International Journal of Power Electronics and Drive System (IJPEDS)
Vol. 17, No. 3, September 2026, pp. 2127~2148
ISSN: 2088-8694, DOI: 10.11591/ijpeds.v17.i3.pp2127-2148

TCM-Former: a transformer with temporal convolution for photovoltaic power forecasting

Sarab Al-Chlaihawi[1], Mohammed A. T. Alrubei[2], Faris A. Alhaddad[3]
[1]Technical Engineering College of Najaf, Al-Furat Al-Awsat Technical University, Al-Najaf, Iraq
[2]Communication Department of Najaf Technical Institute, Al-Furat Al-Awsat Technical University, Al-Najaf, Iraq
[3]Electrical Department of Najaf Technical Institute, Al-Furat Al-Awsat Technical University, Al-Najaf, Iraq

Article Info

Article history:

Received Dec 19, 2025
Revised Aug 6, 2026
Accepted Aug 19, 2026

Keywords:

Photovoltaic power forecasting
Seasonal–trend decomposition via LOESS
Solar energy prediction
STL decomposition
Time series forecasting

ABSTRACT

To tackle the problem of modeling long-term trends and short-term and high-frequency variations in PV time series, a strong and efficient forecasting model of photovoltaic (PV) power generation is advanced. This paper presents STL-TCM-Former, a hybrid model that breaks down the raw PV signal with seasonal-trend decomposition with LOESS (STL) into trend and seasonal components. These elements are after that processed via a dual path encoder decoder transformer architecture that is improved with a temporal convolutional module (TCM). The period (transformer-based) path is used for capturing the global, long-range dependencies in seasonal component and temporal (TCM) path is used to extract the localized, short-term dynamics. Trend component is processed with a dedicated TCM branch, minimizing component interferences and providing a multiscale temporal representation, specific to PV generation patterns. The proposed model was evaluated on the Yulara (Uluru) solar dataset under short-term (60-hour) and long-term (300-hour) forecasting horizons. Compared with benchmark models including LSTM, WOA-LSTM, VMD-LSTM, WOA-VMD-LSTM, and hybrid WOA/VMD/LSTM configurations, STL-TCM-former achieved superior performance with $R^2 = 99.76\%$, MAPE $\approx 2.24\%$, RMSE = 16.63, and MAE = 10.36. In addition, the PJM interconnection dataset was also employed to evaluate the generalization capability of the proposed method. The results demonstrate high accuracy, stability, and strong generalization capability under varying environmental conditions.

Corresponding Author:

Sarab Al-Chlaihawi
Technical Engineering College of Najaf, Al-Furat Al-Awsat Technical University
31001 Al-Najaf, Iraq
Email: inj.srb@atu.edu.iq

1. INTRODUCTION

As the world undergoes a rapid transition towards the use of renewable energy sources, photovoltaic (PV) power generation has become an essential element of renewable energy portfolios [1]. The growing nature of solar energy into modern power grids, however, poses a great challenge as it entails inherent intermittency and varying reliance on other factors like solar irradiance, temperature, and cloud cover [2]. Stable, efficient, and economical functioning of power systems therefore requires accurate and reliable PV power forecasting. Precise PV power prediction is highly practical, which can directly support major operational processes, including grid scheduling, energy trading, and optimization of storage. Higher quality forecasts will allow operators to manage the balance between supply and demand more efficiently, make better decisions in the market, and manage storage assets more efficiently. Therefore, they require high accuracy with regard to predicting the system performance and enabling a large-scale integration of PV [3], [4]. Despite the large body of research on this topic, PV power forecasting is a complex problem with stochasticity, nonlinearity and close temporal correlations. Conventional statistical models, like autoregressive integrated moving average (ARIMA) and exponential smoothing models are not always sufficient to describe nonlinear dynamics and multi-scale variability on solar power data [5], [6]. Machine learning (ML) systems, like random forests (RF), support vector machines (SVM) and the artificial neural network (ANN) have increased the accuracy of prediction by learning nonlinear trends, but these systems continue to fail in capturing the long-term relations and in temporal changes [7]. Gated recurrent units (GRU) and long short-term memory (LSTM) are recurrent neural network (RNN) and deep learning models, which were majorly utilized for predicting PV power [8]. Such architectures are very good at capturing time series sequential dependencies; nevertheless, they have a number of limitations. LSTM-based models are prone to the problem of vanishing gradients, high computational costs, and a reduced capacity to effectively learn long term temporal patterns. Besides, their effectiveness reduces with complicated periodicities or abrupt alterations as a result of meteorological conditions [9]. Transformer-based architecture has become a strong competitor to time series modeling over the past few years because of their self-attention mechanism (AM) that allows them to effectively model long-range correlations and complicated temporal interactions [10].

Though the recently suggested LSTM- and transformer-based PV forecasting models have already shown the promising performance, a number of research gaps exist. Specifically, the vast majority of current methods cannot effectively and simultaneously capture both short-term local dynamics and long-term temporal dependencies on unified platform. The standard transformer models do not usually provide sufficient representation of localized temporal variations that are very important in solar power generation resulting in lowered accuracy in highly dynamic settings [11], [12]. Consequently, these challenges can be resolved by creating a model that can collectively learn local temporal- and long-range dependencies in one architecture, and thereby enhance PV power forecasting performance. Even though various hybrid transformer based models discussed in the literature have tried to improve forecasting accuracy by integrating transformer models with convolutional networks, recurrent networks, or time series decomposition, most of such models include other modules in an ad hoc manner without a specific mechanism of modeling various forms of temporal dependencies. The main difference of the proposed method is the design and integration of a temporal convolutional module (TCM) as a special purpose component to learn the trend subseries. In this design, TCM is not just an auxiliary convolutional block; it is explicitly tasked with the role of extracting long term, non-stationary, and multi scale variations that are embedded within the trend component. Temporal convolutional structures enable TCM models to model local and continuous dependencies more effectively than AMs. Conversely, the transformer is only tasked with learning long range dependencies and periodic patterns that exist in the seasonal component. This learning task decoupling of TCM and transformer ensure that each subnetwork focuses on a specific class of temporal pattern, without the information interference that is common to most hybrid transformer designs. The proposed approach aims not only at integrating a number of existing elements, but at creating a purpose-driven architecture in which TCM is the central element in trend modeling and the complement of the transformer. It is this particular architectural design that makes the proposed model unique in comparison to most of the literature's transformer-based hybrid approaches. This paper presents the new hybrid system dubbed STL-TCM-Former to meet the challenges, comprising of STL decomposition and a dual-path transformer model with TCM. It is also notable that the TCM-Former can be viewed as a successful bridge of global modeling and local feature learning, where the transformer captures the long-range dependencies, and TCM learns the fine-grained temporal variations, to unite the two methods in a distinctive and complementary manner that is effective in PV forecasting. The structure would be effective in capturing the long-term periodic and the short-term variation in PV power data. STL decomposition separates the trend and seasonal parts, which are further enhanced to obtain multiscale temporal features. The dual-path architecture reduces component interference, increases model interpretability, and boosts robustness.

The major contributions of the presented study could be summed up as follows:

i) Creation of a new PV power predictive system named temporal convolutional module enhanced transformer (TCM-Former): The suggested model, named TCM-Former, simultaneously captures global and local temporal dependencies with the use of a dual-path structure, which integrates transformer-based blocks and TCM blocks. TCM-Former allows richer and more comprehensive multi scale temporal representation regarding PV power time-series data by joint modelling long-term trends and short-term seasonal variation.
ii) Utilization of seasonal-trend decomposition using loess (STL): Using STL decomposition to divide the time series into two principal components-trend and seasonal- the learning model can capture the long-term variations and short-term periodic variations in the time series independently and precisely. To a large extent, this separation reduces information interference and overlap among components, thus improving the generalization ability of the model to deal with temporally dynamic and climate-dependent datasets.
iii) Integration of complementary periodic and temporal features: The encoder–decoder architecture is used to combine the output of two parallel sub modules period module and the temporal module. The Period module

uses attention in order to extract long-range dependencies and the temporal module, which is based on the TCM, concentrates on the short-term local temporal patterns. This dua-path fusion can provide complete and enhanced multiscale features representations of PV power time-series data.

iv) Application of the TCM module for trend component modeling: It uses a special TCM module to model the trend component (XTrend) that seeks to capture long-term temporal dynamics as well as non-stationary variations of PV power data. Also, the proposed design has better predictive accuracy and robustness to noise as compared to traditional models that approximate the trend by using linear models or shallow layers.

The importance of this research is that it could further develop the research community and the use of PV power forecasting. Precise and reliable PV forecasting is essential in increasing grid stability, energy management optimization, and facilitating the large scale integration regarding renewable energy sources. The proposed approach helps to develop more efficient and robust forecasting frameworks, as it deals with some of the main issues in modeling complex temporal dynamics. As a research perspective, findings give us an insight into how to improve time series learning to properly manage different temporal patterns. Practically, the enhanced forecasting accuracy can guide grid operators, energy planners and smart grid systems for making well-informed decisions, decrease the operational uncertainty and enhance the overall performance and sustainability of the energy systems.

The research is structured as follows: i) Section 2 provides a comprehensive review of the literature in the field of PV power forecasting; ii) Section 3 explains the methodology of the planned STL-TCM-former model; iii) Section 4 provides the experimental findings and includes quantitative and qualitative analysis on model performance; and iv) Section 5 summarizes the research and presents possible future research directions.

2. RELATED WORK

In the recent past, PV power prediction has received much attention on hybrid deep learning (DL) architectures to enhance prediction accuracy and robustness. CGAformer model [13] consists of one-dimensional convolutional neural networks (CNN1D) for extracting local features, a Global Additive AM for capturing global dependencies, and an auto-correlation module to learn periodic patterns. The last prediction is produced by a multilayer perceptron (MLP). Experimental findings indicated that the suggested models are significantly better than the traditional LSTM models in terms of RMSE and MAE, and also in terms of generalization. Nevertheless, the complexity of the model and its dependence on large-scale data may be the only restrictions. A hybrid model that incorporates a better whale variational mode decomposition (IWVMD) with a context-embedded causal convolutional transformer (CCTrans) is presented in [14] to solve the task of long-sequence feature extraction. The optimization strategy of the CDCTF also improves the performance of forecasting.

Comparable, [15] presents a multi-step prediction model that integrates CNN as a short-term dynamics network and transformer for long-term dependencies. Although these methods based on transformers are effective at capturing global temporal interactions, they can have difficulty at capturing short-term or local variation because of their inherent focus on long-range dependencies and global AMs. In order to capture local temporal dynamics, a number of studies integrate convolutional and recurrent architectures. The TCN-ECANet-GRU model [16] combines the time convolutional networks (TCN), efficient channel attention, and GRU to process nonlinear and high-dimension PV data during different weather conditions. The model has an excellent performance in all seasons.

Similarly, [17] suggests a complex hybrid framework (ICEEMDAN-TCN-BiLSTM-MHA) which breaks the signals down into intrinsic mode functions and introduces AMs to amplify the feature representation, leading to greater stability in different weather conditions. Transformer-based architecture has also been expanded to medium- and long-term forecasting. SP-transformer, presented in [18], uses spatiotemporal positional encoding and self-attention for capturing the relations between geographical, meteorological, and PV data.

In [19], the IFTformer model incorporates outlier detection, time-series decomposition, and a ProSparse AM to improve prediction accuracy and robustness. Even though they outperform, these transformer-based models continue to have drawbacks in terms of their ability to capture fine-grained local variations and short-term dynamics, which are instrumental to highly volatile PV generation. The research [20] outlines a systematic research on the significance of combining meteorological data and irradiance modelling with transformer networks, enhancing forecasting accuracy with OpenMeteo information.

A hybrid VMD-WOA-LSTM model is proposed in [21], in which decomposition and reconstruction of PV signals are optimized to yield high accuracy and stability in prediction. In [22], they introduce a new transformer-based framework regarding ultra-short-term PV power forecasting that takes into account weather pattern recognition, in which meteorological features and solar irradiation patterns are modeled together. Through experimental data, it is established that there are great improvements in the forecasting accuracy in very short time horizons but the model performance is very reliant on the quality of the input data. In [23], a transformer model with feature augmentation is enhanced to forecast short-term PV power that uses advanced feature extraction and normalization modules to enhance noise reduction and prediction stability. In this work, it is demonstrated that AMs in combination with feature enhancement are able to successfully reduce RMSE and MAE errors in energy distribution networks, but at the cost of a high computational cost and training time.

In [24], the study provides a deep probabilistic method of solar power forecasting and involves the combination of a transformer network with Gaussian process approximation to quantitatively model prediction uncertainty. It shows that transformer-based models can be used to forecast points as well as estimate a probability distribution of PV power, and can be used in risk management in smart grids. Nevertheless, the main limitations of this approach are the statistical complexity and the large amounts of training data required.

In [25], a thorough comparative analysis of the classical ML algorithms and the transformer-based deep models in solar energy forecasting is carried out on real-world dataset. It has been shown that transformer models are better than traditional ones, especially when using multi-source data and in different weather conditions, but still, high attention should be paid to hyperparameter optimization, otherwise results can deteriorate dramatically.

The research in [26] explores the issue of feature selection as well as hyperparameter optimization in transformer models to forecast PV power with the use of Swordfish movement optimization algorithm (SMOA). This paper demonstrates that efficient feature selection and automatic hyperparameters optimization could significantly increase the predictive accuracy and stability of the model in changeable weather conditions, however, it is still unclear how the convergence behavior of the optimization algorithm and its sensitivity to initial parameters are affected by changing weather conditions. In general, these works confirm that the use of hybrid models that integrate signal decomposition, AMs, and DL models are helpful in PV forecasting, and also indicate that transformer-based models need to be complemented by additional mechanisms to overcome the effect of short-term and local variability. A comparison of the reviewed methods is provided in Table 1.

Table 1. Comparison of related works

Ref.	Method	Advantages	Disadvantages
[13]	CGAformer (CNN1D + GADAttention + Auto-Correlation + MLP)	Effective multi-scale feature extraction; strong global dependency modeling; improved RMSE and MAE; good generalization	High model complexity; requires large datasets; limited interpretability
[14]	IWVMD + CCTrans (transformer) + CDCTF	Powerful multi-scale decomposition; strong long-sequence modeling; high accuracy and robustness	High computational cost; complex architecture; limited ability to capture short-term/local variations due to transformer
[15]	CNN + transformer (multi-step forecasting)	Combines short-term (CNN) and long-term (transformer) forecasting; flexible across time scales	Model balancing is challenging; Transformer part weak in modeling fine-grained local/short-term fluctuations
[16]	TCN-ECANet-GRU	Strong local and temporal feature extraction; good performance under nonlinear and seasonal variations	Sensitive to hyperparameter tuning; may struggle with very long-term dependencies
[17]	ICEEMDAN-TCN-BiLSTM-MHA	Strong noise resistance; effective in different weather conditions; good feature representation	Very complex architecture; high computational cost
[18]	SP-transformer	Captures spatiotemporal dependencies; high accuracy in medium/long-term forecasting	High complexity; requires spatial data; limited performance on short-term/local dynamics due to global attention
[19]	IFTformer (DIF + LOF + Decomposition + ProSparse transformer)	Robust to noise and outliers; strong feature extraction; high accuracy and generalization	Multi-stage preprocessing increases complexity; transformer struggles with local/short-term variations
[20]	Transformer + meteorological data (OpenMeteo)	Improved accuracy using external weather data; comprehensive modeling	Dependent on external data quality; transformer limitation in capturing rapid local fluctuations
[21]	VMD-WOA-LSTM	High accuracy; effective for non-stationary data; stable under fluctuations	Time-consuming optimization; lacks global dependency modeling

3. PROPOSED METHOD

In this research paper suggests a new transformer-based architecture, which is called TCM-Former, to forecast PV power. The major aim of such architecture is addressing the inherent constraints of self-attention (SA)-based models in modeling short-term and localized dependencies in time-series data. Though the AM has a high capability of modeling long-range dependencies, it tends to under-perform in the case of rapid fluctuations, transient variations and fine-grained local patterns. In overcoming this difficulty, the suggested architecture incorporates a TCM in parallel encoder-decoder lines, thus, allowing an intelligent mix of global and local dependency learning. This design facilitates a rich, multiscale temporal representation of PV power data.

In the first stage, the input time-series data are decomposed utilizing seasonal-trend decomposition using loess (STL) into two principal components: trend component (X_{Trend}) and the seasonal/periodic component (X_{Season}). The decomposition is useful in helping the model to independently learn both the long-term variations and the short-term periodic structures in the PV power signal. The de-composed time-series components are then fed into the proposed transformer network simultaneously and at the encoder and decoder stages, two parallel processing paths are used. Two complementary sub modules are used in the encoder, the Period Encoder, which uses the AM to capture long-range dependencies and recurring time patterns in the seasonal component, and the temporal encoder, which is based on the TCN, and extracts short-term and local features like abrupt changes and high-frequency noise.

The outputs regarding such two paths are fused for generating an improved multiscale representation of the seasonal component encapsulating both long-term dependencies and localized temporal structures. The decoder comprises two parallel submodules: period decoder, which applies masked AM to model future long-term dependencies; and temporal decoder, which uses TCN for refining and enriching short-term features for accurate prediction regarding the seasonal component. The outputs from these two decoder submodules are integrated for producing the final predicted PV power sequence. A dedicated TCN network is incorporated for extracting dynamic temporal features from X_{Trend}, ensuring accurate modeling regarding long-term variations and trends. This architecture produces a robust, comprehensive and hierarchically aware model that has the ability to simultaneously capture global and local temporal structures. As a result, the suggested TCM-former shows more stability and accuracy in terms of predicting PV power as opposed to traditional transformer-based frameworks.

The suggested STL-TCM-Former algorithm is effective in increasing the interpretability of models in comparison with the baseline models, which is mostly explained by the incorporation of the STL decomposition approach and a dual-path mechanism. The STL module breaks down the input signal explicitly into trend, seasonal, and residual components, allowing the model to analyze each structural component of the data separately rather than analyze a single entangled time series. This decomposition gives a clear picture of the role of various temporal patterns towards the final prediction.

As an example, the trend component will capture long-term drift, and the seasonal component will capture periodic trends. Dual-path architecture also enhances interpretability through separating short term local dynamics and long-term dependencies into two parallel processing streams. The convolutional receptive fields provided by the TCM path pay focuses on local temporal changes, and thus, it is simpler to attribute model decisions to short term fluctuations or abrupt changes. By contrast, the transformer models high-level interactions between the full sequence, and its attention maps provide a clear and inspectable view of how distant time steps impact one another. The general structure of the suggested approach is shown in Figure 1.

3.1. STL decomposition

In step 1 of the proposed approach, to increase the ability of the model to analyze the multiscale dynamics of PV power time-series data, the STL approach is used. Incorporating STL into the proposed TCM-Former framework allows the model to isolate long-term and stable trends as well as rapid and periodic variations, which will allow the model to learn the specifics of each temporal dynamics efficiently. In addition, the ability to select only the significant data in the raw data and filter the noise and other non-structural elements leads to a more efficient and stable learning process. This means that the encoder and decoder modules of the transformer can capture long-term and short-term dependencies more accurately and make PV power predictions more accurate. The length of T of the input time series is given as (1).

$$X = \{x_1 . x_2 . \dots .\ x_T\} \tag{1}$$

This time series can be broken down into three major components in the STL framework as (2).

$$x_t = T_t + S_t + R_t \qquad .t = 1.2. \dots . T \tag{2}$$

Here, T_t is the trend component, which is the long-term change or the general direction of the data; S_t denotes the seasonal component, which represents the repetitive or periodic changes over a given run; and R_t is the residual component, or the random changes or the unpredictable noise. The trend and seasonal components are estimated during the STL decomposition process by locally smoothing with the help of locally estimated scatterplot smoothing (LOESS). The above iterative filtering and updating of each other between the two main components result in stable convergence. This process has the following major stages:

i) Trend estimation with the use of the LOESS filter in order to smooth the data:

$$S_t = LOESS(X_t - T_t) \tag{3}$$

ii) Seasonal extraction of the residual data, after eliminating the trend:

$$T_t = LOESS(X_t - S_t) \tag{4}$$

iii) Residual computation: the difference between the observed data and the total of the estimated components:

$$R_t = X_t - (T_t + S_t) \tag{5}$$

Once the smoothing process and iterative convergence are completed, the two important elements $X_{Trend} = \{T_t\}$ and $X_{Season} = \{S_t\}$ are used as the main inputs to the suggested TCM-Former network. The trend component (X_{Trend}) and seasonal component (X_{Season}) are used in this structure to capture the long-term information and slow changes in PV power and periodic daily or seasonal changes, respectively. This decomposition not only simplifies the model and facilitates training but also allows processing each component separately in separate encoder and decoder streams. Finally, this design makes a contribution towards the accuracy and strength of PV power forecasting. By contrast, the residual component (R_t) is not part of the proposed framework because of its noise dominated, non-structured and inherently stochastic characteristics. This component can be regarded as the main indicator of irregular variations and noise in measurements that do not have any stable temporal dependencies, so its presence can cause an unnecessary uncertainty and have an adverse effect on the learning stability and predictive performance regarding the model. Therefore, it is intentionally excluded from the input pipeline to ensure that the model focuses on the most informative and structurally meaningful components of the time series. Figure 2 presents an example of a time series and its components after the STL decomposition process. Figure 2(a) shows the original time series. Also, the decomposed trend, seasonal, and remainder components are shown in Figures 2(b)-2(d), respectively.

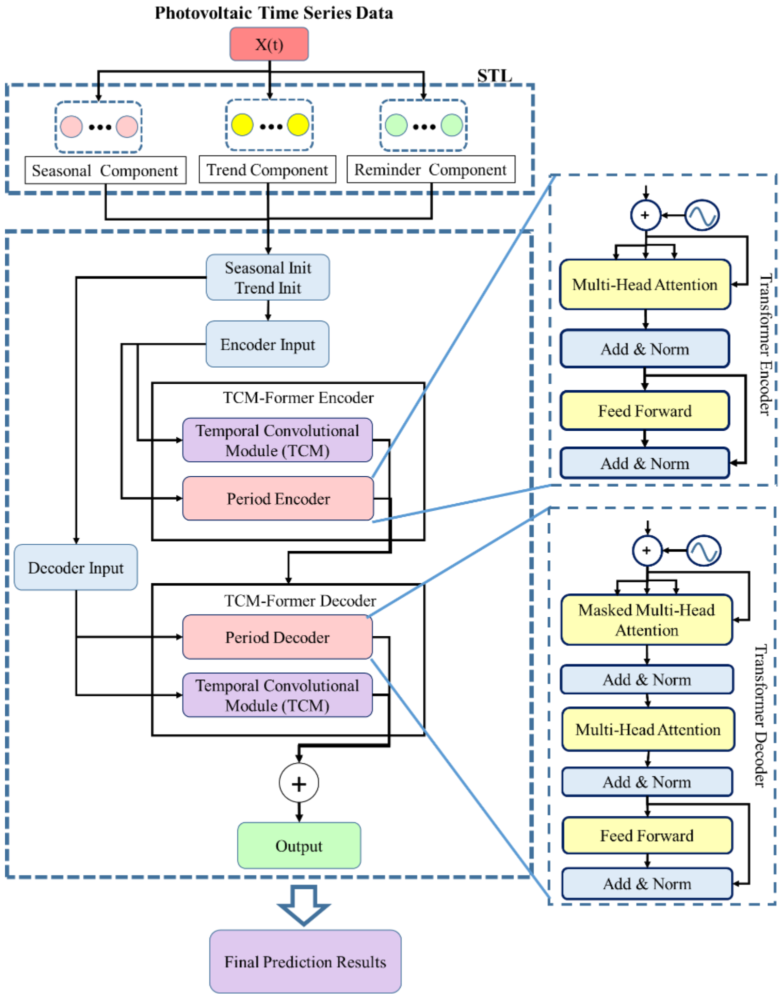


Figure 1. Diagram of the proposed TCM-former framework

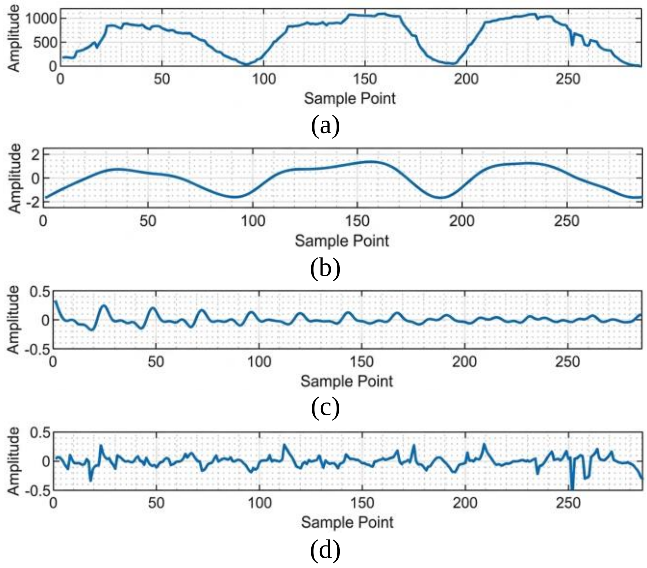


Figure 2. Example of a time series and its decomposed components using the STL algorithm: (a) photovoltaic time series signal, (b) trend component, (c) seasonal component, and (d) remainder component

3.2. TCM-former encoder

The proposed TCM-Former encoder architecture is developed to capture multiscale information from the seasonal component of the time-series data, enabling the system to represent short- and long-term temporal dependencies simultaneously. This part comprises two parallel paths, one transformer encoder and the other TCM path, which concentrate on long-range global dependencies and local features and short-term correlations, respectively. The model has a dual-branch structure that enables it to gain a more comprehensive insight into the temporal dynamics and, therefore, enhances the accuracy and stability of PV power forecasting. The architecture of the proposed TCM-Former encoder is depicted in Figure 3.

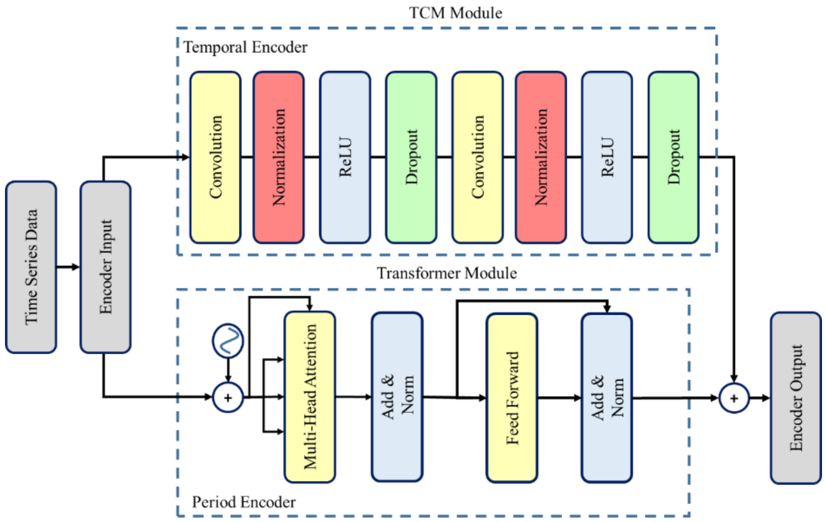


Figure 3. Structure of the encoder in the proposed TCM-Former network

The input sequence of the seasonal data in the transformer encoder Path, $X_{Season} \epsilon \mathbb{R}^{T \times d}$ is fed into a regular transformer encoder that consists of multi-head attention (MHA), add and norm, and feed-forward layers. The multi-head attention mechanism allows the model to acquire long-range dependencies between temporal points at various scales. The output of the attention is calculated as in (6).

$$Attention(Q.K.V) = Softmax(\frac{QK^T}{\sqrt{d_k}}) \quad (6)$$

Where the matrices are query, key, and value ($Q = XW_Q$, $K = XW_K$ and $V = XW_V$) and W_Q, W_K, W_V are parameters to be trained. The results of all attention heads are concatenated and linearly projected:

$$MultiHead(Q.K.V) = Concat(head_1. \ldots . head_h)W_0 \tag{7}$$

A residual connection is incorporated, followed by layer normalization:

$$Z_1 = LayerNorm(X + MultiHead(Q.K.V)) \tag{8}$$

Then, the normalized output passes through a feed-forward network (FFN) composed of two linear transformations with a ReLU activation:

$$Z_2 = ReLU(Z_1W_1 + b_1)W_2 + b_2 \tag{9}$$

After another residual connection and normalization, the ultimate output of the transformer encoder path is obtained:

$$H_{Trans} = LayerNorm(Z_1 + Z_2) \tag{10}$$

This branch focuses on extracting long-term periodicities and global dependencies in the seasonal information.

In the TCM Path, the model represents local and short-term dependencies. It has two successive convolutional blocks, each containing the following sequence of layers:

Convolution → Normalization → ReLU → Dropout.

Each block has a one-dimensional convolutional layer that extracts local temporal features. At time step (t), the convolutional operation is defined as (11).

$$C_t = W * X_t + b \tag{11}$$

In which ($*$) is the temporal convolution, (W) is the weights of the filter, and (b) is the bias term. After that, feature distribution is stabilized utilizing batch normalization:

$$\tilde{C}_t = BN(C_t) \tag{12}$$

The ReLU activation introduces nonlinearity:

$$R_t = ReLU(\tilde{C}_t) \tag{13}$$

Lastly, a dropout layer is added to avoid overfitting:

$$D_t = Dropout(R_t) \tag{14}$$

This step is repeated on the second convolutional block. The initial convolutional layer captures low level temporal information such as high frequency and localized variations, and the subsequent convolutional layer (with deeper receptive field) learns higher level inter period dependencies in a more abstract form. The final outcome of the TCM path is given as (15).

$$H_{TCM} = D_t^{(2)} \tag{15}$$

The results of the transformer encoder and the TCM module are then combined to form the final encoder representation:

$$H_{Enc} = Concat(H_{Trans}.H_{TCM}) \tag{16}$$

The resulting vector (H_{Enc}) is a summation of complementary information between the long-term dependencies (self-attention) and short-term dynamics (temporal convolution). This dual-path encoding architecture offers a potent and flexible structure for multiscale time-series modeling, which significantly enhances the quality and validity of photovoltaic power forecasting.

3.3. TCM-Former decoder

The decoder of the proposed TCM-Former network takes a dual-path design similar to the encoder, which allows it to model both the long-term global and the short-term local characteristics at the same time, while the encoder reconstructs and predicts the future seasonal components. The first branch follows the architecture of a transformer decoder, and the second branch employs the TCM as explained previously. Through the integration of the two complementary mechanisms, the decoder is provided with the capability to forecast the future seasonal trends accurately and consider both global and local temporal dependencies. The proposed TCM-Former decoder has the following structure, as shown in Figure 4.

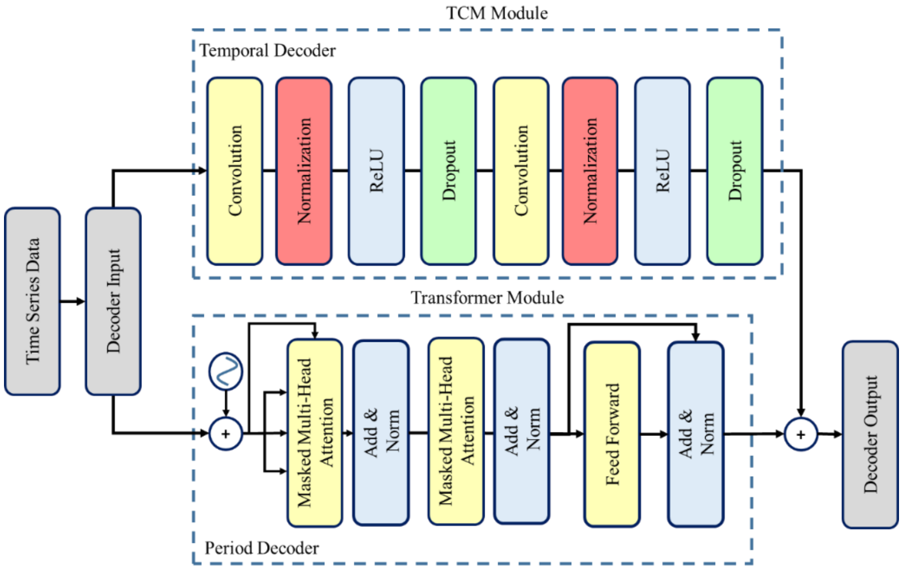


Figure 4. Structure of the decoder in the proposed TCM-Former network

In the first path, the input to the decoder, i.e., the predicted seasonal components up to time step t, is processed by the masked multi-head attention (MHA) layer. This layer operates similarly to the attention mechanism in the encoder, with the key difference being that it applies a lower-triangular mask to prevent information leakage from future time steps. This ensures that each position can only attend to its preceding positions, maintaining causal (autoregressive) modeling. Mathematically, it is formulated as (17).

$$MaskedAttention(Q.K.V) = Softmax(\frac{QK^T}{\sqrt{d_k}} + M) \tag{17}$$

Where (M) refers to the mask matrix, the values of which are set to $(-\infty)$ in future positions to apply zero attention weights. This operation guarantees sequential step-by-step prediction and is in line with causal-temporal modelling. The result of this layer is added to the input through a residual connection, and then layer normalization (Add & Norm) is performed to increase the stability of the training. Next, the output is passed into a second multi-head attention layer, where the Query (Q) is the output from the previous decoder layer, while the key (K) and value (V) are derived from the encoder's final representation (H_{Enc}). This cross-attention aligns the decoder intermediate representations with the contextual features of the encoder on many scales, thus providing the decoder with the ability to effectively utilize the historical temporal context to predict future seasonal components.

A second Add&Norm step is subsequently used, which is then followed by an FFN comprising two linear transformations separated by a ReLU activation function. The subnetwork provides nonlinear feature transformations and remapping that enable the model to learn sophisticated relations between temporal patterns. Once the final Add&Norm operation is done, the output of this branch is represented as $H_{Dec-Trans}$.

Similar to the encoder, in the second path a TCM is added to capture local variations and short-term dependencies in the predicted seasonal component. The TCM in the decoder consists of sequential convolution blocks (normalization blocks), ReLU blocks, and dropout blocks. The first convolutional layer focuses on learning fine-grained temporal features and high-frequency local variations, while the second convolutional layer, with a deeper receptive field, captures medium-term dependencies and broader contextual transitions. The result of the TCM path is represented as $H_{Dec-TCM}$. The results of the two decoder paths are combined to produce the final decoder representation:

$$H_{Dec} = Concat(H_{Dec-Trans}. H_{Dec-TCM}) \tag{18}$$

This combined representation is then passed through a fully connected (FC) layer that maps the learned feature space to the final prediction space:

$$\hat{Y}_{\text{Season}} = \sigma(W_f * H_{Dec} + b_f) \tag{19}$$

In which (W_f) and (b_f) are the parameters to be trained in the FC layer, and (σ) is the output activation function. The resultant enhanced seasonal forecast $\hat{Y}_{\text{Season}}$ combined with the trend component, which has been processed by the specialized TCM, is the ultimate power prediction of photovoltaic output. This design allows the TCM-Former decoder to attain a balanced representation of global and local temporal structures, which provides robust and highly accurate PV power forecasting outcomes.

4. RESULTS EVALUATION

For assessing the performance regarding the suggested approach in estimating the output power of PV systems, a series of experiments were carried out on both short-term and long-term bases. These experiments are aimed at determining the accuracy, the stability, and the generalization potential of the proposed model in different temporal conditions and environmental variations. To this end, the Yulara Solar Power System in Uluru (Ayers Rock), Australia was selected as the test data. Two-time horizons were forecasted: a short-term horizon (60 hours) and a long-term horizon (300 hours). The entire computational experiment was carried out in MATLAB on a computer running Windows 11, with 8 GB RAM, an Intel Core i7 processor, an 8 GB graphics card, and a 1 TB hard disk.

4.1. Dataset description

The data in this experiment was derived from the Yulara Solar Power Station in Uluru (Ayers Rock), Australia, covering the timeframe of 1 January 2018 to 31 December 2018 (DKA Solar Centre, 2018). This data is accessible on the DKA Solar Centre website [27]. PV arrays do not produce electricity at night time, and therefore, the data that fell within the range of 07:00 AM and 07:00 PM were analyzed. As a result, 13 samples were taken daily, and for 365 days, this resulted in 4745 hourly samples. The data temporal resolution is one hour, which produces 4745 time points. The dataset offers a realistic and representative standard of assessing short and long-term prediction of PV power models in different solar and climatic conditions.

4.2. Evaluation metrics

To quantitatively determine the performance and accuracy of the proposed approach in predicting PV output power, four metrics of statistical evaluation have been used: RMSE, MAE, mean absolute percentage error (MAPE), and coefficient of determination (R^2). All these metrics assess the model performance from a different perspective as defined below:

i) Root mean square error (RMSE)

The RMSE is used to measure the extent of the error in predictions by computing the squared differences between the predicted and actual values. This measure is especially sensitive to large errors, thus effectively highlighting significant deviations. A smaller RMSE indicates higher accuracy of the forecast. Where, (y_i) and ($\hat{y}_i$) are the actual and predicted values of PV power, respectively.

$$RMSE = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{20}$$

ii) Mean absolute error (MAE)

The MAE is an average of the magnitude of the absolute deviations between actual and predicted values. In comparison to RMSE, this measure is less sensitive to outliers and provides a direct measure of average prediction deviation.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} /y_i - \hat{y}_i/ \tag{21}$$

iii) Mean absolute percentage error (MAPE)

The MAPE is used to express the absolute error of prediction in percentages of the actual value. It is especially applicable in comparing the performance of the models on datasets of different scales. A lower MAPE indicates better predictive precision.

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \left/ \frac{y_i - \widehat{y_i}}{y_i} \right/ \tag{22}$$

iv) Coefficient of determination (R²)

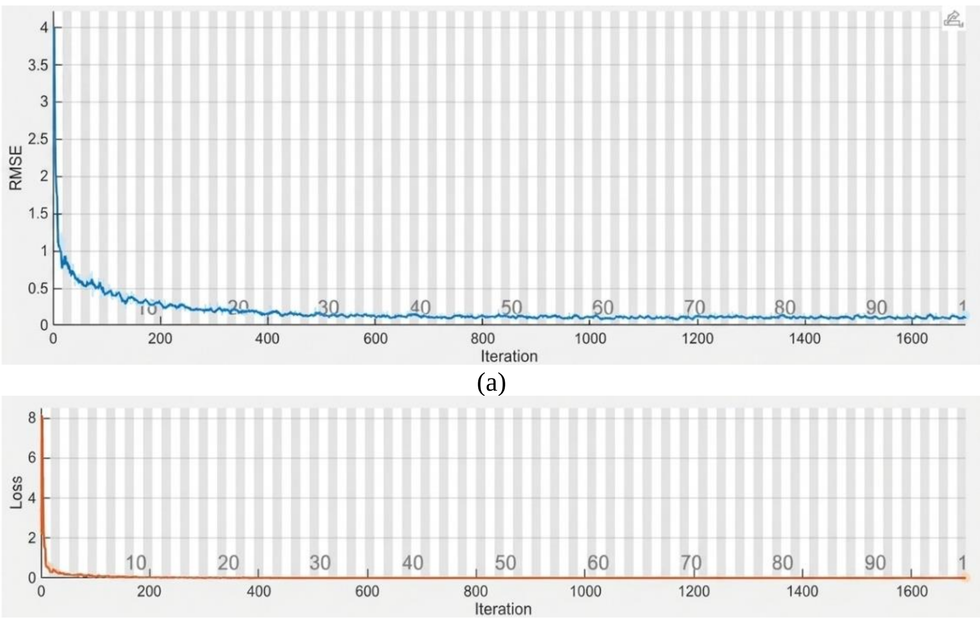
The coefficient of determination is used to determine the percentage of the actual data that can be determined by the predictions of the model. The R² value is between 0 and 1 and the higher the value, the greater the agreement between the forecasted and the actual data.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{23}$$

Where ($\bar{y}$) represents the average of the actual PV power values.

4.3. Evaluation of the proposed method during training

Figure 5 illustrates the point of convergence of the technique proposed during training. The plots are used to represent iterations on the horizontal axis and the upper and lower vertical axes represent RMSE (blue curve) and loss (orange curve), respectively. The value of RMSE is large at the start as depicted in Figure 5(a) indicating that in the initial training stages, there are large errors in prediction. The more the number of iterations, the smaller the RMSE becomes and it approaches a near constant value near to zero after a few hundred iterations. This trend indicates a progressive learning of this model with respect to the temporal dependencies in the data, which results in the maximized accuracy of prediction. Moreover, the stability of RMSE curve toward the end of the training indicates that the model is not overfitting. Likewise, the change in training of the loss function is shown by the Figure 5(b). The initial loss is very large, but, with the training process, it can be quickly minimized and its value approaches the same level near zero in several dozens of steps. Such sharp decrease shows that optimization of the network was achieved successfully and model parameters were updated for minimizing the objective function. On the whole, the steady reducing trends in both the RMSE and loss curves reveal the stability of the proposed method, high accuracy and great convergence during the training process. These results confirm the effectiveness of the suggested model in attaining a low error state, and the high generalization capability in predicting PV power.



(a)

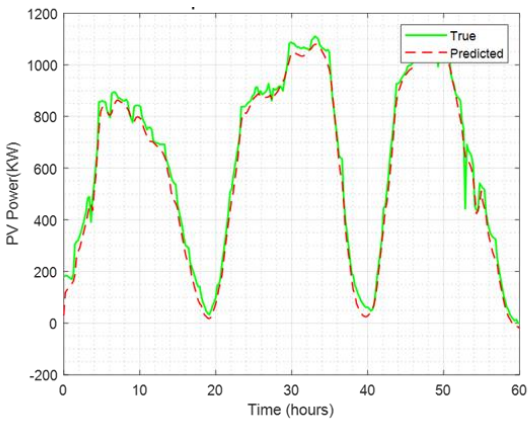
(b)

Figure 5. Learning curve of the proposed model: (a) learning curve in terms of RMSE and (b) learning curve in terms of loss

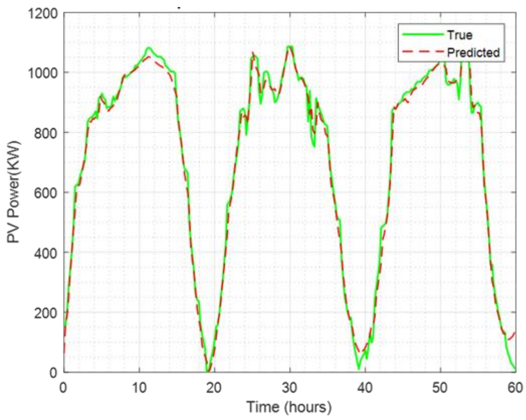
4.4. Photovoltaic power forecasting results for the short-term horizon (60 hours)

Figure 6 shows the comparisons between actual power and the predicted PV output power in a 60-hour short-term forecast and evaluated at two time intervals. The green curve in both plots shows the actual PV power and the red dashed curve shows the values that were produced by the proposed TCM-Former model. The model can capture the temporal dynamics and the general behavior of PV power generation and this is shown by the large overlap between the predicted and the actual curves as well as the small deviation at most time steps. Figure 6(a) is specifically connected with the initial 60-hour interval and the prediction tendency of the model is very stable. This prediction curve is extremely near the smooth up-and-down trend of the actual PV output particularly during midday peaks and zero generation at night. The fact that the modeling capability of this period is near perfect, highlights the ability to learn the dominant periodic pattern of solar production and experience small prediction error in moderately constant irradiance situations. The other 60-hour window that indicates more variations in PV output is shown in Figure 6(b).

This interval has sharper ramps, abrupt declines and irregular changes that are usually brought along by transient weather conditions like a passing cloud. Irrespective of such fast changes, TCM-former model is very responsive: it is able to accurately follow sudden increases and decreases in power and the prediction errors are small even in high-variability periods. This behavior demonstrates robustness of the model in the short term local fluctuations and the larger temporal dependencies. Both of the figures confirm the efficiency of the given architecture to learn periodic and non-periodic temporal features of PV power time series. The steady correlation between the predicted and actual values across different operating conditions implies that the TCM-Former network is providing stable, precise and realistic short-term PV power predictions that could be easily compared with observed values.



(a)

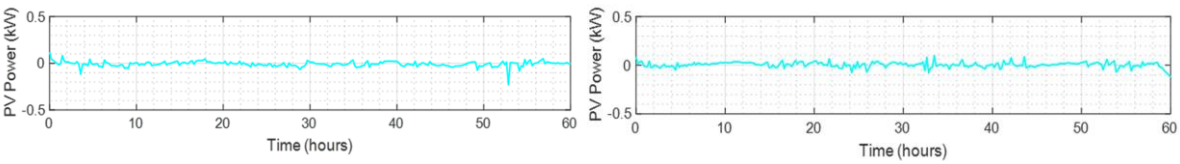


(b)

Figure 6. Comparison of actual and predicted PV output power over short-term intervals (60-hour horizon): (a) comparison for first observation period and (b) comparison for second observation period

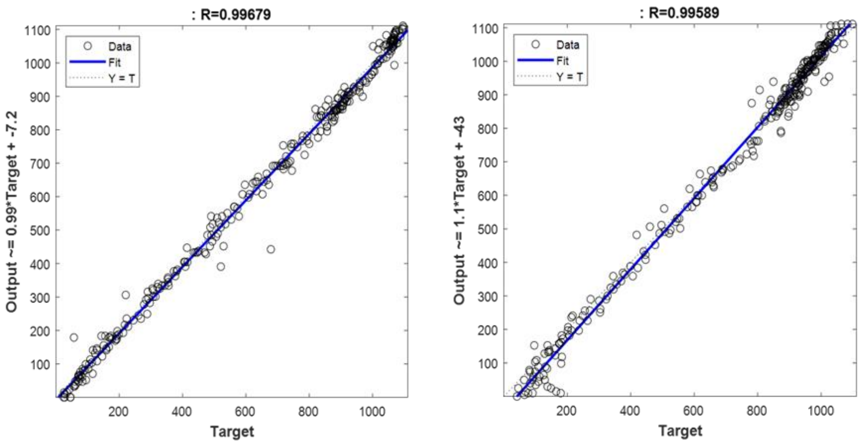
Figure 7 shows the normalized prediction error of the proposed model on a 60-hour short term forecasting horizon, at two different time intervals. As demonstrated, the prediction error within a very limited range and exhibits very stable pattern within the forecasting period, which demonstrates that the model has high capability of replicating the operational behavior of the PV system by not deviating much with the actual values. Figure 7(a) shows that the first 60-hour interval (Figure 7(a)) is highly smooth with the error profile being concentrated around zero. The deviations are minimal and that is, the model is highly accurate with relatively stable environmental conditions. The fact that there are no spikes of errors during this interval is an indication of the capacity of the model to follow the underlying PV generation pattern without being influenced by small changes in irradiance or temperature. Figure 7(b) on the other hand is another 60-hour segment that has a slightly higher variation in the PV output. With these changes, the normalized error curve is uniformly bounded and there is no significant drift or instability. The model can maintain its prediction accuracy when there are minor irregularities in the PV power and this proves the robustness and balanced sensitivity to dynamic environmental factors. This range of error distribution also indicates that the model would generalize to unseen data as well as not being overfitted to a specific set of conditions of the training data. The two subfigures together make sure that the proposed architecture of TCM Former provides a stable, low magnitude and a uniformly distributed error profile across the different operating periods. This action highlights the fact that the model has a high generalization capacity and is highly reliable in real world situations. On the whole, the results of Figure 7 suggest that the TCM Former model provides highly accurate and consistent short term PV power predictions and hence it is a good candidate to get practical use including energy management, generation scheduling and operational optimization of a renewable energy system.

Figure 8 has shown the regression analysis of short term PV output power over two-time intervals and actual measurements (target values) are plotted against the predicted outputs of the proposed TCM Former model. The data points in both subfigures are closely clustered around ideal correlation line (Y = T), which means that the predicted and actual values of PV power are in agreement. The first interval to be evaluated would be Figure 8(a). The regression line is very close to Y = T reference line and the data points are compactly distributed with minimal scatter in this plot. The fitted regression parameters (in particular, the slope very close to unity and a small intercept) indicate that the model creates almost unbiased forecasts within this interval. The correlation coefficient (R > 0.99) shows strong linear relation between the true PV power values and model outputs, which indicates the ability of the model for accurately capturing the underlying generation pattern. Figure 8(b) is a second time interval that has a little greater variability in the PV output. Regardless of these differences, the regression plot does not lose the high degree of accuracy: data points are again concentrated densely around the best correlation line, and the fitted regression line has again a slope that is close to one and an intercept that is close to zero. This consistency over intervals shows that the model does not show systematic bias even when the temporal conditions vary. The fact that the correlation coefficient is greater than 0.99 also confirms the robustness regarding the model and its great generalization ability to unseen data. Stability, reliability, and predictive power of the proposed TCM Former architecture are pointed out by the regression results in Figure 8. The close clustering of the data points, the near odeal regression parameters, and the high level of correlation indicating that the model provides very good predictions of short term PV power at various periods. The results support the model in the context of its application in practical scenarios like energy management, generation scheduling, and operational optimization in renewable energy systems.



(a) (b)

Figure 7. Normalized prediction error of the proposed model for 60-hour short-term forecasting across two-time intervals: (a) normalized error for the first observation period and (b) normalized error for the second observation period

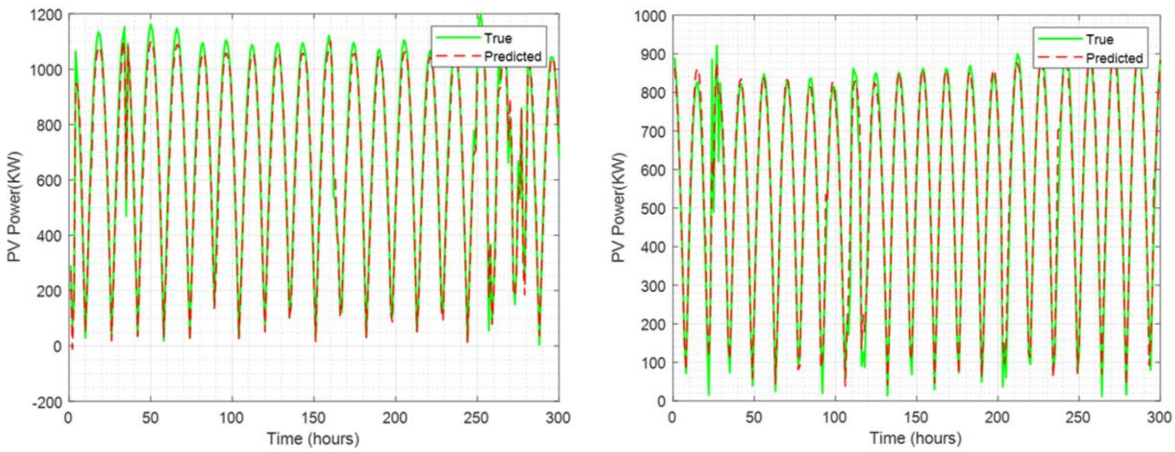


(a) (b)

Figure 8. Regression curves for short-term PV power forecasting across two different time intervals: (a) regression curve for the first observation period and (b) regression curve for the second observation period

4.5. Long-term photovoltaic power forecasting results (300-hour horizon)

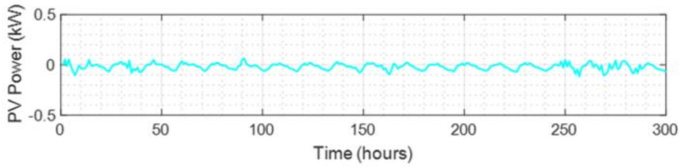
Figure 9 comparatively evaluates the measured and estimated output power of a PV during extremely long term forecasting horizon of 300 hours, at the two different intervals. The actual PV power data on the two plots is the green curve and the red dashed curve is the predicted values regarding the suggested TCM Former model. As shown, the predicted curves are very close to the actual data over the whole forecasting window indicating that the model can capture long term variations in PV output with great accuracy. Figure 9(a) relates to the first 300-hour interval where the model is able to follow the complete amplitude and phase of PV generation cycles. The forecasted curve fits practically with the actual values, particularly at the daily peak generation periods, and zero generation intervals during the night. This high value of the overlap indicates that the model can learn the periodic structure of solar power production and can be used with long horizons without apparent drift or degradation. A second interval of 300 hours (Figure 9(b)) is that which covers slightly different irradiance patterns and lower peak values. Even with these changes, the TCM Former model still manages to reproduce the temporal behavior of the PV output with significant accuracy. The predicted curve is well correlated with the actual signal even at the segments where the power of the PV exhibits the irregularities because of the changes of the environment. The model accurately follows both the peaks and trough of the power curve, attesting to its strength and flexibility to natural changes in solar conditions. The two subfigures reveal that the proposed architecture is very predictive at different long term intervals. The fact that the real and forecasted signals are highly similar with stable performance of amplitude and phase properties, showing generalization capability of the model and long-horizon stability. These results confirm that the TCM Former model provides accurate, consistent and robust PV power predictions both in short- and long-terms, and this indicates that the model is practical in the real world management of renewable energy, generation scheduling and operational optimization.



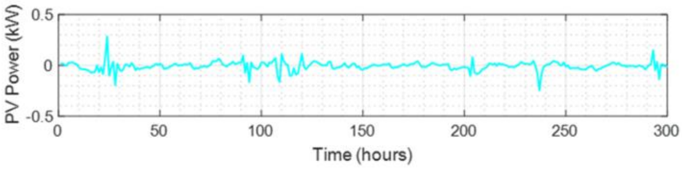
(a) (b)

Figure 9. Comparison of actual and predicted PV output power for long-term (300-hour) forecasting across two different time intervals: (a) comparison for the first observation period and (b) comparison for the second observation period

Figure 10, a normalized prediction error in the proposed model is plotted on a 300-hour prediction horizon as compared with two time segments. As illustrated, the normalized error always stays within a range of ±0.1 kW which is an indication of the high stability of the model, and its capacity to produce prediction that is very close to the actual PV measurements. The error distribution is centered at zero with no visible drift, and this means that the model does not introduce systematic bias to long term predictions. Figure 10(a) relates to the first 300 hours interval, in which the error profile has a smooth and well-confined pattern. The oscillations are also minor and symmetrically distributed around the zero level implying that the model is very predictive in relatively stable environmental conditions. This indicates that this model can capture long term PV generation dynamics without the performance losses by the fact that there was no sudden deviation in this interval. Figure 10(b) is a second 300-hour interval that has more apparent variability in the PV power signal. As the underlying data gets increasingly irregular, the normalized error remains well within the same range of ±0.1 kW. The model still gives unbiased predictions and the error curve does not exhibit any systematic increasing or decreasing trend. This stability over a longer and more complex period, which demonstrates the robustness of the TCM Former model and its adaptability to natural variations in irradiance, temperature or other environmental factors. In general, the two subfigures confirm that the proposed architecture has a stable, low magnitude and unbiased error profile of various long term forecasting segments. Its uniformity and symmetry of the error distribution reflect the great degree of generalization of the model and its resistance to various temporal and environmental forces. The findings in Figure 10 indicate that TCM Former model is a robust, predictable and stable model applicable in the long horizon PV power forecasting and hence is very applicable in practice as an energy scheduling, operational planning as well as maximization of renewable systems.



(a)



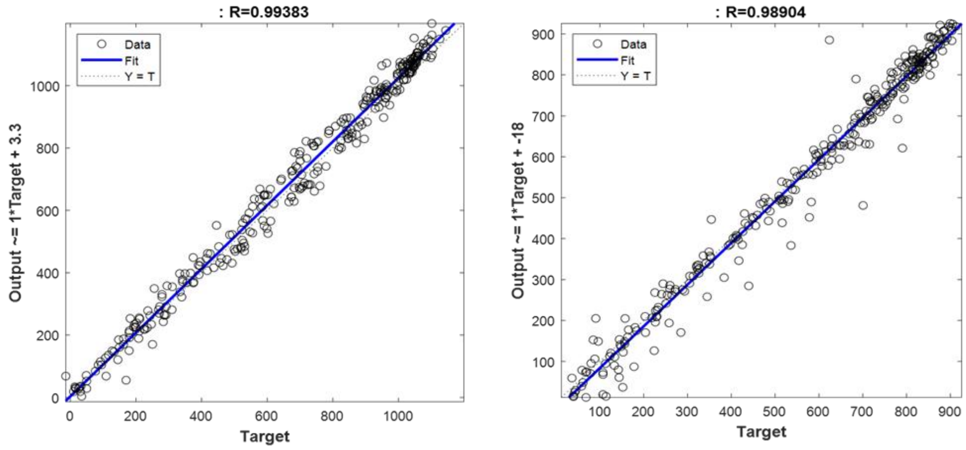
(b)

Figure 10. Normalized prediction error of the proposed model for 300-hour long-term forecasting across two-time intervals: (a) normalized error for the first observation period and (b) normalized error for the second observation period

The results of the regression analysis of the suggested model on long term PV power forecasting within two periods of time are shown in Figure 11. Circular markers in both subfigures indicate the actual target values and the fitted regression line shows the predicted outputs of the TCM Former model. The ideal correlation line (Y = T) is the dashed diagonal line, which is a perfect correlation between actual and predicted values. Figure 11(a) is an indication of the initial long term interval. In this plot, the data points tend to be concentrated closely about the line Y = T and there is slight dispersion. The fitted regression line has a slope that is very close to unity and a small positive intercept meaning that the model has a high predictive accuracy and minimal systematic bias.

The correlation coefficient of R = 0.99383 affirms the high linear consistency of the predicted and actual power of the PV model and this indicates that the model can maintain the accuracy in long forecasting horizons. The second long term interval is in Figure 11(b), where there is a little more variability in the PV power signal. Although this variability has increased, the data points are also closely aligned to the desired correlation line, and the fitted regression line has once again a slope that is close to one. The intercept in this interval is slightly larger, but by no means large to suggest a minimal bias. The fact that the correlation coefficient of R = 0.98904 shows that the model is both robust and that it can generalize well to various temporal conditions further confirm the validity of the model.

It would indicate that the model remains effective in predicting even with more uncertainty associated with the fluctuations in the environment, that the points always have the ideal line. The overall result of the regression in Figure 11 indicates that the proposed TCM Former model is reliable, stable and has good predictive abilities in long term PV power forecasting. The near ideal regression parameters and large value of correlation coefficients in the two intervals confirms that the model can make unbiased and correct predictions and thus justifying the suitability of the model in the actual world applicability such as long term energy planning, operational scheduling and optimization of renewable systems. The closeness of the correlation coefficients to unity shows the high learning ability and the strength of generalization of the proposed model when used to unknown data. Moreover, the fitted regression lines have slopes close to one and intercepts near zero indicating that there is no major systematic bias in the predictions. Collectively, these findings establish that the suggested TCM-Former model is reliable, accurate, and robust for long-term PV power forecasting across varying time periods and environmental conditions, reinforcing its practical utility in energy forecasting and management.



(a) (b)

Figure 11. Regression plots of the presented method for long-term PV power prediction across two different time intervals: (a) regression curve for the first observation period and (b) regression curve for the second observation period

4.6. Comparative analysis

A comparative study of the suggested STL-TCM-Former model with other benchmark models, including WOA-LSTM, LSTM, WOA-VMD-LSTM, VMD-LSTM, the Hybrid (WOA, VMD, LSTM) model [21], VMD-RF-LSTM [28], Informer, Autoformer and FEDFormer, is conducted under long-term (300 hours) and short-term (60 hours) forecasting horizons in Table 2. As it can be seen, all models usually work better in the short-term forecasting situation than the long-term. This is not surprising because short-term predictions involve less uncertainty and temporal variability and hence making them easier to model with high accuracy. Traditional model LSTM and WOA-LSTM have low R^2-values and high error rates in the long-term forecasting setting, and thus they are less effective at capturing long-range dependencies. The use of decomposition method (VMD-LSTM) and optimization methods (WOA-VMD-LSTM and hybrid models) are found to be of great benefit as they minimize errors in prediction and the model's capacity for addressing complex temporal patterns. Transformer models, such as Informer, Autoformer, and FEDFormer, are good at long-term forecasting, because they can model long-term dependencies. Nevertheless, they continue to perform worse than the suggested approach both in terms of accuracy and error reduction. The proposed STL-TCM-Former model gives the highest performance in both forecasting horizons. In the long-term case, it attains the highest R^2 (99.76) and the lowest error values (MAPE = 2.239, RMSE = 16.63, MAE = 10.36). In the short-term case, it further enhances performance as its R2 = 99.88 and considerably smaller error measures (MAPE = 1.684, RMSE = 13.905, MAE = 8.742). The findings indicate the success of the proposed model to explain both short and long-term dependencies. STL decomposition combined with TCM and transformer architecture enables superior feature extraction and robust temporal modeling over various forecasting horizons. In general, the results indicate that the STL-TCM-Former model is very reliable and accurate in both short run and long run PV power prediction task.

For assessing the performance regarding the suggested forecasting model on different datasets, the PV power data in PJM Interconnection database are used in this section and the performance of various approaches are compared to this dataset. The data reflect the PJM West Zone which is one of the main areas on which the PJM Interconnection covers the western parts of its power grid. The states of West Virginia, Pennsylvania (western regions) and Ohio are mainly covered by this zone. It is believed to be one of the most important areas of loads management and power generation because of its geographical position and the variety of energy sources. The dataset covers 34 days in 2020, the interval during which it takes place between January 1, 2020, and February 3, 2020. The data are recorded with an hourly resolution [29]. Table 3 provides a comparative analysis of the suggested STL-TCM-Former model with various recent benchmark approaches on the PJM Interconnection dataset in terms of four common performance measures, i.e., R2, MAPE, RMSE and MAE. As seen, the traditional LSTM model has comparatively poor results and its R2 is 79.012 and its error values are large (MAPE = 25.431, RMSE = 150.132, and MAE = 102.435), thus suggesting its low ability for capturing the complex patterns regarding PV power data. The addition of optimization methods, such as WOA-LSTM, only results in slight improvements, especially in the minimization of MAPE and MAE, but R2 and RMSE do not change significantly. Signal decomposition techniques are utilized for accomplishing a significant improvement. The VMD-LSTM model leads to a significant enhancement in prediction accuracy and increasing R2 to 95.763 and significant reduction of error measurements. Additional combination of the WOA with VMD-LSTM leads to a significant performance improvement, with R2 of 99.032 and much lower error values (MAPE = 5.503, RMSE = 23.039, MAE = 18.860). The hybrid WOA-VMD-LSTM model further improves the findings and achieves R2 of 99.198 with further decreases in all the error metrics. However, the suggested STL-TCM-Former model performs better than all the methods compared on all the evaluation criteria. It achieves the best coefficient of determination (R2 = 99.53) which means it has best goodness of fit and most importantly, it has the lowest error values (MAPE = 2.501, RMSE = 16.99 and MAE = 11.05). These findings reveal how the suggested model is effective in terms of capturing both linear and nonlinear trends of PV power data. All in all, the comparison shows that the suggested STL-TCM-Former model offers the most accurate and robust forecasting performance when compared to all the reviewed methods on the PJM dataset.

Table 2. Comparative performance of the suggested STL-TCM-Former model against benchmark methods based on R^2, MAPE, RMSE, and MAE

Method	Long-term (300 Hours)				Short term (60 Hours)			
	R2	MAPE	RMSE	MAE	R2	MAPE	RMSE	MAE
LSTM	80.805	25.332	148.777	101.257	84.912	21.684	132.541	90.336
WOA-LSTM	80.738	24.618	149.034	99.512	85.103	20.973	130.228	88.917
VMD-LSTM	96.065	14.665	67.335	47.711	97.218	12.304	58.941	41.285
WOA-VMD-LSTM	99.556	5.278	22.617	17.328	99.671	4.102	19.843	14.762
Hybrid (WOA, VMD, LSTM)	99.661	4.405	19.753	15.247	99.742	3.684	17.915	13.228
VMD-RF-LSTM	92.30	-	81.520	63.480	-	-	-	-
Informer	98.842	6.913	28.462	20.185	99.103	5.621	24.307	17.902
Autoformer	99.214	5.487	24.036	17.964	99.418	4.336	20.118	15.406
FEDFormer	99.513	4.962	21.884	16.271	99.635	3.958	18.762	14.113
Proposed (STL-TCM-Former)	99.76	2.239	16.63	10.36	99.88	1.684	13.905	8.742

Table 3. Comparative performance of the proposed STL-TCM-Former model against benchmark methods based on R^2, MAPE, RMSE, and MAE on the PJM Interconnection database

Method	2	APE	MSE	AE
LSTM	9.012	5.431	50.132	02.435
WOA-LSTM	8.921	4.932	51.133	00.001
VMD-LSTM	5.763	5.008	8.308	8.809
WOA-VMD-LSTM	9.032	.503	3.039	8.860
Hybrid (WOA, VMD, LSTM)	9.198	.780	0.391	6.001
Proposed (STL-TCM-Former)	9.530	.501	6.99	1.05

4.7. Ablation study

Table 4 gives an ablation study, aimed at studying the importance of each of the components of the proposed STL-TCM-Former. The analysis is based on assessing the individual and joint performance of the STL and TCM in combination with the transformer architecture. As the table demonstrates, the baseline transformer model is characterized by moderate performance with the R2 of 96.842 and error values that are rather high (MAPE = 9.731, RMSE = 52.184, MAE = 36.907). This shows that, even though the transformer can be used to model temporal dependencies, it does not fully capture non-stationary and highly fluctuating nature of PV power data. With the addition of the STL decomposition to the transformer (Transformer + STL), a significant enhancement is achieved on all measures. R^2 goes to 98.917 whereas the error values become much lower. This gain evidences the success of the decomposition of the original time series into a trend, seasonal, and residual, which makes the learning process easier and the model more effective in capturing underlying patterns. Likewise, using the temporal convolution module (transformer + TCM) results in additional performance improvement over the baseline model. The TCM improves local feature extraction and capture the short term temporal dependence better leading to fewer prediction errors and higher overall accuracy. The entire proposed model (STL-TCM-Former) that integrates both the STL decomposition and TCM with the transformer performs the best among all the configurations. It attains the highest R^2 (99.53) and the lowest error values (MAPE = 2.501, RMSE = 16.99, MAE = 11.05). This proves the fact that the two elements are complementary to each other: STL enhances the data representation through decomposition, and TCM enhances the temporal feature extraction. The results of ablation study show that every component positively affects the model performance, and their integration results in a significant increase in the forecasting accuracy.

4.8. Robustness testing

Table 5 assesses the robustness of the proposed STL-TCM-Former model in demanding conditions, such as the existence of noise and missing data, on the PJM Interconnection dataset. As seen, the performance of the model is strong even when noise is added to the data. In the case when the noise level is low (1%) the performance degradation is not significant in comparison with the initial results, with only minor increases observed in the MAPE, RMSE and MAE. The error metrics are slowly increasing as the noise power is increasing to 5% and 10, still, the overall performance stays within a reasonable range. This means that the model proposed is quite robust to noise and it can be used to capture the underlying signal effectively despite the perturbations. The same trend is seen when there are missing data. When 1% of the data is missing, the model performance is only marginally affected. As the missing rates (5% and 10% increase) the values of the error increase respectively and the degradation is not so severe. This shows that the model is powerful to deal with incomplete datasets. Two main factors can be linked to the robustness of the proposed approach. The STL decomposition assists in isolating noise in the residual component which minimizes its influence on the learning process. Second, the transformer architecture and TCM improve the model to capture both global and local temporal dependencies, allowing the model to predict reliably in the event of corrupted or missing parts of data. Generally, the findings confirm that the STL-TCM-Former model has a high level of robustness to noisy environments as well as to missing data cases, thus it can be considered a strong model when it comes to real-world PV power forecasting applications where imperfections in data are an unavoidable situation.

The results of the suggested STL-TCM-Former model in different weather conditions are shown in Table 6. Three typical weather conditions, sunny, rainy and cloudy, are used to evaluate the model in terms of its adaptability to various atmospheric conditions which is of high importance in determining the PV power generation. As indicated in the table, the model performs optimally when the weather is sunny where the lowest value of error occurs (MAPE = 1.842, RMSE = 13.215, MAE = 8.904). This comes as no surprise since the production of solar energy is more predictable and stable in situations where the solar irradiance is steady and unobstructed. The model is also satisfactory under cloudy conditions, but the error metrics increase moderately (MAPE = 2.915, RMSE = 18.447, MAE = 12.386). The existence of clouds brings variability of solar irradiance and therefore, the prediction task is more difficult; but the proposed model is successful in capturing these changes. When there are rains, the prediction task turns out to be very complex because of extremely unstable and lower solar irradiance. Consequently, the model exhibits higher error values (MAPE = 3.764, RMSE = 22.981, MAE = 15.672). This notwithstanding the performance degradation is still contained meaning that the model is robust even in unfavorable weather conditions. In general, the findings indicate that the suggested STL-TCM-Former model is reliable and accurate in predicting various weather conditions. STL decomposition and TCM integration allow the model to both accommodate stable and highly varying patterns in the generation of PV power, and provide stable operation in the real world.

Table 4. Contribution of each block in the STL-TCM-Former framework for PV forecasting

Method	APE	MSE	AE
Transformer	.731	2.184	6.907
Transformer + STL	.284	8.736	1.503
Transformer + TCM	.112	4.981	8.764
Proposed (STL-TCM-Former)	.239	6.63	0.36

Table 5. Investigating the robustness of the proposed method in noisy and missing data conditions

Conditions	MAPE	RMSE	MAE
Noisy data with noise power 1%	2.873	18.421	12.604
Noisy data with noise power 5%	3.912	21.735	14.889
Noisy data with noise power 10%	5.684	27.918	18.762
Missing data with missing rate 1%	2.691	17.804	11.932
Missing data with missing rate 5%	3.745	20.986	14.203
Missing data with missing rate 10%	5.129	26.447	17.951

Table 6. Investigating the robustness of the proposed method in different weather situations

Weather	MAPE	RMSE	MAE
Sunny	1.842	13.215	8.904
Rainy	3.764	22.981	15.672
Cloudy	2.915	18.447	12.386
Total	2.239	16.63	10.36

4.9. Comparing the efficiency of the proposed model

Table 7 summarizes computational efficiency of the suggested STL-TCM-Former model compared to various benchmark LSTM-based methods both in the training and inference phases. As indicated, the traditional LSTM has a medium training time of 312.45 seconds and inference time of 12.38 seconds. Adding optimization and decomposition methods like WOA and VMD dramatically raises the computational cost, introducing longer run times in all the hybrid versions. In particular, inference latency increases with WOA-LSTM, VMD-LSTM, and WOA-VMD-LSTM models (taking 524.91s, 468.33s, and 689.57s to train, respectively). The fully hybridized structure (WOA, VMD, LSTM) shows the highest computational cost, with the training time of 742.18s and inference time of 17.96s. Conversely, the proposed STL-TCM-Former, that uses a transformer-based framework augmented with STL and TCM, has the lowest execution time of all methods considered. Due to the inherent parallelization property of the transformer model and the lightweight design of TCM block, the model completes training in just 185.42 seconds, with considerably decrease in relation to all the LSTM-based baselines. The inference time has been minimized to 9.73 seconds and this affirms that the model has better computational efficiency when used in real-time or near-real-time applications. In general, the findings clearly indicate that the suggested STL-TCM-Former can not only enhance the accuracy of estimation (as the earlier sections have revealed) but also greatly decrease the computational cost. This is a twofold benefit that underscores the applicability of the model to the real world situation that needs a high precision and fast processing.

Table 7. Comparative performance of the proposed STL-TCM-Former model against benchmark methods based on training and test time

Method	Training time (s)	Test (inference) time (s)
LSTM	312.45	12.38
WOA-LSTM	524.91	14.72
VMD-LSTM	468.33	13.95
WOA-VMD-LSTM	689.57	16.84
Hybrid (WOA, VMD, LSTM)	742.18	17.96
Proposed (STL-TCM-Former)	185.42	9.73

4.10. Statistical analysis

The BoxPlot diagram in Figure 12 shows the statistical analysis of the performance of the different methods, by distribution of the RMSE values obtained after 30 independent runs at varying random seeds. As indicated, the VMD LSTM approach where the average RMSE is 67.97 has the highest prediction error and largest dispersion of all the approaches considered. It shows that, despite the fact that the use of VMD with LSTM is more effective than a standard LSTM in terms of forecasting ability, its performance is sensitive to changes induced by weight initialization and stochastic training conditions, which leads to limited stability. The average of the RMSE in the WOA VMD LSTM approach reduces significantly to 23.10 with the addition of the WOA optimization algorithm. Besides the drastic drop in the error magnitude, the drop in the box height (IQR) in the BoxPlot indicates a decrease in variability in the results and a better model stability. This proves that the optimization of parameters through WOA is a good way of reducing the effect of random factor in the training process. Under the Hybrid method, integrating VMD, WOA and an improved prediction structure, the average of the RMSE values reduces to 20.16 and the distribution of the error values is more compact than the earlier methods.

This lower dispersion means that the Hybrid model is more statistically stable and most likely to produce similar performance on various runs. Lastly, the proposed approach has the lowest average RMSE of 16.79 as compared to all methods. The narrow box range and the short distance between minimum and maximum values show that, besides getting a better accuracy, the proposed method also exhibits better statistical stability. That is, performance of proposed model shows minimal dependence on the selection of random seed and the results of the model are always reliable and closely clustered over several executions.

In general, the statistical analysis performed proves that the proposed approach is not only more effective than comparative approaches in minimizing RMSE but also increases the reliability and reproducibility of the forecasting process by decreasing the variability of the results. This confirms that the proposed model, in addition to enhancing the accuracy of prediction, creates a more stable model in the face of stochastic variations of the training process.

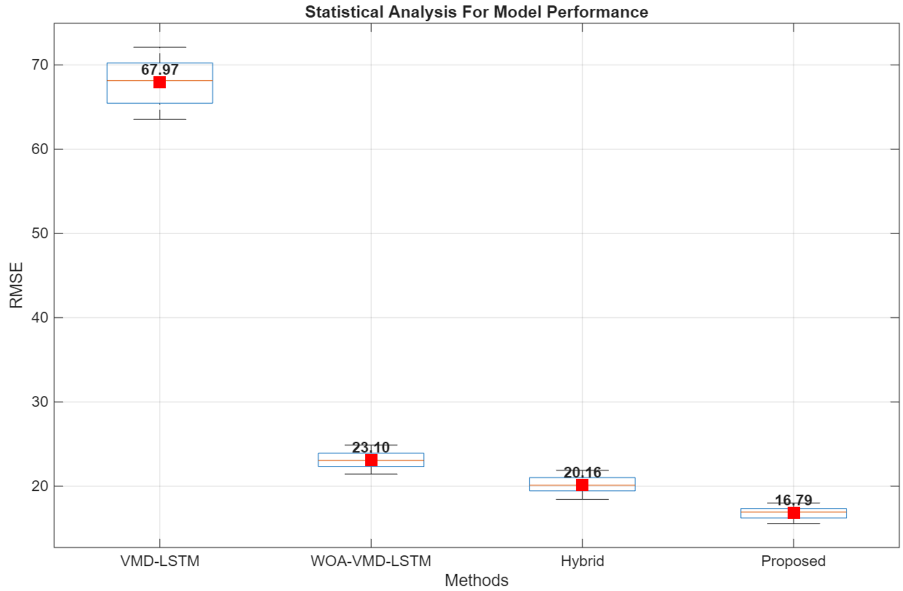


Figure 12. Statistical analysis for model performance

5. CONCLUSION

To ensure stability and performance of modern renewable energy systems, accurate prediction of PV power is crucial. This paper has introduced a new DL model, STL-TCM-Former that combines Seasonal-Trend decomposition through LOESS (STL) with a transformer architecture that is augmented with TCM to efficiently capture the nonlinear and multi-scale dynamics of PV power generation. The model can represent both long-term dependencies and short-term variations in a unified framework by breaking down the input time series into trend and seasonal components and processing them through dual transformer-TCM pathways. Extensive tests were made on the Yulara (Uluru) solar dataset and have shown that the suggested model performs much better than traditional and state-of-the-art procedures. In particular, STL-TCM-Former achieved an R^2 of 99.76%, MAPE of 2.239%, RMSE of 16.63, and MAE of 10.36. The proposed approach showed better predictive accuracy, with the MAPE decreased from 4.405% to 2.239% and RMSE decreased from 19.753 to 16.63 in comparison to the hybrid WOA-VMD-LSTM model. Likewise, the traditional LSTM-based models are significantly more improved, which demonstrates the efficiency of using signal decomposition and hybrid global-local temporal modeling.

Along with its predictive performance, the suggested framework offers an interpretable structure through the separation of trend and seasonal dynamics and minimizing the complexity of their interaction. Moreover, the modular architecture of STL-TCM-Former has a significant educational benefit, since it can be easily applied in laboratory settings of renewable energy and artificial intelligence to illustrate the essential concepts, including time-series decomposition, AMs, and the hybrid DL architecture. Even though the results are promising, many limitations must be noted. To begin with, despite the fact that the model has been tested on two datasets, including Yulara and PJM dataset, the variety of the PV system designs is limited and this could influence the generalizability of the proposed model to other regions and operating conditions. Second, although ablation experiments and robustness tests have been carried out to prove the efficiency and stability of the model, additional validation in more diverse real-world conditions and longer time horizons would offer an even more detailed evaluation of its reliability. The proposed framework will be expanded to multi-site PV forecasting condition in future work in order to enhance spatial generalization. Also, probabilistic forecasting methods can be used to offer uncertainty quantification in order to make more reliable decisions. Lastly, the application of the model in hybrid PV-battery energy management systems is also a promising direction to assist intelligent energy storage optimization and grid-scale uses.

FUNDING INFORMATION

The authors declare that the research presented in this manuscript did not receive any financial support from funding agencies, institutions, or external sponsors. All stages of the study, including data analysis, methodological development, and manuscript preparation, were conducted without the involvement of funded projects or grant-based resources.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Sarab Al-Chlaihawi	✓	✓	✓	✓	✓	✓			✓	✓				
Mohammed A. T. Alrubei	✓	✓	✓	✓	✓	✓			✓	✓				
Faris A. Alhaddad	✓	✓		✓	✓	✓			✓	✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no financial, personal, or professional conflicts of interest related to the research presented in this manuscript. No non-financial competing interests—such as political, academic, ideological, or intellectual influences are associated with the preparation, analysis, or interpretation of the work. The authors state no conflict of interest.

DATA AVAILABILITY

The data supporting the findings of this study are openly available through the DKA Solar Centre. Specifically, the Yulara Solar System dataset used in this research can be accessed at the following link: https://dkasolarcentre.com.au/download?location=yulara. The dataset is publicly accessible, and the authors confirm that all analyses in this manuscript were conducted using this openly available resource.

REFERENCES

[1] J. Yu *et al.*, "Deep learning models for PV power forecasting: review," *Energies*, vol. 17, no. 16, p. 3973, Aug. 2024, doi: 10.3390/en17163973.

[2] Y. Ledmaoui, A. El Maghraoui, M. El Aroussi, R. Saadane, A. Chebak, and A. Chehri, "Forecasting solar energy production: A comparative study of machine learning algorithms," *Energy Reports*, vol. 10, pp. 1004–1012, Nov. 2023, doi: 10.1016/j.egyr.2023.07.042.

[3] M. Husein, E. J. Gago, B. Hasan, and M. C. Pegalajar, "Towards energy efficiency: A comprehensive review of deep learning-based photovoltaic power forecasting strategies," *Heliyon*, vol. 10, no. 13, p. e33419, Jul. 2024, doi: 10.1016/j.heliyon.2024.e33419.

[4] M. Gupta, A. Arya, U. Varshney, J. Mittal, and A. Tomar, "A review of PV power forecasting using machine learning techniques," *Progress in Engineering Science*, vol. 2, no. 1, p. 100058, Mar. 2025, doi: 10.1016/j.pes.2025.100058.

[5] Y. Essam *et al.*, "Investigating photovoltaic solar power output forecasting using machine learning algorithms," *Engineering Applications of Computational Fluid Mechanics*, vol. 16, no. 1, pp. 2002–2034, 2022, doi: 10.1080/19942060.2022.2126528.

[6] M. Alrashidi and S. Rahman, "Short-term photovoltaic power production forecasting based on novel hybrid data-driven models," *Journal of Big Data*, vol. 10, no. 1, p. 26, Mar. 2023, doi: 10.1186/s40537-023-00706-7.

[7] J. Gaboitaolelwe, A. M. Zungeru, A. Yahya, C. K. Lebekwe, D. N. Vinod, and A. O. Salau, "Machine learning based solar photovoltaic power forecasting: a review and comparison," *IEEE Access*, vol. 11, pp. 40820–40845, 2023, doi: 10.1109/ACCESS.2023.3270041.

[8] F. Sun, L. Li, D. Bian, W. Bian, Q. Wang, and S. Wang, "Photovoltaic power prediction based on multi-scale photovoltaic power fluctuation characteristics and multi-channel LSTM prediction models," *Renewable Energy*, vol. 246, p. 122866, Jun. 2025, doi: 10.1016/j.renene.2025.122866.

[9] J. Zhang, Z. Liao, J. Shu, J. Yue, Z. Liu, and R. Tao, "Interval prediction of short-term photovoltaic power based on an improved GRU model," *Energy Science & Engineering*, vol. 12, no. 7, pp. 3142–3156, Jul. 2024, doi: 10.1002/ese3.1811.

[10] J. Kim, J. Obregon, H. Park, and J.-Y. Jung, "Multi-step photovoltaic power forecasting using transformer and recurrent neural networks," *Renewable and Sustainable Energy Reviews*, vol. 200, p. 114479, Aug. 2024, doi: 10.1016/j.rser.2024.114479.

[11] J. Wu, Y. Zhao, R. Zhang, X. Li, and Y. Wu, "Application of three transformer neural networks for short-term photovoltaic power prediction: a case study," *Solar Compass*, vol. 12, p. 100089, Dec. 2024, doi: 10.1016/j.solcom.2024.100089.

[12] F. Mo *et al.*, "A novel multi-step ahead solar power prediction scheme by deep learning on transformer structure," *Renewable Energy*, vol. 230, p. 120780, Sep. 2024, doi: 10.1016/j.renene.2024.120780.

[13] R. Chen, G. Liu, Y. Cao, G. Xiao, and J. Tang, "CGAformer: Multi-scale feature transformer with MLP architecture for short-term photovoltaic power forecasting," *Energy*, vol. 312, p. 133495, Dec. 2024, doi: 10.1016/j.energy.2024.133495.

[14] X. Wang and W. Ma, "A hybrid deep learning model with an optimal strategy based on improved VMD and transformer for short-term photovoltaic power forecasting," *Energy*, vol. 295, p. 131071, May 2024, doi: 10.1016/j.energy.2024.131071.

[15] J. Moon, "A multi-step-ahead photovoltaic power forecasting approach using one-dimensional convolutional neural networks and transformer," *Electronics*, vol. 13, no. 11, p. 2007, May 2024, doi: 10.3390/electronics13112007.

[16] X. Xiang, X. Li, Y. Zhang, and J. Hu, "A short-term forecasting method for photovoltaic power generation based on the TCN-ECANet-GRU hybrid model," *Scientific Reports*, vol. 14, no. 1, p. 6744, Mar. 2024, doi: 10.1038/s41598-024-56751-6.

[17] Y. Li, S. Zhai, G. Yi, S. Pang, and X. Luo, "Short-term photovoltaic power forecasting based on ICEEMDAN-TCN-BiLSTM-MHA," *Symmetry*, vol. 17, no. 10, p. 1599, Sep. 2025, doi: 10.3390/sym17101599.

[18] B. Wang, J. Chen, Y. Zhu, J. Fan, J. Hu, and L. Tan, "SP-transformer: a medium- and long-term photovoltaic power forecasting model integrating multi-source spatiotemporal features," *Applied Sciences*, vol. 15, no. 21, p. 11846, Nov. 2025, doi: 10.3390/app152111846.

[19] X. Liu, Q. Liu, S. Feng, Y. Ge, H. Chen, and C. Chen, "Novel model for medium to long term photovoltaic power prediction using interactive feature trend transformer," *Scientific Reports*, vol. 15, no. 1, p. 6544, Feb. 2025, doi: 10.1038/s41598-025-90654-4.

[20] G. Piantadosi, S. Dutto, A. Galli, S. De Vito, C. Sansone, and G. Di Francia, "Photovoltaic power forecasting: A transformer based framework," *Energy and AI*, vol. 18, p. 100444, Dec. 2024, doi: 10.1016/j.egyai.2024.100444.

[21] Z. Hou, Y. Zhang, Q. Liu, and X. Ye, "A hybrid machine learning forecasting model for photovoltaic power," *Energy Reports*, vol. 11, pp. 5125–5138, Jun. 2024, doi: 10.1016/j.egyr.2024.04.065.

[22] Q. Wu and C. Han, "A novel framework for ultra-short-term photovoltaic power forecasting based on improved transformer and weather pattern recognition," *Solar Energy*, vol. 304, p. 114183, Jan. 2026, doi: 10.1016/j.solener.2025.114183.

[23] M. Liu, S. Rao, M. Huang, and S. Deng, "Short-term photovoltaic power forecasting based on improved transformer with feature enhancement," *Sustainable Energy, Grids and Networks*, vol. 43, p. 101759, Sep. 2025, doi: 10.1016/j.segan.2025.101759.

[24] B. Xiong, Y. Chen, D. Chen, J. Fu, and D. Zhang, "Deep probabilistic solar power forecasting with transformer and Gaussian process approximation," *Applied Energy*, vol. 382, p. 125294, Mar. 2025, doi: 10.1016/j.apenergy.2025.125294.

[25] Z. Arslan, M. E. Sahin, and H. Ulutas, "Comparative evaluation of machine learning and transformer-based deep learning for solar energy forecasting on a real-world dataset," *Arabian Journal for Science and Engineering*, vol. 51, no. 5, pp. 6833–6863, Mar. 2026, doi: 10.1007/s13369-025-10769-8.

[26] E.-S. M. El-kenawy, A. A. Alhussan, E. A. Mattar, and M. Radwan, "Feature selection and hyperparameter tuning in transformer-based deep learning models for photovoltaic power forecasting using the Swordfish Movement Optimization Algorithm (SMOA)," *International Journal of Electrical Power & Energy Systems*, vol. 174, p. 111509, Jan. 2026, doi: 10.1016/j.ijepes.2025.111509.

[27] Z. Hou, Y. Zhang, X. Cheng, and X. Ye, "Photovoltaic power forecasting based on variational mode decomposition and long short-term memory neural network," *Energies*, vol. 18, no. 13, p. 3572, Jul. 2025, doi: 10.3390/en18133572.

[28] H. J. Kim and M. K. Kim, "A novel deep learning-based forecasting model optimized by heuristic algorithm for energy management of microgrid," *Applied Energy*, vol. 332, p. 120525, Feb. 2023, doi: 10.1016/j.apenergy.2022.120525.

BIOGRAPHIES OF AUTHORS

Sarab Al-Chlaihawi is a lecturer in the Department of Communication Engineering at the Technical Engineering College of Najaf, Al-Furat Al-Awsat Technical University. She received her B.Sc. degree in electrical engineering from the University of Technology, Iraq, in 1995. In 2012, she received her M.Sc. degree in electrical and electronic engineering from Jawaharlal Nehru Technological University, Hyderabad, India. She received her Ph.D. degree in electrical engineering from the National University of Science and Technology Politehnica Bucharest, Romania, in 2018. Her research interests include power electronics, electric vehicles, and renewable and sustainable energy systems. She can be contacted at email: inj.srb@atu.edu.iq.

Mohammed A. T. Alrubei has been a lecturer in the Department of Communication Engineering at the Technical Institute of Najaf, Al-Furat Al-Awsat Technical University, Iraq, since 2007. He has held the academic rank of assistant professor since 2025. He received his B.Eng. degree in communication engineering from the Technical College of Najaf, Iraq, in 2007, his M.Eng. degree in information and communication technologies and communication systems from Kazan National Research Technical University, Russia, in 2014, and his Ph.D. degree in electronic engineering and communication systems from Vladimir State University, Russia, in 2024. His research interests include signal processing and digital measurement, digital filter design, wireless communication systems, artificial intelligence and its applications, optical systems, and the use of MATLAB for modeling and simulation. He can be contacted at email: inj.moh3@atu.edu.iq.

Faris A. Alhaddad is a lecturer in the Department of Electrical Techniques at Al-Najaf Technical Institute, Al-Furat Al-Awsat Technical University (ATU), Iraq. He received his B.Sc. degree in electrical engineering from the College of Engineering, University of Babylon, Iraq, in 2001. In 2021, he received his M.Sc. degree in power electronics and electrical machines from the Islamic Azad University, Kermanshah Branch, Iran. His research interests include power electronics, electric vehicles, and renewable and sustainable energy systems. He can be contacted at email: fares.mohammed.inj@atu.edu.iq.