 14.8, and 12.6 in the same region, clearly confirming that variability-aware attention substantially enhances predictive accuracy and robustness across regions.

TABLE I. RESULTS OF DIFFERENT MODELS AT 100 EPOCHS

Central	**RNN**	**BiRNN**	**GRU**	**BiGRU**	**LSTM**	**BiLSTM**	**Proposed**
RMSE	31.8	29.1	24.9	26.2	30.8	24.3	21.6
nRMSE	27.2	28.6	22.4	23.6	27.7	21.9	20.4
MAE	23.9	18.8	12.6	14.8	17.3	14.9	10.9
nMAE	22.3	17.3	11.1	13.2	15.3	13.4	9.5
North	**RNN**	**BiRNN**	**GRU**	**BiGRU**	**LSTM**	**BiLSTM**	**Proposed**
RMSE	30.9	28.7	28.4	29.3	32.8	27.6	22.8
nRMSE	28.3	26.3	26.0	26.9	30.1	25.0	21.3
MAE	24.0	17.2	14.4	15.6	18.9	15.4	12.5
nMAE	22.2	14.6	12.8	13.8	16.7	14.7	11.0
East	**RNN**	**BiRNN**	**GRU**	**BiGRU**	**LSTM**	**BiLSTM**	**Proposed**
RMSE	29.7	28.1	27.5	28.9	31.6	27.0	24.8
nRMSE	25.9	24.6	24.1	25.3	27.6	23.9	21.9
MAE	14.1	13.1	12.9	13.8	15.9	12.7	11.3
nMAE	11.8	10.9	10.7	11.5	13.2	10.5	9.4

IV. CONCLUSION

This work presented a novel variability-aware DL framework for short-term PV power forecasting. The main contributions of this study include the incorporation of variability descriptors to capture short-term PV fluctuations, the design of a variability-aware attention for robust feature fusion, and the use of a BiLSTM backbone to model bidirectional temporal dependencies for improved forecast accuracy. The proposed approach uniquely integrates raw meteorological features and their temporal variability descriptors within a unified and adaptive learning mechanism. The variability gate dynamically modulates the attention assigned to input sequences based on the degree of short-term fluctuations, enabling the network to distinguish stable clear-sky periods from highly transient cloudy intervals. Overall, the key novelty lies in the explicit modeling of feature variability combined with adaptive gated fusion, which provides improved robustness and interpretability compared to existing forecasting approaches. Experimental analysis demonstrated that the proposed VG-BiLSTM with RMSE reduced to 21.6 W and nRMSE to 20.4%, outperforming other models. The ablation study across the Central region showed that, with $C_{s,t}^{f}$enabled, the proposed model achieved RMSE, nRMSE, MAE, and nMAE of 21.6, 20.4, 10.9, and 9.5, respectively.

REFERENCES

[1] A. O. M. Maka and J. M. Alabid, "Solar energy technology and its roles in sustainable development," *Clean Energy*, vol. 6, no. 3, pp. 476–483, June 2022, https://doi.org/10.1093/ce/zkac023.

[2] K. Hasan, S. B. Yousuf, M. S. H. K. Tushar, B. K. Das, P. Das, and Md. S. Islam, "Effects of different environmental and operational factors on the PV performance: A comprehensive review," *Energy Science & Engineering*, vol. 10, no. 2, pp. 656–675, Feb. 2022, https://doi.org/10.1002/ese3.1043.

[3] Y. Guo, Q. Han, T. Li, H. Fu, M. Liang, and S. Zhang, "Robust Photovoltaic Power Forecasting Model Under Complex Meteorological Conditions," *Mathematics*, vol. 13, no. 11, May 2025, Art. no. 1783, https://doi.org/10.3390/math13111783.

[4] H. B. Wolff, "Support Vector Regression for Solar Power Prediction," Ph.D. dissertation, University of Oldenburg, 2017.

[5] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, Jan. 2003, https://doi.org/10.1016/S0925-2312(01)00702-0.

[6] T. A. Nguyen *et al.*, "A Study on Novel Solar Power Forecasting using an XGB-LiGBM-RF Hybrid Model and the L-BFGS-B Optimization Algorithm," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 24516–24522, Aug. 2025, https://doi.org/10.48084/etasr.11308.

[7] J. Yu *et al.*, "Deep Learning Models for PV Power Forecasting: Review," *Energies*, vol. 17, no. 16, Aug. 2024, Art. no. 3973, https://doi.org/10.3390/en17163973.

[8] M. Saffari and M. Khodayar, "Spatiotemporal Deep Learning for Power System Applications: A Survey," *IEEE Access*, vol. 12, pp. 93623–93657, 2024, https://doi.org/10.1109/ACCESS.2024.3424854.

[9] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling." arXiv, Dec. 11, 2014, https://doi.org/10.48550/arXiv.1412.3555.

[10] Y. Li, F. Ye, Z. Liu, Z. Wang, and Y. Mao, "A Short-Term Photovoltaic Power Generation Forecast Method Based on LSTM," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–11, Jan. 2021, https://doi.org/10.1155/2021/6613123.

[11] A. Zameer, F. Jaffar, F. Shahid, M. Muneeb, R. Khan, and R. Nasir, "Short-term solar energy forecasting: Integrated computational intelligence of LSTMs and GRU," *PLOS ONE*, vol. 18, no. 10, Oct. 2023, Art. no. e0285410, https://doi.org/10.1371/journal.pone.0285410.

[12] X. Lei, "A Photovoltaic Prediction Model with Integrated Attention Mechanism," *Mathematics*, vol. 12, no. 13, July 2024, https://doi.org/10.3390/math12132103.

[13] K. Yang, Y. Cai, and J. Cheng, "A deep learning model based on multi-attention mechanism and gated recurrent unit network for photovoltaic power forecasting," *Computers and Electrical Engineering*, vol. 123, Apr. 2025, Art. no. 110250, https://doi.org/10.1016/j.compeleceng.2025.110250.

[14] A. Alcañiz, D. Grzebyk, H. Ziar, and O. Isabella, "Trends and gaps in photovoltaic power forecasting with machine learning," *Energy Reports*, vol. 9, pp. 447–471, Dec. 2023, https://doi.org/10.1016/j.egyr.2022.11.208.

[15] M. Subramaniam, D. Hanafi, and A. Joret, "Neural Network Based Prediction of Solar Power Generation," *Evolution in Electrical and Electronic Engineering*, vol. 6, no. 2, pp. 356–365, Oct. 2025.

[16] T. E. Hoff and R. Perez, "Quantifying PV power Output Variability," *Solar Energy*, vol. 84, no. 10, pp. 1782–1793, Oct. 2010, https://doi.org/10.1016/j.solener.2010.07.003.

[17] B. Nouri *et al.*, "Ramp Rate Metric Suitable for Solar Forecasting," *Solar RRL*, vol. 8, no. 24, Dec. 2024, Art. no. 2400468, https://doi.org/10.1002/solr.202400468.

[18] J. Yan, T. Qu, S. Han, Y. Liu, X. Lei, and H. Wang, "Reviews on characteristic of renewables: Evaluating the variability and complementarity," *International Transactions on Electrical Energy Systems*, vol. 30, no. 7, July 2020, https://doi.org/10.1002/2050-7038.12281.

[19] T. Kealy, "The missing parameter in renewable energy power quality analysis, i.e., the coefficient of variation: Case study of a 3-MW on-site wind turbine project in Ireland," *Journal of Cleaner Production*, vol. 280, Jan. 2021, Art. no. 124699, https://doi.org/10.1016/j.jclepro.2020.124699.

[20] C. Wu, X. P. Zhang, and M. Sterling, "Solar power generation intermittency and aggregation," *Scientific Reports*, vol. 12, no. 1, Jan. 2022, Art. no. 1363, https://doi.org/10.1038/s41598-022-05247-2.

[21] "Solar energy power generation dataset." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/stucom/solar-energy-power-generation-dataset.

[22] D. K. D. S. Devi, "dhanasreek/Solar-Energy-Datasets." July 04, 2025, [Online]. Available: https://github.com/dhanasreek/Solar-Energy-Datasets.