International Journal of Power Electronics and Drive System (IJPEDS)
Vol. 17, No. 3, September 2026, pp. 2086~2100
ISSN: 2088-8694, DOI: 10.11591/ijpeds.v17.i3.pp2086-2100

Adaptive forget-gated BiLSTM enhanced by DTW based feature selection in solar PV forecasting

Are Sambasiva Rao[1], Kunada Dhana Sree Devi[2]
[1]Department of Computer Science and Engineering, GITAM School of Computer Science and Engineering (GSCSE), GITAM Deemed to be University, Hyderabad, India
[2]Department of Artificial Intelligence and Data Science, GITAM School of Computer Science and Engineering (GSCSE), GITAM Deemed to be University, Hyderabad, India

Article Info

Article history:

Received Apr 15, 2026
Revised Jul 22, 2026
Accepted Aug 14, 2026

Keywords:

BiLSTM
Deep learning
Dynamic time warping
Feature selection
PV solar energy

ABSTRACT

Growing exhaustion of fuel reserves and their harmful environmental impacts have driven the shift towards the maintenance of renewable energy sources like solar energy. PV systems, however, still face significant challenges while integrating renewable sources into existing utility systems. Particularly, environmental features like temperature, irradiance, and humidity are not perfectly well coordinated with the power output, are always asynchronous in nature, and affect the power prediction considerably. Close observation of energy datasets from many PV plants revealed many intrinsic environmental variables that are highly asynchronous. Many forecasting models learn redundant features which might seem useful, and thereby the test performance is overfitting. To address the nonlinear and asynchronous behavior of environmental variables, there is a serious requirement for intelligent feature selection algorithms guided by both correlation and temporal alignment metrics. This research proposes a novel adaptive dynamic time warping (A-DWT) feature selection with an adaptive forget gate (AFG-BiLSTM) to address the asynchronous issues. Experiments were conducted with varied environmental asynchronous features, and the results of the proposed model were compared with traditional BiLSTM and stacked BiLSTM models. In all the experiments, the proposed model showed decreased error by (94.9%) on Dataset-1 (0.069, 0.0035), by (94.6%) on Dataset-2 (0.078, 0.0042), and by (90.9%) on Dataset-3 (0.069, 0.0061) when compared to stacked BiLSTM. The MSE loss of the proposed method was observed to be between (0.3%) and (11%) on four datasets.

Corresponding Author:

Are Sambasiva Rao
Department of Computer Science and Engineering
GITAM School of Computer Science and Engineering (GSCSE), GITAM Deemed to be University
Hyderabad, India
Email: sare2@gitam.in

1. INTRODUCTION

The global shift towards the use of sustainable energy sources has remarkably accelerated the development and deployment of photovoltaic (PV) plants. The continual depletion of fossil fuel reserves and intense environmental concerns, solar energy today has emerged as a promising alternative for clean power generation. Modern PV systems today enable both residential and commercial consumers to generate their own electricity, transforming traditional consumers into prosumers. This dynamic drift has necessitated redirection of conventional PV solar energy systems towards the use of more intelligent and smart approaches for energy generation and maintenance of PV systems.

A critical problem with solar power generation systems is its working under varied and intense atmospheric conditions. The output of PV systems changes constantly due to dynamic atmospheric entities, mainly because of asynchronous environmental features like irradiance, cloud cover, temperature, and atmospheric humidity. The natural dynamic variability throughout the environmental cycles creates huge difficulties for maintaining stable and reliable PV systems for efficient energy generation. The fluctuations in PV plant energy can trigger critical issues like extreme voltage deviations, frequency irregularities, and power conversion challenges, all of which compromise the integrity of PV networks.

Precise and timely forecasting of solar power generation has therefore become essential for effective power management from PV plants. On the whole, to one hand, though multiple factors influence PV plant solar energy generation, the primary factor could be understanding the deep-rooted relationships and dynamic nature of atmospheric features such as ambient temperature, cloud cover, atmospheric dew, and technical conditions of PV cells and inverters, which majorly impact the overall system prediction performance. The multifaceted nature of these atmospheric features necessitates the need and usage of sophisticated prediction models capable of capturing deep relationships between environmental entities and solar power output [1].

Deep learning-based bidirectional long-short term memory (BiLSTM) models render advanced methodology for capturing both deep past and future dependencies in time series data, making them highly effective and preferable for solar power forecasting. These models are good at capturing the temporal data dependencies and retaining information over extended long-time sequences, thus making them well suited for modelling the time-dependent nature of solar power prediction. Much of the research on BiLSTMs showed better training accuracy but depleted test accuracy. Some of the explored reasons for depleted test accuracy might be, that temporal misalignments may not have been studied well and that the models learned too many redundant features unknowingly.

Temporal misalignment is especially critical in PV forecasting because solar power follows a tightly deterministic diurnal cycle, so even small lags between meteorological inputs and the target signal can distort the learned mapping. Particularly in PV datasets, there are many redundant features like humidity, point humidity, cloud spread, cloud cover, irradiance, diffuse irradiance, horizontal irradiance, and even certain temperature measurements, which exhibit strong mutual correlations and can introduce redundancy into predictive models if not properly filtered [2]. These explored reasons are the focused, motivated objectives in our proposed work.

This research proposes a model for PV power prediction by stacking multiple BiLSTMs, with a novel feature selection of environmental variables. The BiLSTMs in this study are proposed to capture the past and future time series data dependencies so as to raise the prediction accuracy. A feature selection approach, adaptive dynamic time warping (A-DTW) with variance entropy weight, is proposed to address the temporal misalignments and to achieve better correlations of the time series features with the target. Instead of assigning identical weights to all features, A-DTW dynamically adjusts each feature's contribution according to its local statistical variability. This enables the algorithm to emphasize features with strong temporal relevance, thereby producing a more reliable warping path and improving the accuracy and stability of feature selection for solar power forecasting.

Redundant features that are captured and selected in the A-DTW module are filtered using an adaptive forget gate (AFG) with learnable weights, which is integrated into the proposed stacked BiLSTM [3], [4]. Unlike conventional forget gates, the proposed AFG adaptively regulates memory retention based on temporal feature relevance, thereby preserving informative features while suppressing redundant ones. The novelty of the approach lies in combining A-DTW with AFG for feature selection and redundancy elimination, which is not commonly addressed in prior works.

One of the major gaps in the modeling of deep learning algorithms is that many of the forecasting models assume instantaneous effects of environmental variables in power prediction and fail to model asynchronous and temporally misaligned relationships between environmental features and solar power output [5], [6]. Real-time models lack temporal alignment frameworks that dynamically adjust the lags for delayed impacts [7]. Both feature correlations and temporal dependencies are difficult for traditional forecasting models to adequately capture [8], [9]. Failing to focus on the study of asynchronous and imperfectly coordinated features critically affects the forecasting accuracy and model interpretability.

This study contributes to the field by addressing these gaps through the introduction of a novel learnable dynamic time warping (DTW) module that discovers and adapts optimal temporal alignment paths between environmental features and power output, effectively capturing the varying temporal scales and inherent delays across different meteorological variables. DTW is a powerful algorithmic technique that goes beyond traditional correlation methods.

In line with the identified gap, this research proposes the following objectives: develop a stacked BiLSTM to understand and capture the past and future time series data dependencies from PV data; develop a feature correlation method to capture the correlated environmental features with which the BiLSTM is trained; develop a robust approach to filter the redundant environmental features of the PV data. The significance of the proposed framework, with its clearly defined objectives, lies in its contribution to the research community by offering an efficient approach for temporal feature correlation and redundant feature elimination in high-dimensional environmental datasets. Practically, the framework can enhance PV system monitoring and performance analysis by improving data quality and supporting more reliable solar energy management systems.

Tina *et al.* [10] and Lai *et al.* [11], in their work, presented a systematic assessment of different machine learning methods for forecasting solar power. Their work emphasized the performance and effectiveness of ensemble hybrid models in identifying relevant patterns in PV data. Tai *et al*. [12] in their studies used convolutional neural networks (CNNs) along with bidirectional LSTM networks for capturing time-series temporal relationships of solar energy and irradiance. Their methods showed a marginal (30%) increase in forecasting accuracy compared to traditional algorithms when tested on different datasets.

Piantadosi *et al.* [14] and Kim *et al.* [14], in their studies, proposed a transformer-based solar power prediction model. This model used a multi-head attention mechanism to find temporal patterns and trends in solar data. The work of Stoean *et al.* [15] proposed various hyperparameter tuning approaches for optimizing the performance of deep learning forecasting models. Their experimental studies showed a comparative study where metaheuristic techniques often outperformed conventional evolutionary optimization algorithms, with R2 of (0.604) and MSE of (0.014). Sulaiman and Mustaffa [16] introduced an efficient differential approach for solar power forecasting with recurrent neural networks with optimized parameters. Their work proposed a mutation algorithm that adaptively changes the search parameters according to the optimization progress. The maximal error given by their approach stands at (7.303%). Fine-tuning deep learning models for time series forecasting was proposed in [17]. Their work included an iterative, fine-tuning-based approach for parameter optimization to avoid local minima while triggering a global exploration. The experimental studies showed 92% precision of the model.

Lightweight LSTM models were proposed for forecasting solar power, which has raised the prediction accuracy [18], [19]. To quantize the lightweight model, the works used pruning and quantization methods. Dao *et al*. [20] and Chang *et al.* [21] introduced lightweight embedded. An optimized extreme learning technique was focused on by Behera *et al*. [22], which dynamically modified model complexity based on the parameters. Their models under various operational settings incurred low computational cost, with a MAPE of 2.94% at 15 min of time interval. Niu *et al.* [23], Qu *et al*. [24], and Huang *et al*. [25] used a hierarchical attention model to capture historical trends at various time levels. Their approaches showed greater accuracy with extreme meteorological volatile data. The work of Gan *et al*. [26] presented a unique attention approach with a multi-headed mechanism for solar energy prediction with historical data. Their approaches demonstrated high performance in comparison to traditional LSTM, reducing MAE by 38.53%, and in multistep forecasting, a reduced MAE of 54.26% was observed.

Gensler *et al*. [27] in their work use a hybrid LSTM network for solar energy prediction in comparison to RNN models and showed better performance with LSTMs in understanding the solar irradiation feature. Their approach reduced absolute error, with a better performance of 87%. Yu *et al*. [28] and Awais *et al*. [29] in their works used enhanced LSTM with attention layers, which have the capability of understanding the past and present irradiance features with adaptive weights assigned. These adaptive weights are then used to study the feature relevance. They conducted experiments under extreme weather patterns and highlighted their model's high performance, raised by more than 15% when compared to traditional LSTM methods. Zhong *et al*. [30] presented a BiLSTM model with temporal attention for multistep lag study in forecasting irradiance. Chen *et al*. [31] modelled a sliced attention BiLSTM that uses satellite data for regional solar energy forecasting. Their model exhibited high prediction accuracy with a low MAE of 1.15. The work studied by Yu *et al.* [32] modelled a BiLSTM with double-decomposed layers for solar power prediction. Especially for longer prediction horizons, these stacked layers proved better for feature extraction and predictions, with reduced prediction error by 56.09% on cloudy days.

Studies were presented by Xie *et al.* [33] and Gu *et al*. [34] by setting up a multi-layered stacked BiLSTM with intermediate residual connections. Their experimental demonstrations showed that the proposed model addressed the vanishing gradient and achieved considerable performance. Liu *et al.* [35] in their work introduced a novel feature engineering method in a stacked BiLSTM model, with an accuracy improvement of 12.6%. Wavelet decompositions were used to break down time series into intermediate frequency components for adaptive feature engineering. Sarmas *et al*. [36] and Gao *et al.* [37] explored and presented a transfer learning approach to study highly variant solar data. Their experimental studies conducted on various datasets showcased the model's capability in understanding complex patterns in limited data.

Conventional feature selection approaches such as principal component analysis (PCA) and correlation-based methods have been extensively used for dimensionality reduction and redundancy analysis. Nevertheless, PCA projects the data onto orthogonal components by maximizing variance without preserving temporal dependencies, whereas correlation coefficients evaluate only linear relationships between synchronized observations and fail to capture time-lagged similarities [38], [39]. Consequently, these methods are less effective for solar power forecasting, where meteorological variables often exhibit nonlinear and temporally shifted interactions, motivating the adoption of DTW in this research for temporal alignment and similarity-based feature selection. DTW has emerged as a powerful method for addressing these limitations of distance metrics. Samara *et al*. [40] introduced DTW for time series analysis. Kim and Hong [41], in their studies, included DTW for comparing patterns in time series historical data. They used DTW within their proposed k-Shape clustering algorithm and experimentally showed 89% model accuracy in forecasting time series.

2. DATA AND METHODOLOGY

Many of the current solar energy prediction systems perform poorly on variable data features because the models are trained on one or two datasets with only a few environmental variables. This study includes empirical analysis on 4 different solar datasets, with a considerable number of environmental variables to understand the model behavior and variability with the features of the data. Dataset 1 is our custom dataset collected from a PV plant installed on the rooftop of an organization. The dataset encompasses around (105000) records collected over 2 years with 13 features. Most of the features in this dataset are environment-related. Dataset 2 is real-time data, taken from one of the solar power generation plants. The dataset has around (100000) records and 22 features. This dataset also has major environmental features. Dataset 3 is an open dataset available from a Git repository. The dataset has around 1360 records with 17 features. Dataset 4 is real-time data, taken from a PV plant power generation organization. The dataset has around 69480 records with 38 features. These datasets are made available in the links provided in the references.

Each dataset underwent a rigorous procedure of preprocessing, refining, transforming, and scrutinizing the raw data to generate meaningful relationships among the environmental features. Missing values were identified for each feature and imputed using forward filling for continuous meteorological variables. Missing records constituted 0.6%, 1.4%, 0.7%, and 1.1% of the total observations in Datasets 1–4, respectively. All continuous variables were normalized using min-max normalization, transforming every feature into the range [0,1]. The correlation analysis was conducted to determine the strength of relationships between environmental variables and energy output (power generated), establishing a foundation for subsequent feature selection and model development.

2.1. Proposed methodology

Temperature, irradiance, wind speed, and other environmental variables may change due to weather, geographic location, time of day, and seasonal patterns, making solar energy generation extremely variable. The power output from PV panels can fluctuate rapidly and unpredictably over short or long periods, driven by changes in external environmental conditions, thus causing the environmental variables to change dynamically. Both feature correlations and temporal dependencies are difficult to capture by traditional forecasting models under these variabilities.

Understanding the temporal dependencies is highly necessary to build models that can predict changes in power generation. Our research proposes a novel three-stage hybrid model to understand these temporal dependencies, which combines the contributions of: Stacked BiLSTM for capturing past and present dependencies; A-DTW for feature correlation and selection; novel BiLSTM with AFG for redundant feature elimination. We used two algorithms: one for feature selection and one for model training. Figure1 shows the flow design of the proposed methodology, which is elaborated in further sections. Better time series learning, ideal feature selection, redundant feature elimination, and increased predictive accuracy are guaranteed by these techniques.

2.1.1. DTW and A-DTW

A number of environmental conditions like temperature, cloud cover, and irradiation have a direct impact on solar power output. The impact of these variables on power production is not always precisely coordinated. For example, a scenario where irradiance spikes due to the sun appearing from behind a cloud, but the AC power doesn't instantly spike because of delayed panel heating. This causes reduced efficiency and can be really understood. Such scenarios may be indicators that there are different asynchronous delays and non-linear interactions in how changes in environmental circumstances impact the production of solar energy. Thus, a thorough understanding of the dynamic interactions between environmental features and power generation is essential for accurate and precise forecasting. From the theory of statistics, correlation analysis and lagged dependency exploration are preferred to analyze such behaviors.

As correlation only considers a one-time step, it may give a low result in capturing temporal misalignments. This may end up selecting features that falsely indicate maximal influence on the target power generation variable. This is the case where we need a technique that could wrap the time delays of various features, thus making them ideal for managing the asynchronous impacts of environmental features on power output. As environmental features like 'temperature', 'cloud cover', and 'solar irradiance' do not always have perfectly synchronized effects on power output, the warping technique becomes critical and helpful in the context of solar energy generation and may lead to better model predictions.

DTW is a powerful technique that goes beyond traditional correlation methods. By enabling non-linear alignment of time series data, DTW is a potent approach that surpasses conventional correlation methods. Because of its exceptional ability to align temporally unsynchronized time series data, DTW is ideally suited for managing the asynchronous impacts of environmental variables on power output. To find the best alignment between two feature sequences, the DTW approach will compress or stretch time series, in contrast to conventional correlation methods, and is given by: DTW(X, Y) = min $\sum_{(i,j)\epsilon path} d(x_i, y_i)$.

DTW assumes all features equally contribute to time series similarity, thereby selecting all features. The proposed adaptive DTW (A-DTW) expands on the idea of using dynamic weighting path analysis to pick the most significant features. It offers a deep understanding of the complex relationships between many environmental variables and solar power generation by calculating not only the distance but also examining the alignment of features using a warping path (WP). The value of the warping path from DTW indicates how close the two-time series are. By looking at the DTW distance, we can analyze how differently the time series behave, and the path alignment score explores how consistently the features track each other. As a whole, the algorithm determines correlation strength and provides information about which environmental factors have the biggest effects on solar energy production.

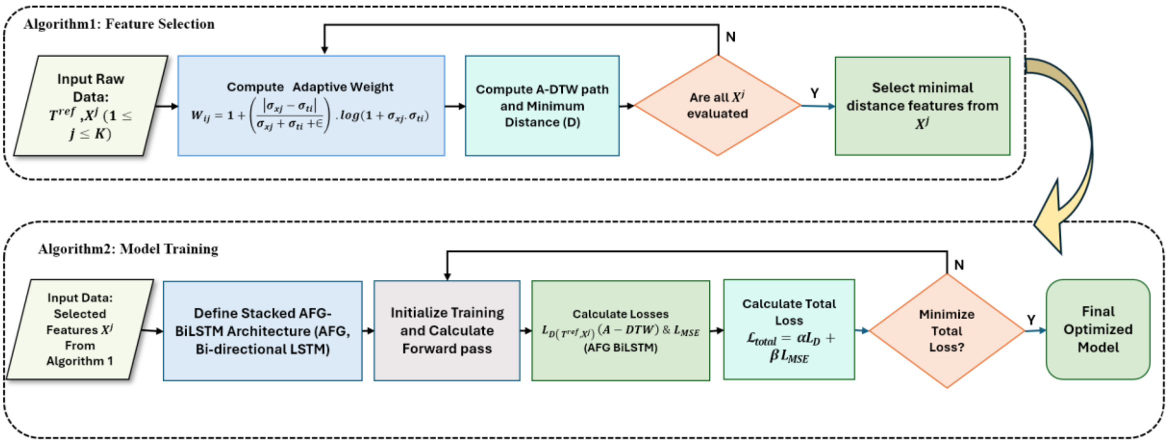


Figure 1. Proposed methodology pipeline

2.1.2. Feature selection using adaptive dynamic time warping (A-DTW) with variance entropy weight

Given a target time series (T^{ref}) and a collection of K time series, X_j $(1 \le j \le K)$, A-DTW aims to find a time series $X_j \in X^k$ that is closest to T^{ref}. Notations in (1) and (2) represent the A-DTW distance and cumulative minimal ADTW distances.

$$dist(t_i, x_j) \tag{1}$$

$$D(T^{ref}, X^j) \tag{2}$$

Mathematically, A-DTW is given by (3).

$$dist(t_i, x_j) = w_{ij}.(t_i - x_j)^2 \tag{3}$$

Where $t_i \in T^{ref}, x_j \in X^k$:

$$W_{ij} = 1 + \left(\frac{|\sigma_{xj} - \sigma_{ti}|}{\sigma_{xj} + \sigma_{ti} + \epsilon}\right) . \log\left(1 + \sigma_{xj}.\sigma_{ti}\right) \tag{4}$$

is the proposed adaptive weight based on variance. The warping path W is then constructed, where the wrapping path W is a sequence of points that align two time series sequences T^{ref} and $X^j$by, minimizing the cumulative distance as shown in (5).

$$W = \{(i_1, j_1), (i_2, j_2), \ldots \ldots, (i_r, j_r)\} \tag{5}$$

Where (i_r, j_r) represents the alignment of t_{ir} and x_{jr}.

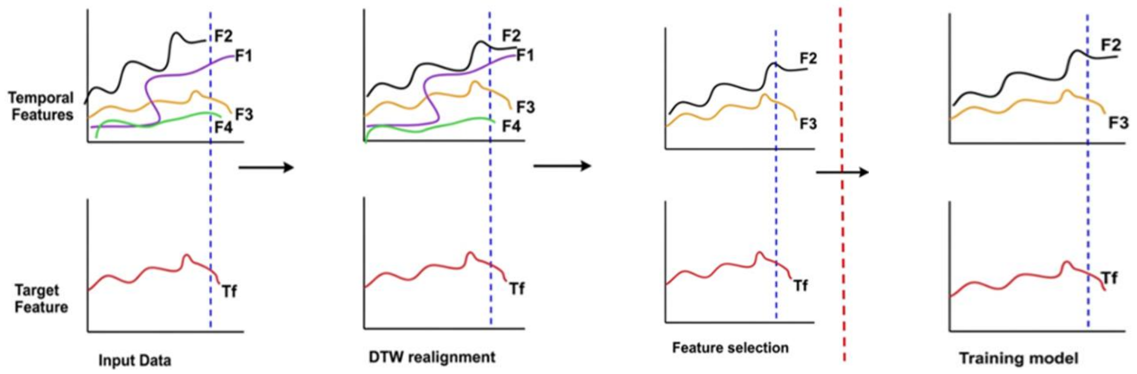
In adaptive DTW, the goal is to find the optimal wrapping path (W) between two time series sequences by minimizing the cumulative distance given by (6).

$$d(i,j) = min \begin{cases} D(i-1,j) + w_{ij} * dist(t_i, x_j) \\ D(i,j-1) + w_{ij} * dist(t_i, x_j) \\ D(i-1,j-1) + w_{ij} * dist(t_i, x_j) \end{cases} \tag{6}$$

These three minimal cumulative correspond to wrapping across vertical step, horizontal step, and diagonal step. The weight w_{ij} as shown in (4), is an adaptive weight that varies with the variance between the features (t_i, x_j). Higher weights are considered for (t_i, x_j) with high variance. $D(i,j)$ of (6) includes such low-variance points to be included in the optimal wrapping path W. The DTW distance between the target and selected time series feature $D(T^{ref}, X^j)$ is the minimum cumulative distance (Euclidean) of aligning all points along the warping path W. It is computed as (7).

$$D(T^{ref}, X^j) = \sum_{(i,j)\in w} dist\big(t_i, x_j\big) \tag{7}$$

This minimal distance from (7) shows $X^j$is closest to T^{ref} and X^j are selected features with importance. Figure 2 shows a sample of four F1, F2, F3, and F4 time series features compared to the target feature T_f for similarity alignment. Figure 2(a) shows initially all features to be completely dissimilar to the target; after DTW realignment shown in Figure 2(b), the features F2 and F3 are selected to be similar to the target, which are shown in Figure 2(c). These selected features are shown in Figure 2(d) and are taken into model training. Table 1 summarizes the symbols and abbreviations used in this paper.



(a) (b) (c) (d)

Figure 2. Sample time series feature selection in DTW alignment: (a) original features, (b) DTW realignment, (c) selected similar features, and (d) features selected for model training

Table 1. Symbols and abbreviations used in the paper

Symbol	Description
T^{ref}	Target time series
w_{ij}	Adaptive weight
W	Warping path
$D(T^{ref}, X^j)$	DTW distance
$f_{adp}(t)$	Proposed approach: forget gate
α	Learnable parameter
δ	Small quantity
γ	Shift smoothing parameter
$L_{D(T^{ref}, X^j)}$	Differentiable loss function
L_{MSE}	Loss from the AFG-BiLSTM

Algorithm 1. Feature selection with A-DTW
Input: Time series target T^{ref} and X^j $(1 \leq j \leq K)$
1) Compute adaptive weights given $W_{ij} = 1 + \left(\frac{|\sigma_{xj}-\sigma_{ti}|}{\sigma_{xj}+\sigma_{ti}+\epsilon}\right) .\log\left(1+\sigma_{xj}.\sigma_{ti}\right)$.
2) Compute warping path W by minimizing $D(i,j)$ using:

$$d(i,j) = \min \begin{cases} D(i-1,j) + w_{ij} * dist(t_i, x_j) \\ D(i,j-1) + w_{ij} * dist(t_i, x_j) \\ D(i-1,j-1) + 2 * w_{ij} * dist(t_i, x_j) \end{cases}$$

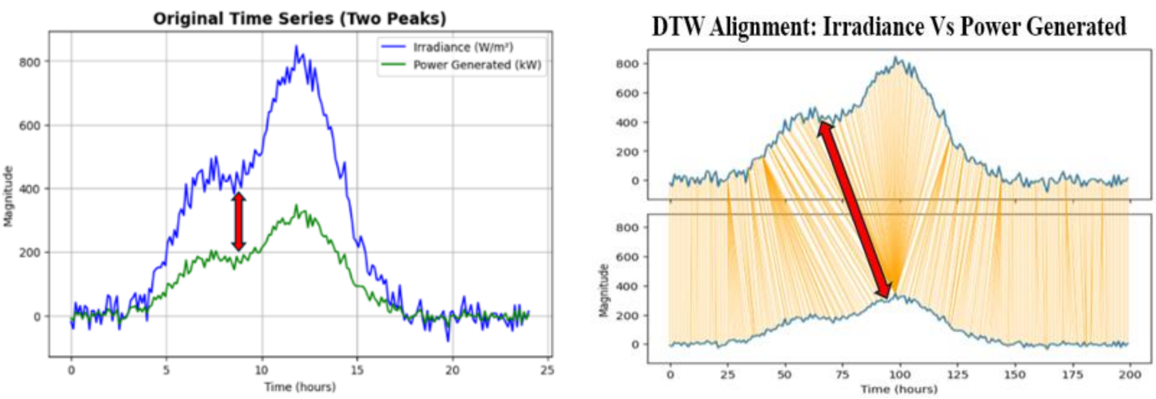
3) Compute $D(T^{ref}, X^j) = \sum_{(i,j)\in W} dist(t_i, x_j)\,; (i,j) \in W$.
4) Select features X^j, with minimal $D(T^{ref}, X^j)$ from step 3.

2.1.3. Need for wrapping in time series solar power (energy) prediction with variance entropy weight

Measuring similarity between two time series features is challenging because of factors like complexity, variability, and temporal nature. Time series data often have shifts in time, meaning the same points may appear at different timestamps. Because of the temporal nature of time series, the features may initially seem to be dissimilar, but close observation by stretching or compressing the time axes may really help in identifying the relationships between these temporal features. In scenarios of solar energy prediction, similarity analysis will fetch strong predictors. A-DTW with variance entropy weight aligns time sequences by compression or stretching for better analysis of the data.

Our proposed A-DTW approach uses variance entropy weight, as solar power generation is uncertain with most of the weather variables. If both variances are high, the weight term grows enormously, which is not stable for the best feature selection. Our proposed approach penalizes this explosive behavior with a log() term in the numerator, as shown in (4), making the model less sensitive to extreme variance.

The graph in Figure 3(a) shows the relationship between irradiance (blue curve) and power generated (green curve), indicating the lag phenomenon. The power curve shows a slight lag, which could reduce its similarity to irradiance. Figure 3(b) shows the realignment of the same lag points after applying wrapping, showing how the peaks could be correlated for data similarities. Also, the differences between these lags are minimized and made uniform.



(a) (b)

Figure 3. DTW alignment of irradiance and generated power: (a) before alignment and (b) after alignment

2.2. Proposed adaptive forget gate stacked BiLSTM (AFG-BiLSTM)

For time series modelling, recurrent neural networks (RNNs) and their variants have been extensively utilized. However, because of vanishing gradients, typical RNNs have problems with short-term memory. By preserving long-range dependencies, long-short term memory (LSTM) networks help to alleviate these problems. The LSTM uses three gates: forget (f_t), input (i_t), and output (o_t) given by (8)-(10).

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (8)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (9)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (10)$$

But many of the current research works on PV data showed low performance with LSTM, as they failed to analyze the past and present data dependencies of time series sequences, particularly when predicting power generated with many of the environmental predictor variables.

In the proposed study, in order to improve the learning of both past and current dependencies, we investigate stacked bidirectional LSTMs (BiLSTMs) to be trained with the features where many environmental variables are involved. Conventional LSTM networks handle input in a forward orientation, efficiently capturing previous dependencies while disregarding information that may become accessible in the sequence in the future. By adding a second LSTM layer that processes the sequence in reverse, bidirectional LSTMs (BiLSTMs) overcome this restriction and provide a more thorough comprehension of sequential dependencies. Furthermore, hierarchical feature extraction is made possible by stacking numerous BiLSTM layers, which enhances the expressiveness of the model.

Given an input sequence $X = [x_1, x_2, x_3, \dots \dots x_T]$, a BiLSTM computes hidden states (h_t) in both directions.

$$\vec{h}_t = LSTM(x_t, \vec{h}_{t-1}) \quad (11)$$

$$\overleftarrow{h}_t = LSTM(x_t, \overleftarrow{h}_{t+1}) \quad (12)$$

Where $\vec{h}_t$ and $\overleftarrow{h}_t$ represent forward and backward hidden states at time t. The final representation at each step is the concatenation of both, given by (13).

$$h_t = [\vec{h}_t, \overleftarrow{h}_t] \quad (13)$$

The proposed approach adds novelty to the stacked BiLSTM model by using a learnable, parametrized forget gate in the LSTM layers, which adapts to the data features selected from the A-DTW module. The conventional forget gate in LSTM controls the amount of previous memory retained or discarded using fixed learned weights and sigmoid activation functions. Its behavior remains generally uniform across input features and does not explicitly consider temporal feature correlation or redundancy. For example, because of multiple lag alignments of the features, in the A-DTW module, some redundant features like humidity and point humidity, and wind speed and wind flow may end up showing a minimal DTW distance and thereby be selected by the A-DTW feature selection module. With a conventional forget gate, these highly correlated or redundant environmental features may still be propagated through the memory states. The presence of redundant features can lead to model overfitting, adversely impacting model performance. Our proposed adaptive forget gate layer uses a learnable parameter (α) that adapts to the data features X^j, selected from the A-DTW module. This parameter penalizes the redundant features. The forget gate $f_{adp}(t)$ in our approach is given by (14):

$$f_{adp}(t) = \alpha\left(D(T^{ref}, X^j)\right) . \sigma(W_f x_t^j + U_f h_{t-1} + b_f) \quad (14)$$

with α as a learnable parameter that adapts to the data features X^j, selected from the DTW module. From (7), $D(T^{ref}, X^j)$ is the DTW distance between a feature (X^j) and the target time series (T^{ref}). The learnable parameter α converts the DTW distance into a weight using (15).

$$\alpha\left(\mathrm{D}(\mathrm{T^{ref}}, \mathrm{X^j})\right) = \frac{1}{1+e^{\mathrm{D}(\mathrm{T^{ref}}, \mathrm{X^j})}} \quad (15)$$

For $\mathrm{X^{j1}}, \mathrm{X^{j2}} \in \mathrm{X^j}$ two similar redundant features selected in the A-DTW module, they show nearly identical weights, i.e $\alpha\left(D(T^{ref}, X^{j1})\right) \approx \alpha\left(D(T^{ref}, X^{j2})\right)$ and are classified as redundant. We define the redundant feature discrimination condition (RFDC) as, for $\mathrm{X^{j1}}, \mathrm{X^{j2}} \in \mathrm{X^j}$ two redundant features:

$$|\alpha\left(\mathrm{D}(\mathrm{T^{ref}}, \mathrm{X^{j1}})\right) - \alpha\left(\mathrm{D}(\mathrm{T^{ref}}, \mathrm{X^{j2}})\right)| \leq \delta \quad (16)$$

where δ (<0.05) is a small quantity.

Unlike the conventional LSTM forget gate (8), which computes the memory retention coefficient solely from the current input and previous hidden state, the proposed adaptive forget gate, as shown in (14), explicitly incorporates the temporal similarity between each selected feature and the target time series through the adaptive coefficient α(D(T^{ref}, X^j)). Consequently, the proposed gate can be expressed as $f_{adp}(t)$ = α(D(Tref, Xj))f_t, indicating that it is a DTW-guided scaling of the conventional forget gate. When a feature exhibits a smaller A-DTW distance to the reference target, indicating stronger temporal correlation, α approaches unity, allowing the feature to retain most of its historical memory. Conversely, features with larger A-DTW distances receive smaller adaptive coefficients, thereby reducing their contribution to the cell state and preventing redundant or weakly correlated information from propagating through successive memory updates. The proposed BiLSTM forget gate is trained to discard(forget) one among these two features with minimal RFDC and maximum correlation between features X^j. The BiLSTM forget gate, which is shown in (14), iteratively trains itself on the error function given in (18). Figure 4 shows the architecture of our proposed model. Figure 5(a) shows one cell state of the AFG BiLSTM forget gate (AFG). Figure 5(b) shows the proposed A-DTW and AFG BiLSTM modules. The steps in the training process of AFG-BiLSTM are as shown in Algorithm 2. The correlation analysis as shown in Table 2 determines the strength of relationships between environmental variables and energy output (power generated). The table includes the analysis on two datasets. The table shows the correlation and A-DTW distances.

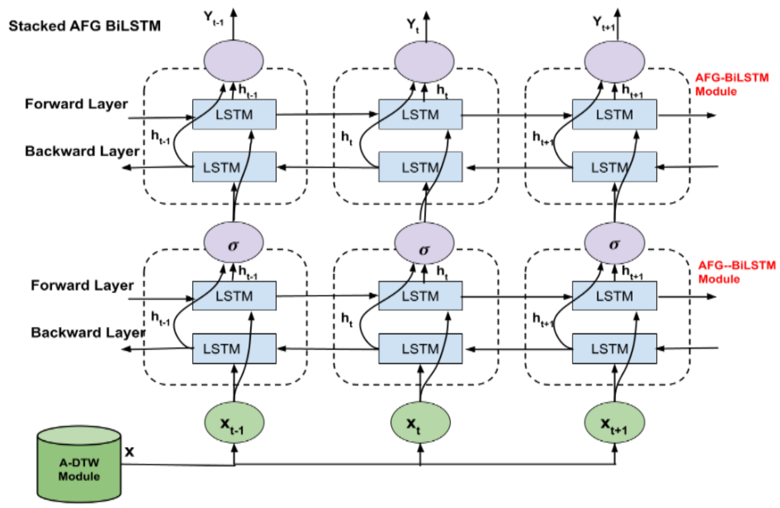
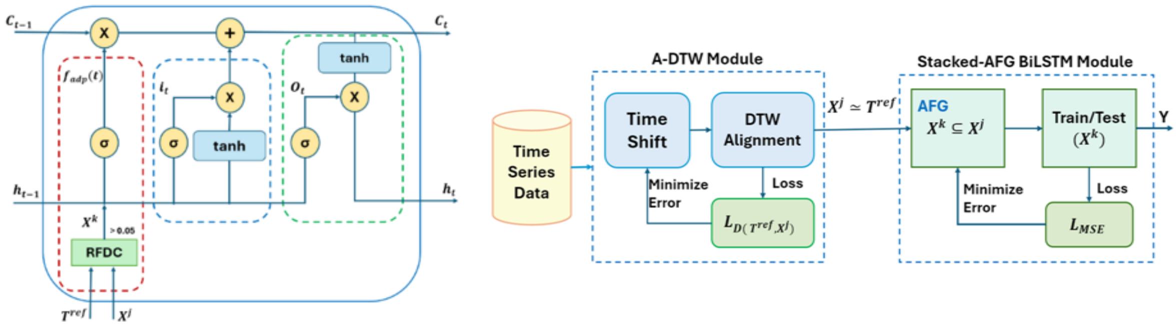


Figure 4. Proposed stacked BiLSTM architecture



(a) (b)

Figure 5. Proposed adaptive forget gate (AFG) BiLSTM architecture: (a) AFG BiLSTM forget-gate cell state and (b) proposed A-DTW and AFG BiLSTM modules

Algorithm 2. Model training: Stacked AFG-BiLSTM
Input: Selected features Xj from Algorithm 1
1. Train BiLSTM with adaptive forget gate given by fadp(t)
2. Calculate loss $L_{D(T^{ref},X^j)}$ from the A-DTW module in (17).
3. Calculate loss L_{MSE} from the AFG BiLSTM module in (18).
4. Calculate total loss, $\mathcal{L}_{total}$ from (19)
5. Repeat training to minimize the total loss

Table 2. Correlation Vs A-DTW distance on Dataset-1 and Dataset-2

Dataset	Feature	Correlation	Feature	A-DTW distance
Dataset-1: Target feature-power_generated	irradiance	0.98	irradiance	0.49
	air_temperature	0.74	humidity	0.83
	ambient_temperatur	0.57	wind_direction	1.72
	surface_temperature	0.53	wind_speed	1.81
	dew_temperature	0.50	air_temperature	2.01
	wind_speed	0.14	surface_temperature	2.14
	atmospheric_pressur	0.13	dew_temperature	2.20
	wind_direction	0.09	atmospheric_pressure	2.21
	humidity	-0.65	ambient_temperature	2.26
	vapour_presure	-0.91	vapour_presure	2.41
	Cloud_cover	-0.96	Cloud_cover	2.69
	relative_humidity	-0.97	relative_humidity	2.88
	cloud_precipitation	-0.98	cloud_precipitation	3.04
Dataset-2: target feature- power generated	Solar irradiance	0.97	Solar irradiance	0.62
	Diffuse solar radiation	0.68	Direct normal irradiance	0.97
	Direct normal irradiance	0.65	Diffuse solar radiation	0.97
	UV Radiation	0.62	UV radiation	0.97
	Wet bulb temperature	0.13	Relative humidity	1.50
	Wind gust	0.004	Atmospheric pressure	1.64
	Visibility	0.0017	Wind speed	1.64
	Dew point temperature	0.0015	Cloud cover (%)	1.64
	Ground surface temperature	0.0014	Air temperature	1.64
	Surface albedo	0.0011	Surface Albedo	1.65
	Precipitation rate	0.0009	Total sky cover	1.65
	Snow depth	0.0009	Heat index	1.65
	Total sky cover	0.0006	Wind gust	1.65
	Air temperature	0.0002	Dew point temperature	1.66
	Wind direction	0.0001	Feels-like Temperature	1.66
	Feels-like Temperature	-0.0005	Precipitation rate	1.66
	Heat index	-0.0014	Visibility	1.66
	Relative humidity	-0.0023	Wind direction	1.66
	Wind speed	-0.0028	Wet bulb temperature	1.66
	Cloud cover (%)	-0.0035	Snow depth	1.67
	Atmospheric pressure	-0.0059	Ground surface temperature	1.67
	Solar Zenith angle	-0.6081	Solar Zenith angle	1.9

3. RESULTS AND DISCUSSION

All experiments are run on the deep learning framework PyTorch, which allowed flexibility in designing, training, and evaluating models. The implementation was carried out in Python, with pandas and Scikit-learn libraries for data preprocessing and evaluation. Experiments were conducted on a computer equipped with an NVIDIA GPU and CUDA support to accelerate training.

The model configurations and their hyperparameters are set as follows: BiLSTM with one BiLSTM layer with 128 hidden units, regular forget gate, dropout 0.3, RMSprop optimizer with fixed 0.001 learning rate, and MSE loss; Stacked BiLSTM uses 128 units per layer, 3 layers of BiLSTM, the usual forget gate, dropout 0.3, RMSprop with 0.001 learning rate. The proposed AFG-BiLSTM is an adaptively optimized stacked BiLSTM with a forget gate, tuned to 128 units, 0.3 dropout, fine-tuned with the Adam optimizer, and trained for 100 epochs. A 70-30 train-test split was adopted for each dataset, with the first 70% of the time-series observations used for training and the remaining 30% used for testing.

The mean squared error (MSE) metric was used to compare the performance of the models. The performance of the Ada-BiLSTM method with AFG was studied through comparison against four datasets as discussed in section 2. The correlations and the A-DTW distance are extensively studied across the target and the feature variables for all four datasets. The feature selection is studied under two scenarios: feature selection with correlations; feature selection with A-DTW distance.

The experimental results of correlations and A-DTW distances on each dataset are recorded in Table 2 highlights the following insights: in Dataset 1, some of the essential features like ‘humidity’, ‘cloud cover’, ‘wind speed’, and ‘wind direction’ were observed to be very weakly correlated with the target power generation and are not among the top-5 strongly correlated features. But upon applying A-DTW alignment, these features are in the Top-5, highlighting the performance of our novel feature selection approach. Whereas in dataset 2, essential features like ‘air temperature’, ‘cloud cover’, ‘relative humidity’, ‘wind speed', and 'wind direction' look very weakly correlated with the target (power generated) and are not in the Top-10 strongly correlated features. Applying A-DTW alignment on the dataset, these features were in the Top-10. These insights show that, for time series data, in order to understand the features that most impact the target solar energy, it is essential to study the warping alignments along with correlations for key feature selection. Errors may occur while arriving at the optimal warping path from among the available time lags. Our proposed approach uses a smooth, differentiable loss function to study the loss of the A-DTW module and is given by (17).

$$L_{D(T^{ref},X^j)} = \underbrace{min}_{w\in\mathcal{P}} \sum_{(i,j)\in w} exp\,(-\frac{D(T^{ref},X^j)}{\gamma})\, D(T^{ref},X^j) \tag{17}$$

Where γ is the shift smoothing parameter, w∈P is a warping path from the set of all possible warping paths P. The loss from the AFG-BiLSTM module is given by (18), where we use MSE to judge prediction accuracy. The total loss of the approach is expressed by (19).

$$L_{MSE} = \frac{1}{T} \sum_{t=1}^{T} (t_{ref} - \hat{t}_{ref})^2\,;\ for\ t_{ref} \in T^{ref} \tag{18}$$

$$\mathcal{L}_{total} = L_{D(T^{ref},X^j)} + L_{MSE} \tag{19}$$

Table 3 shows the loss of the proposed model as per (19), which is the aggregated loss from the A-DTW and AFF-BiLSTM modules. Table 4 presents the training MSE loss of the baseline models and the proposed model, and these results showed our proposed A-DTW feature selection and AFG redundant feature elimination reduced validation error in comparison to the chosen traditional models, emphasizing the importance of relevant features for solar energy modelling.

Table 3. Loss of the proposed model as per (19)

Dataset	A-DTW loss ($L_{D(T^{ref},X^j)}$)	AFG-BiLSTM loss (L_{MSE})	Total loss ($\mathcal{L}_{total}$)
Dataset-1	2.722e-05	0.0035	0.0035
Dataset-2	2.569e-06	0.0042	0.0042
Dataset-3	2.011e-02	0.0061	0.0061
Dataset-4	5.946e-03	0.1052	0.1111

Table 4. Loss of the proposed model vs baseline models

Dataset	Loss of BiLSTM (L_{MSE})	Loss of stacked BiLSTM (L_{MSE})	Proposed model-total loss
Dataset-1	0.086	0.069	0.0035
Dataset-2	0.096	0.078	0.0042
Dataset-3	0.067	0.069	0.0061
Dataset-4	0.185	0.167	0.1111

Figure 6 shows the training loss of the proposed approach in comparison to traditional BiLSTM and stacked BiLSTM models, where these baseline models were trained with all the dataset features. As observed from Figure 6, the proposed model loss was huge across dataset 4. This was due to the features in the dataset. A close observation of Dataset 4 features showed that many features were not relevant to energy prediction. Also, even the correlated and proposed A-DTW alignment technique did not pick the top 10 features to be very relevant compared to other datasets, as observed from Table 2. This could be the reason for the huge loss of the approach with Dataset 4. These comparisons show the importance of environmental feature selection for better solar power prediction.

The superior performance of the proposed A-DTW and AFG-BiLSTM framework stems from the complementary integration of variance-aware temporal feature selection and adaptive memory learning. While A-DTW identifies temporally relevant features, the Adaptive Forget Gate selectively retains informative temporal patterns by eliminating further redundant features from A-DTW, enabling the BiLSTM to learn more discriminative representations and achieve higher forecasting accuracy than the conventional stacked BiLSTM. By removing redundant and correlated features, A-DTW reduces the input dimensionality and mitigates overfitting.

Figure 7 shows the MSE comparisons of the methods on the actual PV power generated on the four datasets, on a common day of 24 hours. Across the four datasets, the proposed model yielded the smallest mean squared error (MSE), on average. In all the experiments, the proposed model decreased error by (94.9%) on Dataset-1 (0.069, 0.0035), by (94.6%) on Dataset-2 (0.078, 0.0042), and by (90.9%) on Dataset-3 (0.069, 0.0061) when compared to the stacked BiLSTM. The most difficult observation is from Dataset-4, where the error of the proposed approach increases significantly. This was due to the features in the dataset. These comparisons show the importance of environmental feature selection for better solar power prediction. Even on Dataset 4, the proposed model is, however, optimal with a (33.5%) decrease compared to the stacked BiLSTM (0.167,0.1111), Taken together, the findings show the proposed approach generalizes well across all the datasets considered.

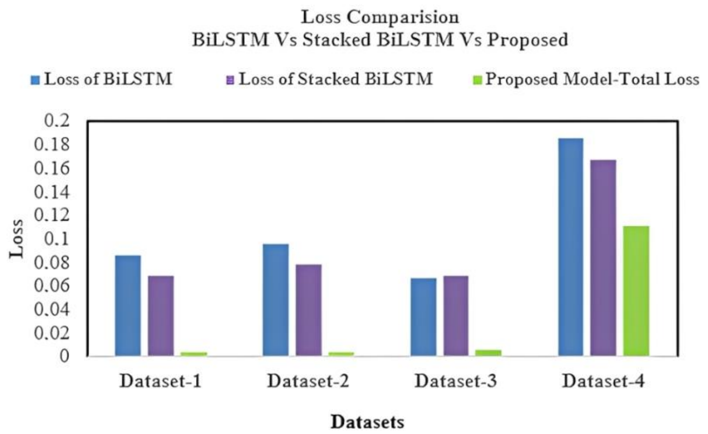


Figure 6. MSE loss of proposed model compared to BiLSTM and Stacked BiLSTM

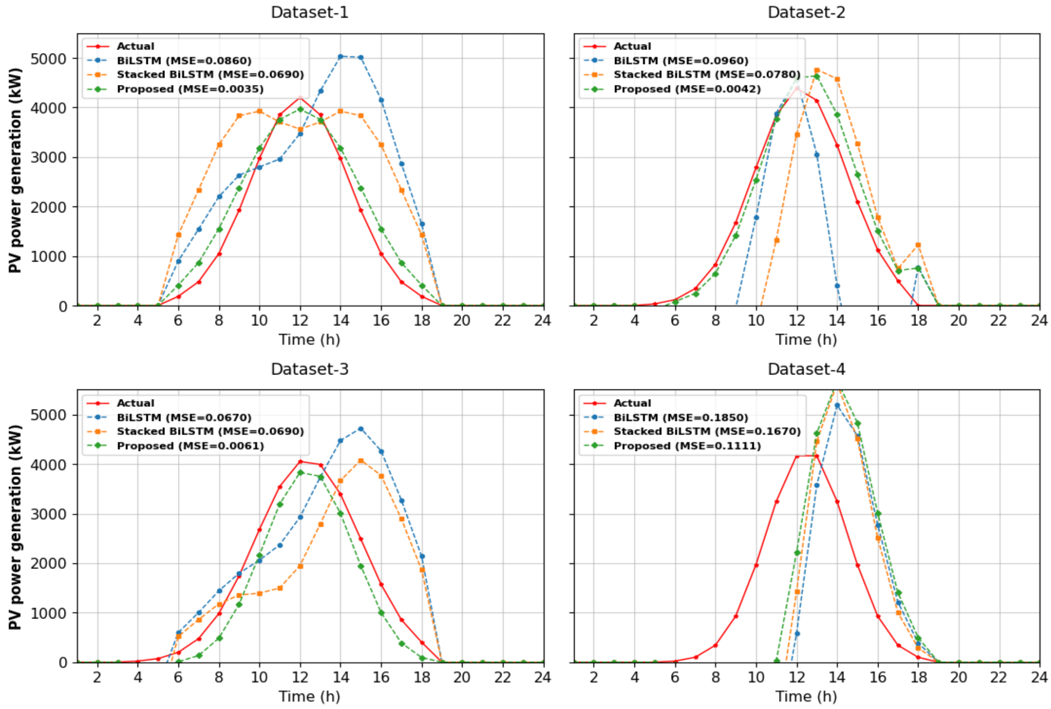


Figure 7. Validation loss of models in comparison to the baseline on four datasets

3.1. Limitations of the proposed study

Although the proposed AFG-BiLSTM demonstrates superior forecasting accuracy across four benchmark datasets, several limitations should be acknowledged. First, the framework has been validated using offline experimental datasets and has not yet been deployed in real-time photovoltaic systems or hardware platforms such as smart converters, embedded controllers, or energy management systems. Second, the comparative analysis primarily focuses on LSTM-based architectures, as the objective of this study is to investigate the effectiveness of bidirectional temporal dependency learning using the proposed AFG-BiLSTM framework. Consequently, recent Transformer and attention-based forecasting models were not included in the comparative evaluation. Their integration and benchmarking will be considered in future work to further assess the generalization capability of the proposed framework. Although consistent improvements were observed across multiple datasets, further validation on geographically diverse solar plants and unseen operating conditions is required to establish broader generalization.

4. CONCLUSION

A close observation of Solar Energy datasets from many PV plant organizations revealed many intrinsic environmental variables that are too asynchronous in nature. Whether all such features are essential for energy prediction systems is the point of motivation for this research. Also, many forecasting models end up learning redundant features that might seem useful, and thereby the test performance is bitter in overfitting, thereby stating the need for intelligent feature selection algorithms guided by both correlation and temporal alignment metrics. In this research, we developed a framework to address asynchronous features with an adaptive dynamic time warping (A-DWT) feature selection approach and model overfitting using stacked BiLSTM with AFG-BiLSTM.

Experiments were conducted on four datasets with varied environmental asynchronous features, and the results of the proposed model were compared with baseline BiLSTM and Stacked BiLSTM models. In all the experiments, the proposed model showed decreased error by (94.9%) on Dataset-1 (0.069, 0.0035), by (94.6%) on Dataset-2 (0.078, 0.0042), and by (90.9%) on Dataset-3 (0.069, 0.0061) when compared to stacked BiLSTM. The loss of the proposed method was between (0.3%) and (11%) when trained on the four datasets. With the proposed approach and the experimental results conducted, we support the importance of adaptive feature selection methods for solar energy modelling it still has limitations.

The proposed framework is evaluated using specific PV environmental datasets, which may limit its direct generalization to different geographical conditions and diverse sensor configurations without further validation of this heterogeneity. As a future scope, the proposed approach may be applied to more heterogeneous datasets in terms of geographic and climatic regions, thus testing its flexibility. It could be further improved by integrating hybrid feature selection frameworks using A-DTW to increase the relevance and strength of features. This research conducted experiments on feature selection using A-DTW and an AFG-based BiLSTM framework. Future work will extend this approach to broader renewable energy applications, including solar forecasting, grid stability analysis, and energy storage scheduling, with integration into smart converters, renewable energy control systems, and power electronic converters for load balancing and real-time optimization. Future work will extend the proposed framework to multi-step forecasting, evaluate its scalability on large-scale PV farms with heterogeneous sensor networks, and compare its performance with Transformer-based temporal alignment and multi-head attention architectures for capturing long-range temporal dependencies. In addition, real-time deployment on edge devices, smart converters, and renewable energy control systems will be explored to support intelligent energy management and load balancing.

FUNDING INFORMATION

This research did not receive any specific grant from a funding agency in the public, commercial, or not-for-profit sector.

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
Are Sambasiva Rao	✓	✓	✓			✓			✓	✓	✓			
Kunada Dhana Sree Devi				✓		✓			✓	✓				

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 no conflicts of interest.

DATA AVAILABILITY

The datasets and materials used and analyzed during the current study are made available and can be accessed using the link below: Datasets: https://github.com/dhanasreek/Solar-Energy-Datasets.

REFERENCES

[1] H. Mauladdawilah, M. Balfaqih, Z. Balfagih, M. del C. Pegalajar, and E. J. Gago, "Deep feature selection of meteorological variables for LSTM-Based PV power forecasting in high-dimensional time-series data," *Algorithms*, vol. 18, no. 8, p. 496, Aug. 2025, doi: 10.3390/a18080496.

[2] S. Yang and Y. Luo, "Short-term photovoltaic power prediction based on RF-SGMD-GWO-BiLSTM hybrid models," *Energy*, vol. 316, p. 134545, Feb. 2025, doi: 10.1016/j.energy.2025.134545.

[3] M. Liu, X. Wang, and Z. Zhong, "Ultra-short-term photovoltaic power prediction based on BiLSTM with wavelet decomposition and dual attention mechanism," *Electronics*, vol. 14, no. 2, p. 306, Jan. 2025, doi: 10.3390/electronics14020306.

[4] W. Guo, S. Liu, L. Weng, and X. Liang, "Power grid load forecasting using a CNN-LSTM network based on a multi-modal attention mechanism," *Applied Sciences*, vol. 15, no. 5, p. 2435, Feb. 2025, doi: 10.3390/app15052435.

[5] D. Zhou, Y. Liu, X. Wang, F. Wang, and Y. Jia, "Combined ultra-short-term photovoltaic power prediction based on CEEMDAN decomposition and RIME optimized AM-TCN-BiLSTM," *Energy*, vol. 318, 2025, doi: 10.1016/j.energy.2025.134847.

[6] H. Lin, L. Gao, M. Cui, H. Liu, C. Li, and M. Yu, "Short-term distributed photovoltaic power prediction based on temporal self-attention mechanism and advanced signal decomposition techniques with feature fusion," *Energy*, vol. 315, p. 134395, Jan. 2025, doi: 10.1016/j.energy.2025.134395.

[7] A. Alorf and M. U. G. Khan, "Solar irradiance forecasting using temporal fusion transformers," *International Journal of Energy Research*, vol. 2025, no. 1, Jan. 2025, doi: 10.1155/er/3534500.

[8] Y. Wang *et al.*, "A wind and solar power prediction method based on temporal convolutional network–attention–long short-term memory transfer learning and sensitive meteorological features," *Applied Sciences*, vol. 15, no. 3, p. 1636, Feb. 2025, doi: 10.3390/app15031636.

[9] H. Tang, F. Kang, X. Li, and Y. Sun, "Short-term photovoltaic power prediction model based on feature construction and improved transformer," *Energy*, vol. 320, p. 135213, Apr. 2025, doi: 10.1016/j.energy.2025.135213.

[10] G. M. Tina, C. Ventura, S. Ferlito, and S. De Vito, "A state-of-art-review on machine-learning based methods for PV," *Applied Sciences*, vol. 11, no. 16, p. 7550, Aug. 2021, doi: 10.3390/app11167550.

[11] J.-P. Lai, Y.-M. Chang, C.-H. Chen, and P.-F. Pai, "A survey of machine learning models in renewable energy predictions," *Applied Sciences*, vol. 10, no. 17, p. 5975, Aug. 2020, doi: 10.3390/app10175975.

[12] B. Tai *et al.*, "Power prediction of photovoltaic power generation based on LSTM model with additive Attention mechanism," in *2023 7th International Conference on Smart Grid and Smart Cities (ICSGSC)*, Sep. 2023, pp. 474–480, doi: 10.1109/ICSGSC59580.2023.10319231.

[13] 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.

[14] 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.

[15] C. Stoean *et al.*, "Metaheuristic-based hyperparameter tuning for recurrent deep learning: application to the prediction of solar energy generation," *Axioms*, vol. 12, no. 3, p. 266, Mar. 2023, doi: 10.3390/axioms12030266.

[16] M. H. Sulaiman and Z. Mustaffa, "Forecasting solar power generation using evolutionary mating algorithm-deep neural networks," *Energy and AI*, vol. 16, p. 100371, May 2024, doi: 10.1016/j.egyai.2024.100371.

[17] C. Saigustia and P. Pijarski, "Time series analysis and forecasting of solar generation in Spain using extreme gradient boosting: a machine learning approach," *Energies*, vol. 16, no. 22, p. 7618, Nov. 2023, doi: 10.3390/en16227618.

[18] C.-H. Loh, Y.-C. Chen, C.-T. Su, and H.-Y. Su, "Establishing lightweight and robust prediction models for solar power forecasting using numerical–categorical radial basis function deep neural networks," *Applied Sciences*, vol. 14, no. 22, p. 10625, Nov. 2024, doi: 10.3390/app142210625.

[19] A. A. Alguhi and A. M. Al-Shaalan, "LSTM-based prediction of solar irradiance and wind speed for renewable energy systems," *Energies*, vol. 18, no. 17, p. 4594, Aug. 2025, doi: 10.3390/en18174594.

[20] L. A. Dao, L. Ferrarini, and D. La Carrubba, "Improving solar and PV power prediction with ensemble methods," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 12829–12834, 2020, doi: 10.1016/j.ifacol.2020.12.1974.

[21] X. Chang, W. Li, and A. Y. Zomaya, "A lightweight short-term photovoltaic power prediction for edge computing," *IEEE Transactions on Green Communications and Networking*, vol. 4, no. 4, pp. 946–955, Dec. 2020, doi: 10.1109/TGCN.2020.2996234.

[22] M. K. Behera, I. Majumder, and N. Nayak, "Solar photovoltaic power forecasting using optimized modified extreme learning machine technique," *Engineering Science and Technology, an International Journal*, vol. 21, no. 3, pp. 428–438, Jun. 2018, doi: 10.1016/j.jestch.2018.04.013.

[23] Y. Niu, J. Wang, Z. Zhang, Y. Cao, P. Yan, and Z. Li, "Amplify seasonality, prioritize meteorological: Strengthening seasonal correlation in photovoltaic forecasting with dual-layer hierarchical attention," *Applied Energy*, vol. 394, p. 126104, Sep. 2025, doi: 10.1016/j.apenergy.2025.126104.

[24] J. Qu, Z. Qian, and Y. Pei, "Day-ahead hourly photovoltaic power forecasting using attention-based CNN-LSTM neural network embedded with multiple relevant and target variables prediction pattern," *Energy*, vol. 232, p. 120996, Oct. 2021, doi: 10.1016/j.energy.2021.120996.

[25] S. Huang, Q. Zhou, J. Shen, H. Zhou, and B. Yong, "Multistage spatio-temporal attention network based on NODE for short-term PV power forecasting," *Energy*, vol. 290, p. 130308, Mar. 2024, doi: 10.1016/j.energy.2024.130308.

[26] J. Gan *et al.*, "Improving short-term photovoltaic power generation forecasting with a bidirectional temporal convolutional network enhanced by temporal bottlenecks and attention mechanisms," *Electronics*, vol. 14, no. 2, p. 214, Jan. 2025, doi: 10.3390/electronics14020214.

[27] A. Gensler, J. Henze, B. Sick, and N. Raabe, “Deep learning for solar power forecasting — an approach using autoencoder and LSTM neural networks,” in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct. 2016, pp. 002858–002865, doi: 10.1109/SMC.2016.7844673.
[28] D. Yu, W. Choi, M. Kim, and L. Liu, “Forecasting day-ahead hourly photovoltaic power generation using convolutional self-attention based long short-term memory,” *Energies*, vol. 13, no. 15, p. 4017, Aug. 2020, doi: 10.3390/en13154017.
[29] M. Awais *et al.*, “Short-term photovoltaic energy generation for solar powered high efficiency irrigation systems using LSTM with Spatio-temporal attention mechanism,” *Scientific Reports*, vol. 14, no. 1, p. 10042, May 2024, doi: 10.1038/s41598-024-60672-9.
[30] Y. Zhong, T. He, and Z. Mao, “Enhanced solar power prediction using attention-based DiPLS-BiLSTM model,” *Electronics*, vol. 13, no. 23, p. 4815, Dec. 2024, doi: 10.3390/electronics13234815.
[31] W. Chen, H. He, J. Liu, J. Yang, K. Zhang, and D. Luo, “Photovoltaic power prediction based on sliced bidirectional long short term memory and attention mechanism,” *Frontiers in Energy Research*, vol. 11, Mar. 2023, doi: 10.3389/fenrg.2023.1123558.
[32] M. Yu *et al.*, “Short-term photovoltaic power point-interval forecasting based on double-layer decomposition and WOA-BiLSTM-Attention and considering weather classification,” *Energy*, vol. 275, p. 127348, Jul. 2023, doi: 10.1016/j.energy.2023.127348.
[33] Q. Xie, L. Ma, Y. Liu, Q. Fu, Z. Shen, and X. Wang, “An improved SSA-BiLSTM-based short-term irradiance prediction model via sky images feature extraction,” *Renewable Energy*, vol. 219, p. 119507, Dec. 2023, doi: 10.1016/j.renene.2023.119507.
[34] B. Gu, X. Li, F. Xu, X. Yang, F. Wang, and P. Wang, “Forecasting and uncertainty analysis of day-ahead photovoltaic power based on WT-CNN-BiLSTM-AM-GMM,” *Sustainability*, vol. 15, no. 8, p. 6538, Apr. 2023, doi: 10.3390/su15086538.
[35] W. Liu and Z. Mao, “Short-term photovoltaic power forecasting with feature extraction and attention mechanisms,” *Renewable Energy*, vol. 226, p. 120437, May 2024, doi: 10.1016/j.renene.2024.120437.
[36] E. Sarmas, N. Dimitropoulos, V. Marinakis, Z. Mylona, and H. Doukas, “Transfer learning strategies for solar power forecasting under data scarcity,” *Scientific Reports*, vol. 12, no. 1, p. 14643, Aug. 2022, doi: 10.1038/s41598-022-18516-x.
[37] Y. Gao, Z. Hu, S. Shi, W.-A. Chen, and M. Liu, “Adversarial discriminative domain adaptation for solar radiation prediction: A cross-regional study for zero-label transfer learning in Japan,” *Applied Energy*, vol. 359, p. 122685, Apr. 2024, doi: 10.1016/j.apenergy.2024.122685.
[38] S. Shan *et al.*, “SolarFM: Spatio-temporal solar irradiance forecasting based on time series foundation model,” *Applied Energy*, vol. 410, p. 127532, May 2026, doi: 10.1016/j.apenergy.2026.127532.
[39] K. Saidani, N. Essaddi, and M. Besbes, “Harnessing Prophet model with rigorous cross validation for long term solar insolation prediction,” *Results in Engineering*, vol. 28, p. 107915, Dec. 2025, doi: 10.1016/j.rineng.2025.107915.
[40] E. Samara *et al.*, “Dynamic time warping as a means of assessing solar wind time series,” *The Astrophysical Journal*, vol. 927, no. 2, p. 187, Mar. 2022, doi: 10.3847/1538-4357/ac4af6.
[41] T. Kim and J. Hong, “Optimized model selection for wind farm power forecasting using dynamic time warping,” in *Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing*, Mar. 2025, pp. 738–745, doi: 10.1145/3672608.3707947.

BIOGRAPHIES OF AUTHORS

Are Sambasiva Rao is a research scholar in the Department of Computer Science and Engineering at GITAM Deemed to be University, Hyderabad, India. He received his B.Tech. degree from Jawaharlal Nehru Technological University (JNTU), Hyderabad, and his M.E. degree from Osmania University, Hyderabad. His research interests include machine learning applications in photovoltaic (PV) power forecasting and deep learning. He can be contacted at email: sare2@gitam.in or sambasiva.ee@gmail.com.

Kunada Dhana Sree Devi is an associate professor in the Department of Artificial Intelligence and Data Science (AI & DS) at GITAM Deemed to be University, Hyderabad, India. She has published more than 40 technical papers in international journals and conference proceedings, as well as book chapters. She is associated with several professional organizations, including CSI, ISTE, IAENG, and IEEE. Her research interests include machine learning, deep learning, natural language processing, data mining, database security, and network security. She can be contacted at email: dkunada@gitam.edu.