PHOTOVOLTAIC POWER PREDICTION BASED ON IMPROVED PYRAFORMER

Guodong LI [1], Wenhao CAI [2]*

The stochastic and high volatility of PV power generation poses a great challenge to the operational management of the electrical grid, while precise anticipation of PV power yield can reduce the impact of its uncertainty. Hence, a photovoltaic prediction model called PDGformer, based on an improved Pyraformer model, is proposed. The encoder of the PDGformer model employs a unique dual-branch structure, wherein the local branch captures the local information of the photovoltaic power sequence, enabling the simultaneous capture of temporal and dimensional dependencies of the photovoltaic data. Conversely, the global branch captures the global information of the photovoltaic power sequence. The model's decoder incorporates an Attention mechanism to effectively integrate both local and global information and generate the final prediction results. Additionally, an MSE reweighting framework is introduced to alleviate the interference caused by abrupt changes in predictions. This framework reduces the loss caused by mutations while increasing the loss for normal states. Employing actual photovoltaic data from a specific location for illustrative analysis, the experiments demonstrate the superior performance of the proposed model compared to others, such as Pyraformer, in effectively predicting photovoltaic output power.

Keywords: PV power prediction; Pyraformer; PDGformer; MSE reweighting framework.

Glossary

NWP	Numerical Weather Prediction
RNN	Recurrent Neural Networks
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Networks
LRA	The Long-Range Arena
PAM	Pyramidal Attention Module
CSCM	Coarser-Scale Construction Module
CVM	Cross-Variable Module
Gconv	Global Convolution Kernel Module
Re-MSE	The Reweighted Mean Squared Error
LD	Local Difference

[1] Associate Prof., School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China, e-mail: lgd@ncepu.edu.cn
[2] Master's Degree Student., School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China, e-mail: 1780147977@qq.com

1. Introduction

Solar energy has assumed an increasingly pivotal role in global power systems in recent years [1]. Among the diverse applications of solar energy, converting it into electricity through photovoltaic installations stands out as the most widely recognized method for harnessing solar power [2]. Not only does it offer the world a source of clean energy and diminish reliance on fossil fuels for societal advancement, but it also boasts a remarkably economical operation and maintenance cost, resulting in substantial economic benefits [3]. Despite these benefits, PV power generation faces challenges due to its stochastic nature and high volatility [4]. The increasing share of installed capacity amplifies uncertainty, leading to scheduling and operational challenges. Accurate prediction of PV power output is crucial for mitigating this impact, holding significance in grid scheduling, power generation technology advancement, and economic optimization of power plants [5].

Presently, PV power prediction methods are categorized into three groups: physical [6], statistical [7], and artificial intelligence [8] techniques. Physical methods use PV system design and NWP for forecasting without historical data. Statistical methods extract features from input data for future outcome prediction. Artificial intelligence methods, including machine learning and deep learning, have gained prominence. Machine learning techniques like neural networks [9], random forests [10], and support vector machines [11] outperform traditional statistics, but recent research focuses on advanced deep learning due to overfitting and generalization challenges [12]. The literature [13] proposes a CNN-LSTM model that effectively uses weather variables to predict photovoltaic power plant output with superior accuracy compared to various ML and DL models.

In 2017, Google introduced the Transformer model, leveraging the attention mechanism [14]. Unlike RNNs, Transformers excel at capturing long-range dependencies, making them prevalent in time series forecasting, including photovoltaic power forecasting [15]. A novel Transformer-based model is proposed for one-hour-ahead photovoltaic power prediction [16]. Despite its effectiveness, the Transformer faces challenges of quadratic time complexity and memory usage, leading to the emergence of variants like Informer with lower complexity [17]. Pyraformer is another approach, a low-complexity pyramidal attention model for time series forecasting [18]. It primarily focuses on temporal dependencies but overlooks inter-variable connections. Incorporating related dimensions can enhance predictions in a specific dimension. Literature [19] underscores solar radiation, temperature, and other factors' significant impact on photovoltaic energy generation, aiding effective output prediction.

The Long-Range Arena (LRA) benchmark [20] systematically evaluates sequential models across extensive contexts, spanning from 1K to 16K tokens.

Notably, Transformer-based models have found limited success in LRA, leading to the emergence of global convolutional networks. Particularly, the S4 model [21] draws from state-space models, akin to a global convolutional kernel. SGconv [22] re-evaluates global convolution and underscores its importance in modeling extended sequences. It emphasizes two design principles for global convolutional kernels: a sublinear (logarithmic) relationship between learnable parameters and input length, and a weight decay structure for the global kernel.

Considering the pros and cons of the mentioned prediction methods, we introduce PDGformer, an innovative photovoltaic power prediction model. It's built upon an enhanced Pyraformer design, featuring a dual-branch encoder structure: local and global branches. The local branch captures local photovoltaic generation information. It employs the Cross-Variable Module (CVM) to learn variable dependencies and Pyraformer for temporal dependencies, effectively encompassing both dimensions. The global branch uses the Global Convolution Module (Gconv) to capture the global information of photovoltaic generation data. The decoder integrates both local and global information for prediction. Prior to input, we normalize photovoltaic generation using Dish-TS [23] and denormalize after predictions. Additionally, we propose an MSE reweighting framework to reduce mutation-induced loss while enhancing normal state loss.

2. Pyraformer-based photovoltaic power forecasting

The Pyraformer model is a neural network that utilizes a multi-resolution pyramid attention mechanism, which effectively captures dependencies in time series forecasting. It achieves linear time and space complexity, making it an efficient solution. Fig. 1 illustrates the model structure of Pyraformer.

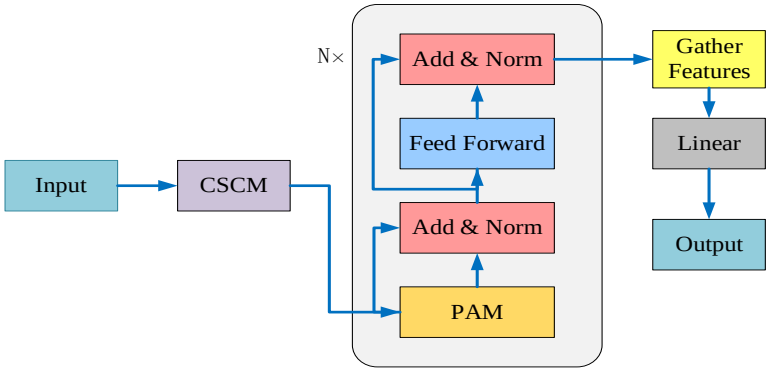


Fig. 1. Structure diagram of Pyraformer

2.1 Pyramid Attention Module (PAM)

PAM serves as the fundamental module within the Pyraformer model. The module employs pyramid diagrams to depict temporal interdependencies within historical sequences with a multi-resolution approach, classifying connections into two types: connections within the same scale, connections between different scales. Connections between different scales lead to a C-fork tree summarizing diverse resolution features, while intra-scale adjacency connections capture diverse ranges of temporal dependencies. In the pyramid graph structure, lower nodes represent time points in the original sequence, and upper-layer nodes extract features to effectively represent characteristics at lower resolutions. By establishing connections between nodes within each layer, meaningful relationships among the nodes can be established.

2.2 Coarse-Scale Construction Module (CSCM)

The CSCM module's primary objective resides in the initialization of coarse-scale nodes within the pyramid graph, enabling efficient information exchange between these nodes in the subsequent PAM module. Specifically, in the temporal dimension, a series of convolutional layers is sequentially applied to the embedded sequence. These convolutional layers have a C-sized kernel and employ a C-sized step. This sequential convolutional operation results in coarse-grained sequences with a scale of s and a length of L/Cs. Before feeding these sequences into the stacked convolutional layers, the dimension of each node undergoes reduction via a fully connected layer, which is subsequently restored after the completion of all convolutional operations. Prior to entering the PAM module, these sequences, spanning from fine to coarse granularity, are interconnected. This meticulously crafted architecture significantly diminishes the parameter count within the module, thereby mitigating the risk of overfitting.

3. PDGformer-based PV power prediction

When predicting photovoltaic (PV) power, apart from historical power generation data, certain meteorological variables like historical wind speed, temperature, and others also contribute to the prediction. The Pyraformer model effectively captures temporal dependencies, yet it falls short in fully utilizing the interrelationships among the variables, thereby limiting its predictive capacity. Additionally, recent research [22] has highlighted the remarkable capability of global convolution kernels in capturing long-range dependencies.

To address the aforementioned limitations of the Pyraformer model and improve the capture of long-range dependencies in PV power generation data, this paper proposes the PDGformer model. This model adeptly captures both inter-temporal and cross-dimensional dependencies while effectively incorporating both

global and local information, maximizing their synergistic effects. Moreover, this paper introduces the Re-MSE, a mean squared error (MSE) reweighting framework, and employs it as the loss function during model training.

The model's configuration, as expounded upon herein, is illustrated in Fig. 2. The encoder adopts a unique two-branch design, with each branch dedicated to capturing and extracting specific types of information: local and global information, respectively. Subsequently, the model decoder integrates these two types of information to generate accurate prediction results. Since the generalized neural paradigm Dish-TS proves to be effective in removing and recovering non-stationary information from time series [23] we add Dish-TS layers before and after the proposed model.

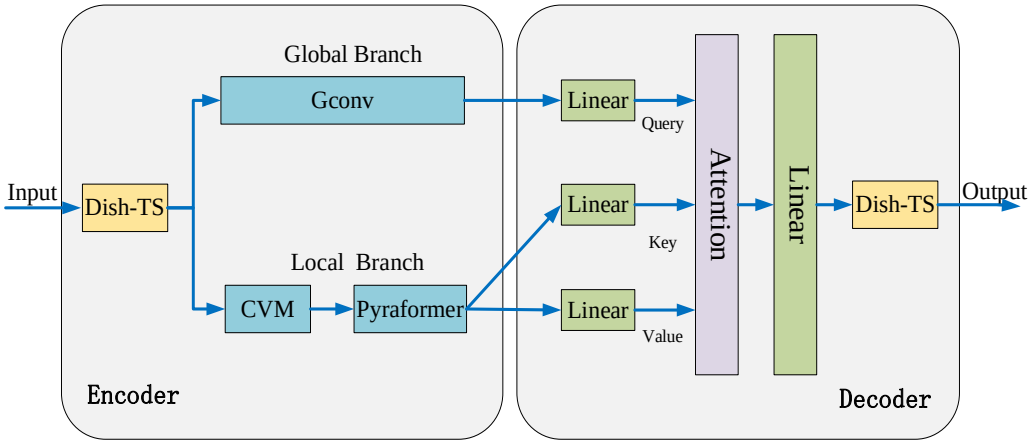


Fig. 2. PDGformer Model

3.1 Encoder

The model's encoder has two parallel branches. The global branch focuses on extracting global information Z_{global}, representing long-term dependencies in the PV generation sequence. This branch utilizes a global convolution kernel (Gconv) to process the entire input sequence $X \in R^{N \times d}$. Notably, the complexity of the global convolution kernel is sublinear (logarithmic) in relation to the length of the sequence.

Conversely, the local branch is dedicated to capturing the immediate proximate information Z_{local}, i.e., the dependencies between neighboring time steps. The local branch consists of two modules: CVM and Pyraformer. While CVM is responsible for capturing cross-dimensional dependencies, Pyraformer handles cross-time dependencies. To reduce overall complexity without compromising prediction accuracy, only the tail of the input sequence $X_{tail} \in R^{N' \times d} (N' < N)$ is fed into the local branch. Thus, the encoder can be succinctly represented as follows:

$$Z_{global} = Branch_{global}(X), Z_{local} = Branch_{local}(X_{\text{tail}}) \quad (1)$$

3.2 Decoder

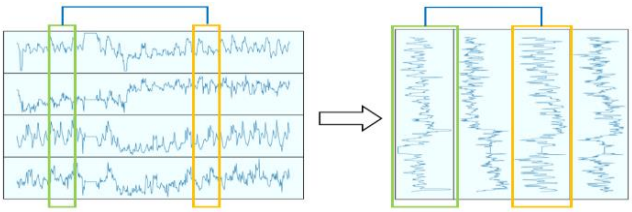
In order to augment the exploitation of both global and local information, the decoder assumes the responsibility of integrating the global information (Z_{global}) and the local information (Z_{local}), and it outputs the prediction results. The decoder module primarily consists of the cross-attention module, designed to ensure an efficient representation of the historical information in the photovoltaic generation sequence. Initially, the encoder linearly maps the global and local information, subsequently employing the global information as a query vector (q), while the local information serves as both key vector (k) and value vector (v). These vectors are then inputted into the Attention layer, thereby achieving the following representation:

$$q = MLP(Z_{global}), k = MLP(Z_{local}), v = MLP(Z_{\text{local}}) \quad (2)$$

$$Attention(q,k,v) = softmax\left(\frac{qk^T}{\sqrt{d_q}}\right)v \quad (3)$$

3.3 Cross-Variable Module (CVM)

Cross-time attention typically involves embedding all data points across different dimensions for a specific time period into a feature vector, with a primary focus on capturing the interdependencies existing across distinct time periods. While this approach effectively captures cross-time dependencies, it falls short in fully leveraging cross-dimensional dependencies, potentially constraining its predictive capabilities. As a solution to this limitation, the CVM module introduces cross-variable attention, which facilitates the learning of dependencies between variables, as illustrated in Fig. 3.



a. Cross time attention b. Cross dimensional attention

Fig. 3. Cross-Variable Attention

The CVM module comprises two crucial components: a multi-head attention (MHA) mechanism and a feed-forward network (FFN). The input sequence X, representing the PV data, is a two-dimensional tensor characterized by the shape L × C, where C signifies the number of variables. Prior to processing, the

sequence X is required to be flipped. The cross-variable attention serves as a pivotal aspect of the MHA and is precisely defined as follows:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{C}}\right)V \tag{4}$$

In the equation, Q, K, and V are the query, key, and value matrices, respectively, while C denotes the quantity of variables. Q, K, and V are typically obtained through a linear transformation of the original input X. Due to the detrimental effects of additional embedding layers on temporal information and subsequent performance degradation, the input sequence is directly fed into the MHA module without any embeddings. Furthermore, as there is no temporal ordering among different variables, there is no need for positional encoding of the input sequence.

3.4 Global Convolution Kernel (Gconv)

Previous research has demonstrated the ability of global convolutional kernels to effectively capture and model long-range dependencies. Specifically, global convolutional kernels employ elongated filters that extend across the entire input sequence, thereby enabling the capture of prolonged interdependencies. Given the input sequence $u \in R^{n\times d}$, the learnable global kernel $k \in R^{n\times d}$, and the output $y \in R^{n\times d}$, the operation of global convolution can be accomplished through the utilization of the rapid Fourier transform, indicated by F, resulting in a computational intricacy of O(NlogN). The specific details are expounded below:

$$y = u * k = F^{-1}\left(F(u)\cdot F(k)\right) \tag{5}$$

Compared to local convolutions with fixed kernel sizes, global convolution requires kernels of the same size as the input sequence length, denoted as L. When dealing with long sequences, parametrizing the convolutional kernels in a straightforward manner, as done in local convolutions, becomes challenging. Therefore, it is crucial to have an efficient kernel parameterization method. SGConv [22] addresses this issue by constructing the global convolution kernel through a composition of sub-kernels, each with increasing sizes. The size of each subsequent sub-kernel is twice that of the previous sub-kernel. Importantly, all sub-kernels are upsampled from the same number of parameters, establishing a logarithmic relationship between the number of parameters and the input length. Additionally, a weighted combination of sub-kernels with weight decay is utilized, assigning smaller weights to larger sub-kernels. The global convolution kernel is defined as follows:

$$k = \frac{1}{Z}\left[k_0, k_1, \cdots, k_{N-1}\right], k_i = \alpha^i Upsample_{2^{\max[i-1,0]}d}\left(w_i\right) \tag{6}$$

Within the equation, $w_i \in R^d$ represents the parameters for the i-th subkernel k_i, while $N = \lceil \log_2 (L/d) \rceil + 1$ represents the number of scales. To create sub-kernels of different scales, an upsampling operation using linear interpolation is employed. This operation, denoted as $Upsample_{2^{\max[i-1,0]}d}(w_i)$, involves upsampling w_i to a length of $2^{\max[i-1,0]}d$. Z is a normalization constant that ensures the convolution operation does not alter the scale of the input, while α is a decay coefficient that governs the rate of decay.

3.5 Re-MSE

In PV power prediction, unexpected or unknown events (e.g., PV sensor failures) may lead to drastic changes in PV power data. Despite their infrequent occurrence in the training set, the losses caused by these mutations can significantly impact the overall loss, thereby limiting the generalization performance of prediction models during the testing phase.

To mitigate the influence of mutations, this paper introduces a reweighting framework that reduces the weight of losses caused by mutations while increasing the weight of losses caused by normal states. Subsequently, the training is conducted using the reweighted Re-MSE loss function. Given a photovoltaic power dataset, it can be partitioned, resulting in sets for training, validation, and testing. Let $D = \{(X_t, Y_t)\}_{t=1}^{N}$ indicates the training set, with X_t signifying the input sequence and Y_t representing the output sequence. This framework primarily addresses the issue of loss imbalance, which arises from significant differences between adjacent input-output pairs (X_a and Y_a) compared to other input-output pairs (X_t and Y_t). Here, a represents the timestamp associated with a mutation.

To address this issue, the concept of Local Differences (LD) is introduced to quantify the discrepancy between two adjacent input-output sequences, X_t and Y_t. It is denoted as:

$$LD(X_t, Y_t) = \frac{\bar{X}_t - \bar{Y}_t}{\sqrt{\frac{S_{\bar{X}_t}^2}{I} + \frac{S_{\bar{Y}_t}^2}{O} + \varepsilon}} := v_t \tag{7}$$

Within the equation, $\bar{X}_t$ signifies the input series mean, while $S_{\bar{X}_t}$ denotes its standard deviation, I signifies the input series length, and O signifies the output series length.

Drawing inspiration from the work introduced in [24] regarding deep imbalance regression, the next step involves computing the LD density, which serves as an indicator of the frequency of temporal changes. This is achieved by utilizing kernel density estimation with the LD values obtained from the training samples. The calculation process for determining the LD density is as follows:

$$\tilde{p}(v') \coloneqq \int_V k(v,v')p(v)dv \tag{8}$$

In the equation provided, the term p(v) represents v's frequency of occurrence within the training data. The function k (v, v') represents a symmetric kernel function.

After estimating the LD density $\tilde{p}(v)$, weights w_t are assigned to each training sample using the formula $w_t = c \cdot \tilde{p}(v_t) \propto \tilde{p}(v_t)$, where c is a constant that acts as a scaling factor. The primary aim of this reweighting procedure is to mitigate the impact of mutations while also attending to the challenge of imbalanced loss. The reweighted mean squared error (Re-MSE) loss function can be defined as follows, incorporating the assigned weights:

$$Re\text{-}MSE_w\left(Y_t, \hat{Y}_t\right) = \frac{1}{O} \cdot w_t \sum_{i=0}^{O-1}\left(Y - \hat{Y}_t\right)^2 \tag{9}$$

In this expression, $\hat{Y}_t$ represents the predicted output sequence obtained by using X_t as input, while Y_t represents the corresponding ground truth values. O signifies the length of the output sequence, denoting the predicted extent.

4. Case study

4.1 Experimental dataset and pre-processing

The text utilizes a photovoltaic energy generation dataset obtained from the Yulara Solar System 1 site, located at the DKASC in Australia. The dataset encompasses the timeframe spanning from January 1, 2017, to December 31, 2021, with data recorded at an hourly resolution.

For the accuracy and effectiveness of the predictive model, the raw dataset undergoes preprocessing. Occasionally, data loss may occur due to maintenance issues or equipment failures at the solar site. In such cases, missing values are filled using linear interpolation. Additionally, any negative values in the generated power are replaced with zero for consistency. For model training, evaluation, and testing, the dataset is then partitioned into subsets: 70% for training, 20% for validation, and 10% for testing.

4.2 Experimental parameterization and evaluation metrics

The meticulous choice of suitable model parameters greatly impacts the model's predictive efficacy. In this experiment, the local branch of the PDGformer model maintains a constant input length of 96, while the global branch adopts an enlarged window (336) to encompass more information. Prediction lengths are set at 24, 48, 72, 96, and 192, respectively. A batch size of 32 is specified, with the epoch set at 10, and a learning rate established at 0.0001. Throughout the training process, the loss function introduced in this text, namely RE-MSE, is employed.

Within the PDGformer model's encoder, the local branch comprises two modules: CVM and Pyraformer. These modules are respectively dedicated to capturing inter-dimensional dependencies and temporal interdependencies. CVM is configured with 2 layers, while Pyraformer's pyramid attention consists of 4 layers and 4 attention heads. In CSCM, the convolutional kernel size C is designated as 4, with a stride of 4. PAM entails that A, the number of adjacent nodes attended to by nodes within the same scale, is set to 3. The global branch Gconv within the PDGformer model's encoder is established as a single layer.

With the intent to evaluate the predictive precision of the model, the study utilizes four different evaluation metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-Squared (R^2).

4.3 Experiment and Analysis

To assess the effectiveness of the suggested model in predicting solar energy generation, this manuscript compares it with benchmark models, including LSTM, Transformer, Informer, and Pyraformer. All models are used to forecast with five different prediction horizons: 24, 48, 72, 96, and 192. The results obtained from these models are presented in Table 1 for comparison and analysis.

Based on the analysis of Table 1, it is evident that the PDGformer model surpasses various benchmark models in relation to different assessment criteria across distinct prediction lengths (24, 48, 72, 96, and 192). (1) Compared to the traditional Pyraformer model, PDGformer exhibits higher prediction accuracy, with reductions in MSE of 13.5% (24), 10.9% (48), 21.2% (72), 22.8% (96), and 32.4% (192). Additionally, MAE is reduced by 10.9% (24), 11.6% (48), 11.5% (72), 17.4% (96), and 20.6% (192). These results underscore the effectiveness of PDGformer's dual-branch design, Cross-Variable Module (CVM), and MSE reweighting framework (Re-MSE). (2) Notably, as the prediction length extends, there is a typical decline in prediction accuracy across all models. However, the PDGformer model shows a relatively slower decline, maintaining high R^2 values even as the length extends from 96 to 192. This observation suggests that the dual-branch structure allows the PDGformer model to concurrently capture global and local information, maximizing their complementary strengths and maintaining competitiveness in long-term forecasting tasks. (3) Moreover, compared to Informer, Transformer, and LSTM models, PDGformer achieves an average MSE reduction of 29.6% (24), 27.7% (48), 31.9% (72), 38.0% (96), and 45.5% (192), indicating superior predictive performance across various prediction lengths.

Fig. 4 illustrates the prediction curves of all models for a prediction length of 192. It is evident that the PDGformer model excels in capturing and reconstructing the intricate details of the fluctuations in the photovoltaic power generation. While transformer-based models (Pyraformer, Informer, and Transformer) accurately capture overall fluctuation patterns, they fall short in

predicting certain crucial elements. In contrast, PDGformer demonstrates exceptional performance, accurately reconstructing subtle variations and turning points, outperforming other models in capturing finer details of the fluctuations.

Table 1

Evaluation of the accuracy of photovoltaic power prediction

Prediction Length (h)	Metrics	PDGformer	Pyraformer	Informer	Transformer	LSTM
24	MSE	0.134	0.155	0.178	0.175	0.226
	MAE	0.212	0.238	0.248	0.249	0.313
	RMSE	0.366	0.393	0.422	0.418	0.475
	R^2	0.860	0.838	0.814	0.817	0.749
48	MSE	0.156	0.175	0.196	0.193	0.276
	MAE	0.222	0.251	0.259	0.244	0.335
	RMSE	0.395	0.419	0.443	0.439	0.525
	R^2	0.838	0.815	0.793	0.797	0.709
72	MSE	0.164	0.208	0.219	0.216	0.307
	MAE	0.239	0.270	0.273	0.275	0.344
	RMSE	0.406	0.456	0.468	0.464	0.554
	R^2	0.826	0.781	0.768	0.772	0.683
96	MSE	0.173	0.224	0.259	0.257	0.334
	MAE	0.242	0.293	0.301	0.294	0.367
	RMSE	0.416	0.474	0.509	0.507	0.578
	R^2	0.817	0.762	0.725	0.727	0.646
192	MSE	0.186	0.275	0.309	0.315	0.419
	MAE	0.258	0.325	0.334	0.329	0.401
	RMSE	0.432	0.524	0.556	0.561	0.647
	R^2	0.800	0.706	0.670	0.664	0.532

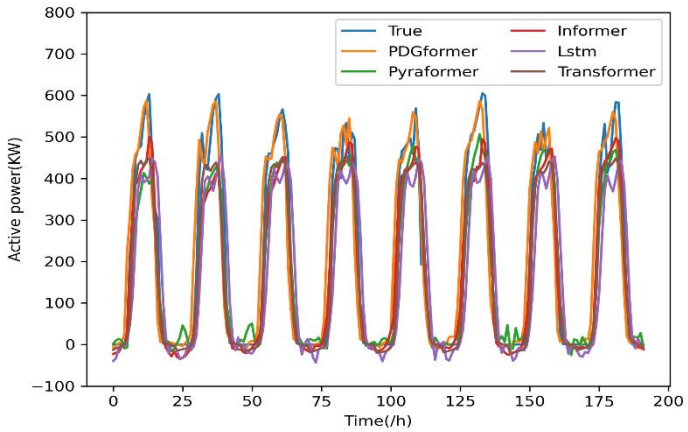


Fig. 4. All models PV power prediction curves

Furthermore, the methods proposed in the literature [25] to [28] were selected for a comprehensive comparative analysis, underscoring the heightened

performance of the proposed method. Table 2 presents distinct prediction outcomes for each method when forecasting at a length of 96.

Table 2

Comparison of different prediction methods in the published literature

Prediction method	MSE	MAE	RMSE	R^2
Ref. 25	0.221	0.278	0.470	0.765
Ref. 26	0.206	0.274	0.454	0.781
Ref. 27	0.194	0.269	0.441	0.795
Ref. 28	0.191	0.256	0.437	0.799
Proposed	0.173	0.242	0.416	0.817

The superior predictive performance of the proposed PDGformer model is evident when compared with CNN-Informer in reference [25], VMD-CNN-TransNN in reference [26], CNN-LSTM-Transformer in reference [27], and RevIN-DLinear in reference [28], as illustrated in Table 2.

4.4 Ablation Experiment

In this section, a ablation study is conducted with a prediction length of 96 to validate the efficacy of various enhancements within the PDGformer model. Throughout the experimental process, utilizing PDGformer as the foundation, models are derived by removing specific components: PDGformer-C (removal of CVM module), PDGformer-G (removal of global convolution kernel), and PDGformer-R (absence of Re-MSE). A comparative analysis of performance is conducted between PDGformer and PDGformer-C, PDGformer-G, and PDGformer-R, with the empirical findings tabulated in Table 3.

Table 3

Outcomes of predicting performance

	PDGformer	PDGformer-C	PDGformer-G	PDGformer-R
MSE	0.173	0.182	0.192	0.189
MAE	0.242	0.258	0.267	0.260
RMSE	0.416	0.427	0.438	0.435
R^2	0.817	0.805	0.796	0.801

From the provided table, it is evident that removing the global convolutional module leads to an 11.0% increase in MSE, emphasizing its significant impact on result prediction accuracy. When using ordinary MSE instead of Re-MSE, the MSE increases by 9.2%, suggesting that the reweighting framework proposed by Re-MSE can effectively balance the loss caused by mutations, reducing model loss and improving prediction accuracy. Moreover, removing the CVM module results in a 5.2% increase in MSE, showcasing its role in capturing cross-dimensional dependencies and improving prediction accuracy.

Statistical analysis reveals the PDGformer model, with its performance enhancement techniques, outperforms others. The global convolutional module proves most effective in enhancing accuracy, followed by the Re-MSE framework and the CVM module.

5. Conclusions

This paper presents the PDGformer model, an improved version of the Pyraformer model, to address photovoltaic power prediction. The PDGformer model's encoder captures both global and local information of the photovoltaic power sequence, while the decoder integrates these two types of information to maximize their complementarity. During model training, a novel MSE loss function called Re-MSE is introduced. Using actual PV data as an example, the PDGformer model is compared with other competitive models for photovoltaic power prediction, yielding the following main conclusions:

(1) The Pyraformer model focuses on capturing temporal dependencies but overlooks inter-variable dependencies. The introduction of the cross-variable module (CVM) in the PDGformer model effectively captures the relationships between variables, leveraging the correlated information from other dimensions to enhance the precision of photovoltaic power prediction.

(2) The Global convolutional module (Gconv) in the PDGformer model is pivotal in capturing global dependencies within the input sequence. By utilizing attention mechanisms, the Gconv module facilitates the integration of global and local information, leading to a significant improvement in prediction accuracy.

(3) To balance the loss caused by mutations, this paper introduces Re-MSE, a novel mean squared error (MSE) loss function. Re-MSE reduces the weight of losses caused by mutations while increasing the weight of losses caused by normal states, thereby further enhancing the model's prediction accuracy.

R E F E R E N C E S

[1]. *Akhter, et al*. "An hour-ahead PV power forecasting method based on an RNN-LSTM model for three different PV plants". Energies, **vol. 15**, no. 6, Mar. 2022, pp. 2243-2264.

[2]. *MA. Dobrea, M. Vasluianu, G. Neculoiu, and N. Arghira*. "Modeling and simulation of a 3 kW photovoltaic system for an autonomous consumer". U.P.B. Sci. Bull., Series C, **vol**. 83, no. 2, 2021.

[3]. *Y. K. Wu, C. L. Huang*, "Completed review of various solar power forecasting techniques considering different viewpoints". Energies, **vol. 15**, no. 9, May. 2022, pp. 3320-3342.

[4]. *P. Gupta, and R. Singh*, "PV power forecasting based on data-driven models: a review". International Journal of Sustainable Engineering, **vol. 14**, no. 6, Nov. 2021, pp. 1733-1755.

[5]. *J. Zhang, et al*, "Baseline and target values for regional and point PV power forecasts: Toward improved solar forecasting". Solar Energy, **vol. 122**, Dec. 2015, pp. 804-819.

[6]. *D. Yang*, "History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining", Solar Energy, **vol. 168**, Jul. 2018, pp. 60-101.

[7]. *M. Q. Raza, M. Nadarajah, and C. Ekanayake*, "On recent advances in PV output power forecast", Solar Energy, **vol. 136**, Oct. 2016, pp. 125-144.
[8]. *A. Mellit, A. Massi Pavan, and E. Ogliari*, "Advanced methods for photovoltaic output power forecasting: A review", Applied Sciences, **vol. 10**, no. 2, Jan. 2020, pp. 487-509.
[9]. *X. Xue*, "Prediction of daily diffuse solar radiation using artificial neural networks". International Journal of Hydrogen Energy, **vol. 42**, no. 47, Nov. 2017, pp. 28214-28221.
[10]. *D. Liu, and K. Sun*, "Random forest solar power forecast based on classification optimization", Energy, **vol. 187,** Nov. 2019.
[11]. *H. S. Jang*, "Solar power prediction based on satellite images and support vector machine", in IEEE Transactions on Sustainable Energy, **vol. 7**, no. 3, Mar. 2016, pp. 1255-1263.
[12]. *M. Elsaraiti, and A. Merabet*,"Solar power forecasting using deep learning techniques", IEEE Access, **vol. 10**, Mar. 2022, pp. 31692-31698.
[13]. *A. Agga, A. Abbou, M. Labbadi, Y. El Houm, and I.H.O. Ali*, "CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production", Electric Power Systems Research, **vol. 208**, Jul. 2022.
[14]. *A. Vaswani, N. Shazeer, and N. Parmar*, "Attention is all you need". Advances in neural information processing systems, **vol. 30**, 2017.
[15]. *X. Huang, et al*, "Time series forecasting for hourly photovoltaic power using conditional generative adversarial network and Bi-LSTM". Energy, **vol. 246**, May. 2022.
[16]. *Q. T. Phan, Y. K. Wu, and Q. D. Phan*, "An approach using transformer-based model for short-term PV generation forecasting", in 2022 8th International Conference on Applied System Innovation (ICASI), IEEE, 2022, pp. 17-20.
[17]. *H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang*, "Informer: Beyond efficient transformer for long sequence time-series forecasting", in Proceedings of the AAAI conference on artificial intelligence, **vol. 35**, No. 12, May. 2021, pp. 11106-11115.
[18]. *S. Liu, H. Yu, C. Liao, J. Li, W. Lin, A. X. Liu, and S. Dustdar*, "Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting", in International conference on learning representations, Oct. 2021.
[19]. *MG. Osman, D. Ciupageanu, and A. Stan*. "Analysis of Solar Radiation in Sudan and Optimal Location of Photovoltaic Panels". U.P.B. Sci. Bull., Series C, vol. 84, no. 4, 2022.
[20]. *Y. Tay, M. Dehghani, and S. Abnar*, "Long range arena: A benchmark for efficient transformers", arXiv preprint arXiv, 2011.04006, 2020.
[21]. *A. Gu, K. Goel, and C. Ré*, "Efficiently modeling long sequences with structured state spaces", arXiv preprint arXiv, 2111.00396, 2021.
[22]. *Y. Li, T. Cai, Y. Zhang, D. Chen, and D. Dey*, "What Makes Convolutional Models Great on Long Sequence Modeling?", arXiv preprint arXiv, 2210.09298, 2022.
[23]. *W. Fan, P. Wang, D. Wang, Y. Zhou, and Y. Fu*, "Dish-TS: A General Paradigm for Alleviating Distribution Shift in Time Series Forecasting", in Proceedings of the AAAI Conference on Artificial Intelligence, **vol. 37**, No. 6, jun. 2023, pp. 7522-7529.
[24]. *Y. Yang, K. Zha, and Y. Chen*, "Delving into deep imbalanced regression", International Conference on Machine Learning, PMLR, Jul. 2021, pp. 11842-11851.
[25]. *Z. Wu, F. Pan, D. Li, and H. He*, "Prediction of photovoltaic power by the Informer model based on convolutional neural network." Sustainability. vol. 14, No. 20, 2022.
[26]. *TN. Trong, HVX. Son, H. Do. Dinh*, et al. "Short-term PV power forecast using hybrid deep learning model and Variational Mode Decomposition." Energy Reports. Vol 9, 2023, pp: 712-717.
[27]. *EM. Al-Ali, Y. Hajji and Y. Said*. "Solar Energy Production Forecasting Based on a Hybrid CNN-LSTM-Transformer Model." Mathematics, vol. 11, no. 3, 2023.
[28]. *G. Wang, Y. Liao, L. Guo*. "DLinear photovoltaic power generation forecasting based on reversible instance normalization." in 2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS). IEEE, 2023.