
A New Hybrid Short Term Solar Irradiation Forecasting Method Based on CEEMDAN Decomposition Approach and BiLSTM Deep Learning Network with Grid Search Algorithm

Anuj Gupta^{1,*}, Sharad Sharma¹ and Sumit Saroha²

¹*Maharishi Markandeshwar (Deemed to be University), Mullana-Ambala, Haryana, India*

²*Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India*
E-mail: annu11gupta@gmail.com

**Corresponding Author*

Received 29 July 2022; Accepted 07 November 2022;
Publication 16 May 2023

Abstract

An accurate and efficient forecasting of solar energy is necessary for managing the electricity generation and distribution in today's electricity supply system. However, due to its random character in its time series, accurate forecasting of solar irradiation is a difficult task; but it is important for grid management, scheduling and its balancing. To fully utilize the solar energy in order to balance the generation and consumption, this paper proposed an ensemble approach using CEEMDAN-BiLSTM combination to forecast short term solar irradiation. In this, Complete Ensemble Empirical Mode Decomposition with adaptive noise (CEEMDAN) extract the inherent characteristics of time series data by decomposing it into low and high frequency Intrinsic Mode Functions (IMF's) and Bidirectional Long Short

Distributed Generation & Alternative Energy Journal, Vol. 38.4, 1073–1118.

doi: 10.13052/dgaej2156-3306.3842

© 2023 River Publishers

Term Memory (BiLSTM) used as a forecasting tool to forecast the solar Global Horizontal Irradiance (GHI). Furthermore, using extensive experimental analysis, the research minimizes the number of IMF's by integrating the CEEMDAN decomposed component (IMF1–IMF14) in order to increase the prediction accuracy. Then, for each IMF subseries, the trained standalone BiLSTM network are assigned to carry out the forecasting. In last stage, the forecasted results of each BiLSTM network are aggregate to compile final results. Two year data (2012–13) of Delhi, India from National Solar Radiation Database (NSRDB) has been used for training while one year data (2014) used for testing purpose for the same location. The proposed model performance is measured in terms of root mean square error (RMSE), mean absolute percentage error (MAPE), Correlation coefficient (R^2) and forecast skill (FS). For the comparative analysis of proposed model, several others models: persistence model, unidirectional deep learning models: long short term memory (LSTM), gated recurrent unit (GRU), BiLSTM and two CEEMDAN based BiLSTM models are developed. The proposed model achieved lowest annual average RMSE (18.86 W/m², 22.24 W/m², 26.25 W/m²) and MAPE (2.19%, 4.81%, 6.77%) among the other developed models for 1-hr, 2-hr and 3-hr ahead solar GHI forecasting respectively. The maximum correlation coefficient (R^2) obtained by the proposed model is 96.4 for 1-hr ahead respectively; on the other hand, forecast skill (%) of 89% with reference to benchmark model. Various test such as: Diebold Mariano Hypothesis test (DMH) and directional change in forecasting (DC) are used to analyze the sensitivity with reference to the difference in forecasted and observed value.

Keywords: Deep learning network, complete ensemble EMD with adaptive noise, gate recurrent unit, long short term memory, bidirectional long short term memory, Diebold Mariano Hypothesis test, directional change in forecasting, hyper parameters.

1 Introduction

Because of the greenhouse effect, pollution and the depletion of natural resources, it is now more vital than ever to use renewable energy resources (RES) that do not pollute the environment and free to use for electricity generation. Among renewable energy resources, solar energy is one of the most popular energy resources for generating electricity with zero carbon emission and its market has grown significantly in recent decades due to its

long-term viability and support [1, 2]. Almost every year, the earth's surface receives around 1.5×10^{18} KWh/area of solar energy which is nearly ten times the current global usage. Among all Asian countries, China got the highest annual average daily global solar radiation (20.2 MJ/(m².d)) while India just (18 MJ/(m².d)). The renewable energy sector in India, as an example of emerging countries, has grown at an exponential rate during the last two decades. India has even established a special ministry for RES; Ministry of New and Renewable Energy (MNRE), with a goal of generating 175 GW of energy from RES by the end of 2022; with 100 GW from solar alone [2]. According to the International Energy Agency (IEA), the overall capacity of photovoltaic installations will reach 1700 GW by 2030. However, according to the world energy states report, this power capacity increased from 8 GW in 2007 to 402 GW in 2017 [3]. Furthermore, according to several studies, the power grid will be completely functioning on the renewable energy source (RES) by the end of 2050 [4]. As we know electricity produced by the photovoltaic power plant is directly proportional to the Global Horizontal Irradiance (GHI) falling on earth surface [3, 4]. However, due to the changes of weather condition, the intensity of GHI is unstable in which directly impact the photovoltaic power plant output [5, 6]. This will affect the reliability of photovoltaic power plant. Hence, in order to make it more reliable there is requirement of highly accurate solar GHI forecasting model.

1.1 Literature Review

Over the past two decades, a great deal of efforts has gone into forecasting solar irradiance with positive outcomes. Different time horizon forecasting models have already been developed in past. Forecasting models can be divided into three types based on time horizon: short term, midterm, and long term forecasting. Short-term forecasting is defined as a look ahead of one week or less, and it is widely used to assist power systems. Midterm forecasting refers to one-week to one-month forecasting that is used for mid-term electricity dispatch. Finally, long-term forecasting analyses power generation, transmission, and distribution forecasts for one month to one year.

As per existing literature, numbers of forecasted models were developed with an aim to increase the forecasting accuracy. As per reference [7–10], the solar irradiance forecasting technologies are divided into five categories: (1) Persistence approach (2) Physical method (3) Machine learning approach (4) statistical method (5) Hybrid approach. The persistence approach is also known as naïve predictor in which calculate value of solar irradiance at time

t is considered as the forecasted value at time $t + h$ where, h represents the forecast interval. This model is generally used to make a comparison with the prediction results of proposed model. The author of [11] suggested that the forecasting results of proposed model need to be compared with Persistence model. On the other hand, the physical model used meteorological and geographical parameter to forecast the GHI instead of time series data and set up a mathematical relation between a meteorological data and forecasted GHI. Due to complexity, less precisions and high cost these models are not popular. European center for medium range weather forecast (ECMWF) and weather research and forecasting (WRF) are the two main methods in physical approach to forecast the atmospheric and operational research [12–14]. To forecast vertical irradiance and design sky radiance using irregular atmospheric condition [15, 16] developed a new unified framework of radiance design (UFRD) which chosen the impact of clouds field and aerosol parameters. The UFRD is based on theory and improves the accuracy of radiance and irradiance calculations. The statistical methods such as: Gaussian Progress Regression (GPR) [17] autoregressive integrated moving average (ARIMA) [18] Seasonal autoregressive integrated moving average (SARIMA) [19] Multilinear regression (MLR) [20] improve forecasting accuracy and set up a mathematical relation between meteorological data and global horizontal irradiance; but, a robustness is weak when a lack of correlation is arisen between input data and irradiance [21]. The author of [18] used the ARIMA model to anticipate sun radiation on a daily basis. In continuation of this, SARIMA model was developed by the author [19] to forecast solar radiation using Phillips-Perron test to identify the suitable model learning parameters. But, due to incomplete data, sparse and lack of correlation these models not provide a satisfactory result for the highly variable non-stationary time series data.

The learning based models such as; artificial neural network (ANN) [21, 22], elman neural network [23] and support vector machine [24], multi layer perceptron [25], Extreme learning Machine [26] have a capability to learn itself and reduce the gap between forecasted data and measured data; but, due to uncertain behavior of global horizontal irradiance, sometimes single learning model stuck in local minima and not perform efficiently. Several academics: Gupta et al., Singla et al., Voyant et al. have already published survey studied on machine learning approaches. So, one strategy is to create a hybrid model to improve the accuracy of solar irradiation estimation. The combination of a data decomposition technique and a forecasting model is one of the mostly used hybrid models.

On the other hand, optimization methods are used to optimize the learning parameter and to enhance the accuracy of a forecasted model. Various authors use different-2 optimization algorithm: particle swarm optimization, genetic algorithm, fuzzy logic, whale algorithm, sine cosine algorithm etc. In recent times, the author [58] used Bat and whale algorithm to optimize the hyperparameter of SVM model. In continuation of this, Genetic algorithm used by the author [59, 60] to select the learning parameters of LSTM and BiLSTM deep learning model in a particular search band.

Apart from this, In literature various data decomposition techniques used as a preprocessing step to decompose the irradiance data, clean up and define the input data according to the specifications. The self organizing map (SOM), wavelet transforms (WT), empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), normalization, Kalman filter, principal component analysis (PCA) are often used in solar irradiance forecasting. These tools not only extract the hidden characteristics of time series data but also remove the trends, noise and randomness. The authors [27] decomposed the clear sky index data using EMD and EEMD and then utilized auto regressive (AR) and ANN models to estimate irradiance. The experimental results shows that when hybrid model compared to a single AR and ANN model, the hybrid model improved prediction performance even more; on the other hand, the author of [28] implement a unique clustering hybrid learning approach which use EEMD for preprocessing. Support vector regression for ensemble learning approach and k-mean used for clustering analysis. The result of developed hybrid model is much better than standalone model. Likewise, the EEMD and SOM-back propagation (BP) network were combined by the author of [29] to forecast the solar irradiance. EEMD decompose the input data and decomposed subseries used as input of SOM-BP network. The output of each SOM-BP networks is aggregate to form the output while the author of [30] used a wavelet decomposition (WD) technique as a preprocessing technique and support vector machine used as a machine learning model. Likewise, the author of [31] used a combination of WD and ANN. The time series decomposed by the WD were modeled by the neural network model.

In addition, deep learning emerged as a powerful technique to forecast the solar GHI and its performance is much better than conventional models in all aspects. In literature a number of researchers used deep learning technique with preprocessing strategy to enhance the accuracy of forecasting model. The author of [32] use long short term memory (LSTM) network to forecast the global horizontal irradiance in which weather data was used as an input

parameter of the LSTM network. The performance of the LSTM network is much better than the conventional model like a BPNN, linear regression model in terms of RMSE. A hybrid model of LSTM and gradient boosting algorithm was implemented by the author of [33] to prevent the situation of over fitting and proposed model is compared with naïve predictor and support vector regression model. The performance of ensemble approach shows that proposed model is effective and excellent in terms of RMSE. The authors of [34] develop an ensemble approach to forecast the solar irradiance using a combination of CNN and LSTM. The historical properties of the input data were acquired by using an LSTM network and the geographical data were obtained using CNN. The forecasting of short term solar irradiance was performed by the author [35] using CNN network which use sky images as input and proposed model performance is measured in terms of RMSE, MBE and forecast skill score. Further, a short term photovoltaic power forecasting was performed by author [36] using a hybrid approach of residual network and convolution neural network. To divide the input dataset into trends, cyclical, and random components, the variational mode decomposition (VMD) was utilized as a preprocessing method. The author of [37] developed an ensemble approach of WT with LSTM model to solar irradiance forecasting for 1-hr to 1-day ahead. The result shows that WT improve forecasting accuracy likewise the author [38] proposed GRU technique for GHI forecasting for day ahead. The model utilized meteorological and historical data as input of the proposed model and its performance is measured in terms of RMSE and forecast skill.

1.2 Research Gap and Motivation

(a). The data processing method is one of the most important requirements for increasing the accuracy of solar irradiance forecasting models. The empirically chosen wavelet basis function determines the deconstructed results of wavelet decomposition (WD). The better the decomposition results, the better the basis function is for data. The lack of universality of WD in multi-regional forecast may be due to the empirical necessities for preset basis function. Furthermore, because wavelet convolution assumes frequency stationary over the wavelet time span, WD is seen as inferior alternatives for estimating the instantaneous frequency. These concerns are theoretically solved by EMD and EEMD which are data adaptive multi scale decomposition. However, in practice mode mixing issue affect EMD performance and the Gaussian white noise introduced to EEMD requires a significant amount of effort to remove. The aforementioned issues would have an impact on the accuracy

and practicality of EMD and EEMD. (b) In literature number of research related to GRU, LSTM, CNN are available for the forecasting of wind and photovoltaic power forecasting [39–41]. But a very few experiments related to Bi-LSTM are available for the forecasting of GHI. However, several researchers use Bi-LSTM for the forecasting of wind speed, price, load forecasting, stock exchange forecasting as well as in covid-19. The author of [42] use standalone BiLSTM to forecast the irradiance using meteorological data as input and the result indicates the superiority of BiLSTM model over LSTM, GRU and LR models in terms of RMSE. An ensemble approach of CNN-BiLSTM proposed by [43] to forecast the irradiance. The proposed model performance is better than CNN-LSTM and CNN-GRU model. The characteristics of the input data time series were extracted using CNN; whereas, the correlations of the time series were accessed using BiLSTM.

Based on the problems faced by previous methodologies, this study is motivated by the following assumption: (a) CEEMDAN used as a preprocessing strategy to effectively extract the randomness, trends in the historical data features. (b) The deep learning BiLSTM model process information twice (forward and backward direction) to capturing the nonlinear characteristics of time series data. (c) Grid search algorithm used to optimize the parameters: concealed units, epochs, drop factor, learning rate etc. of deep learning model which help to improve the forecasting accuracy. Therefore, to further enhance the forecasting accuracy, a hybrid model, namely, CEEMDAN-BiLSTM(SCF) is proposed.

1.3 Contribution and Paper Organization

Based on the literature findings, it has been observed that, the solar irradiance forecasting accuracy can be increased by using (a) appropriate signal processing algorithm (SPA) (b) deep learning network having a capability to handle large amount of data along with them optimization technique is used to precisely select the hyperparameters.

Number of authors used various combinations of SPA and learning network to increase the forecasting accuracy [39–46]. These studies attest to the improvement in the prediction performance by accurately selection of preprocessing technique, learning network and optimization technique. But, the drawback of previous published models is using all decomposed component as input of learning network which increase complexity and simulation time. This study performs various experimental scenario to overcome the

disadvantage and proposed model used only selected component to forecast the solar GHI. The proposed model (CEEMDAN-BiLSTM (SCF)) performance is compared with the Persistence model, and other well known deep learning techniques such as LSTM, GRU, BiLSTM, Standard CEEMDAN based BiLSTM (CEEMDAN-BiLSTM(Standard)) and modified CEEMDAN based BiLSTM (CEEMDAN-BiLSTM(modified)).

The following are the study's significant contributions and innovations in brief:

- (1) Firstly, a brief literature of deep learning techniques is discussed in all aspects.
- (2) Reproduction of persistence model, well known deep learning models: standalone LSTM, GRU and Bi-LSTM model.
- (3) In order to cover the gap of CEEMDAN-BiLSTM process, the numerous tactics of the CEEMDAN based BiLSTM models are studied. For this the conventional CEEMDAN is combined with BiLSTM (CEEMDAN-BiLSTM(Standard)) to forecast irradiance while in the alternative case the CEEMDAN scope is broadened by merging distinct CEEMDAN decomposed components and feeding them into BiLSTM network (CEEMDAN-BiLSTM(modified)) to predict GHI.
- (4) The proposed model is developed to increase the forecasting accuracy using CEEMDAN-BiLSTM (SCF). In the proposed work, the best combination of CEEMDAN component is taken and allocate an individual BiLSTM network to each combination and results of individuals are aggregate to produce the final forecasted value.
- (5) The performance of the proposed experiment is compared with well known deep learning techniques and benchmark model in terms of RMSE, MAPE and R^2 . The response of proposed model is better in all perspectives with lesser annual RMSE of (18.86 W/M², 22.24 W/m², 26.25 W/m²), MAPE of (2.19%, 4.81%, 6.77%) and R^2 of (96.4, 95.4, 93.74) respectively for 1-hr, 2-hr and 3-hr ahead solar GHI forecasting respectively. The forecast skill of proposed model is also observed which 89% with respect to persistence model.

The remaining paper is structure as follows: Section 2 explains the theoretical background of CEEMDAN, LSTM and BiLSTM deep learning network. Section 3 present the proposed model framework; whereas, Section 4 describes numerous experimental scenarios; Section 5 discusses result analysis; Section 6 discusses forecasting model ability using hypothesis test. Finally, study is concluded in Section 7.

2 Theoretical Methodology of CEEMDAN and Data Driven Model

This part presents a brief discussion of decomposition technique i.e., CEEMDAN and deep learning network related to proposed work i.e., LSTM and Bi-LSTM network.

2.1 CEEMDAN

(Complete Ensemble EMD with adaptive noise)

The EMD is proposed by Huang in 1998. The basic idea is EMD decompose the non-linear and non-stationary data into IMFs and residue. However, research has revealed that EMD has a mode mixing constraint [44] Mode mixing means similar elements exist in IMFs. To address the mode mixing issue in EMD, the improved procedure EEMD is introduced. Even though the EEMD addresses the mode mixing issue, the Gaussian white noise that was added may not be eliminated during reconstruction, leading to an error [45]. The authors of [46] suggest CEEMDAN technique which is more advance form of EEMD to solve the aforementioned difficulty. CEEMDAN divide the original data sequence into fifteen IMFs and one residue which is shown in Figure 4. The steps followed in CEEMDAN are given as:

- (1) The original data sequence $k^n(t)$ is added with Gaussian noise $w^n(t)$ and noise standard error (ε) which can be expressed as

$$k^n(t) = k(t) + \varepsilon_o w^n(t) \quad \text{Where } n = 1, 2, 3 \dots m \quad (1)$$

- (2) The EMD decompose the data and the first IMF is evaluate by averaging all the decomposition component

$$IMF_1(t) = \frac{1}{x} \sum_{i=1}^x IMF_1^i(t) \quad (2)$$

The residual is calculated as

$$r_1(t) = k(t) - IMF_1(t) \quad (3)$$

- (3) Further, the signal $r_1(t) + \varepsilon_1 EMD_1 w^n(t)$ are decomposed using EMD to obtain second IMF and residue can be stated as follows:

$$IMF_2(t) = \frac{1}{x} \sum_{n=1}^x EMD_1(r_1(t) + \varepsilon_1 EMD_1(w^n(t))) \quad (4)$$

$$r_2(t) = r_1(t) - IMF_2(t) \quad (5)$$

- (4) At per following stages, the x_{th} residual and $(x + 1)_{th}$ decomposed components can be calculated as

$$r_x(t) = r_{x-1}(t) - IMF_x(t) \quad (6)$$

$$IMF_{x+1}(t) = \frac{1}{x} \sum_{n=1}^x EMD_1(r_x(t) + \varepsilon_x EMD_x(w^n(t))) \quad (7)$$

$IMF_{x+1}(t)$ represent the $(x + 1)_{th}$ IMF obtained by CEEMDAN
Repeat Equations (6) & (7) till the residual meets the requirement for stopping

$$\sum_{q=0}^Q \frac{|r_{x-1}(t) - r_x(t)|^2}{r_{x-1}^2(t)} \leq SD_x \quad (8)$$

Where Q represent the length of sequence $K(t)$ & $r_x(t)$ denote the sequence after x_{th} decomposition and the value of SD is set to 0.2.

- (5) Finally, the original signal $K(t)$ can be computed as:

$$y(t) = \sum_{i=1}^T IMF_i(t) + R(t) \quad (9)$$

Where $R(t)$ represent the final residual value

2.2 Long Short Term Memory Neural Network (LSTM)

J.J. Hopfield developed a Recurrent Neural Network (RNN) in 1982. In this network, the RNN output is related to the input via feedback acting like a dynamic memory [47]. For short term forecasting this network worked best, but for long term forecasting it becomes unstable. This inconsistency caused by gradient bursting i.e., substantial changes in training weights in a short period of time [48]. This problem was solved by LSTM to permit using of memory cells in a hidden layer. These memory cells are utilized to store information in an appropriate manner. Each memory cell having a forget gate (f_t), input gate (i_t) and output Gate (O_t) to accept or reject any information [49]. The architecture of LSTM network is shown in Figure 1. The LSTM network has three inputs $SI(t)$, previous memory previous memory cell output h_{t-1} and bias e_f . As a result, the activation value can be written as [50]

$$f_t = \text{sigmoid}(z_f \cdot [h_{t-1}, SI_i(t)] + e_f) \quad (10)$$

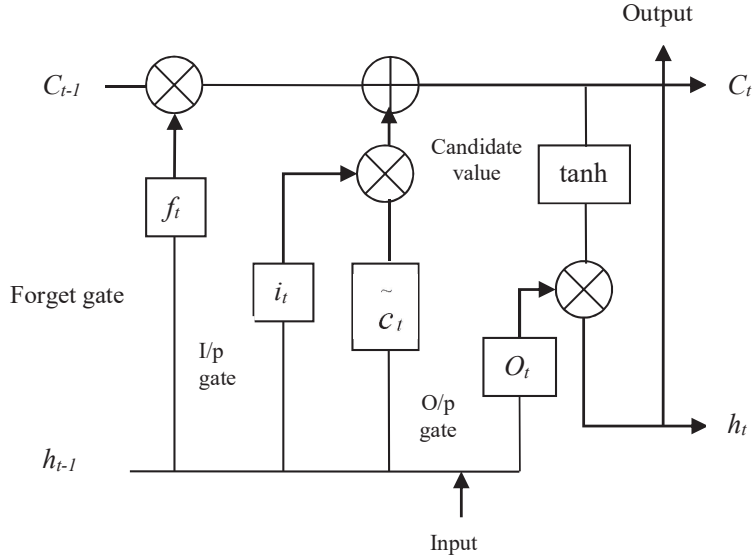


Figure 1 Basic configuration of LSTM network.

The LSTM network use the equation below to determine whether data information should be discarded or maintained [48]

$$i_t = \text{sigmoid}(z_i \cdot [h_{t-1}, SI_i(t)]) + e_i \quad (11)$$

$$c_t = \tanh(z_c \cdot [h_{t-1}, SI_i(t)] + e_c) \quad (12)$$

$$c_t = f_t * c_{t-1} + i_t * c_t \quad (13)$$

Now, the memory cell output represented as [48]

$$o_t = \text{sigmoid}(z_o \cdot [h_{t-1}, SI_i(t)] + e_o) \quad (14)$$

$$h_t = o_t * \text{softsign}(c_t) \quad (15)$$

Here e_f , e_i , e_c and e_o represents the bias voltage of LSTM and Z_f , Z_t , Z_C , Z_o is the weight factor of LSTM network & value of sigmoid lie from 0 to 1.

2.3 Bi-Directional Long Short Term Memory Neural Network (Bi-LSTM)

Bi-LSTM is one type of neural network consist of two LSTM model which having a capability to transfer information in past to future (forward

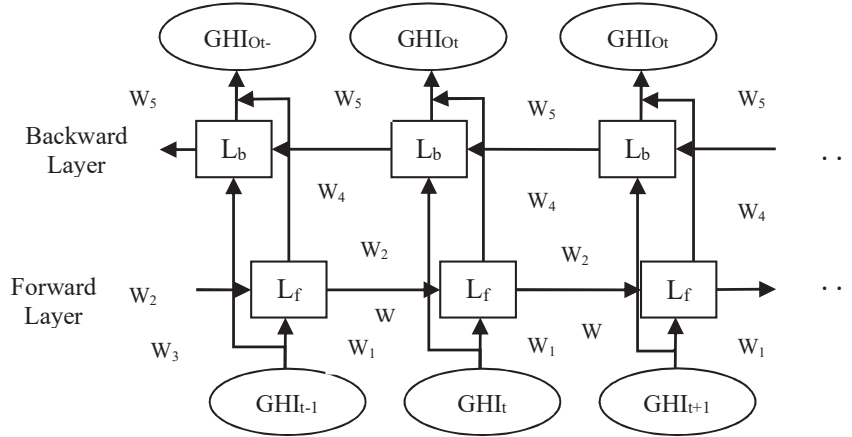


Figure 2 Basic architecture of Bi-LSTM network.

direction) and future to past (backward direction) [51]. Due to processing of input in both directions, twice training of data is possible and prediction accuracy is better than single LSTM model [52]. The basic architecture of Bi-LSTM is shown in Figure 2.

The Bi-LSTM network is updated with the help of parameter i.e., forward hidden layer (H_f), backward hidden layer (H_b) and output sequence $SI_o(t)$. The parameter of the Bi-LSTM is represented mathematically [49]

$$H_f = \text{sigmoid}(w_1 SI_i(t) + w_2 H_{f-1} + a_{H_f}) \quad (16)$$

$$H_b = \text{sigmoid}(w_3 SI_i(t) + w_5 H_{b-1} + a_{H_b}) \quad (17)$$

$$SI_o = w_4 H_f + w_6 L + a_{SI_o} \quad (18)$$

H_f , H_b & $SI_o(t)$ represent the forward parameter, backward parameter and output sequence while w denotes the weight factor.

3 Structure of the Proposed CEEMDAN-Bi-LSTM-Grid Search Algorithm

The goal of this project is to increase the accuracy of Solar GHI forecasting by employing a CEEMDAN-based BiLSTM network with various CEEMDAN pre-processing scenarios. Figure 3 show the schematic diagram of the developed model and its steps is discussed below:

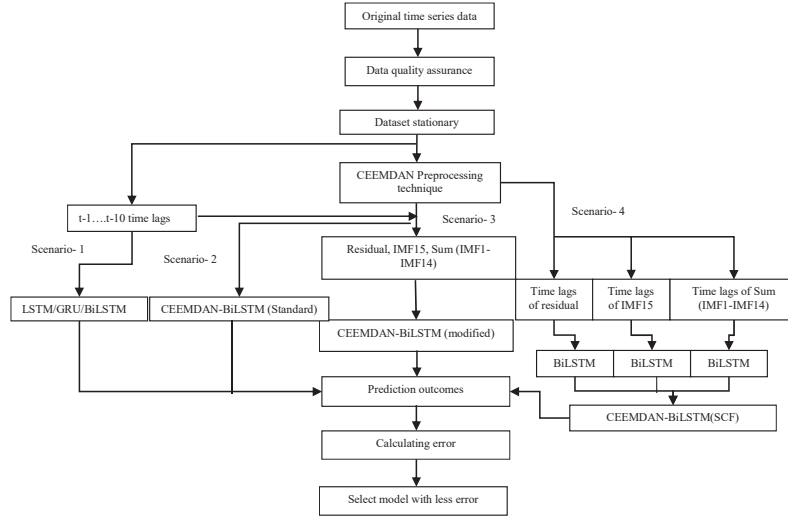


Figure 3 Schematic diagram of the developed model.

3.1 Data Quality Assurance and Data Stationarity

The input data has great impact on the model performance. Primarily, the collected data is available in its raw form which is random and non-linear in nature and has a great influence on the effectiveness of the forecasted model. Due to the weak pyranometer reaction, there is a chance of finding incomplete and negative data recording. So, these types of data recordings must be deleted before feeding to forecasting model [52]. Furthermore, the lack of solar radiation throughout the night, the night hour's data is omitted from the dataset and due to the cosine error of sensor the data just before and after sunset is also a perpetrator element in the model performance. Therefore, to enhance the effectiveness of forecasted model the data converted into stationary form before applying to application. To enhance the quality of input data, this paper calculates CSI index of data in which convert the data in stationary form. The CSI is calculated as follows [53]

$$GHI_{CSI} = \frac{GHI_i(t)}{CS_i(t)} \tag{19}$$

Where $CS_i(t) = I_0 \exp^{-\frac{\tau}{\sin^b(h(t))}} \sin(h(t))$

Here, b indicates the fitting parameter, $h(t)$ represents the solar height and I_0 denote the Extra-terrestrial radiation. From literature, it is found that clear sky GHI is similar to $CS_i(t)$. So, we replace both values with each other and

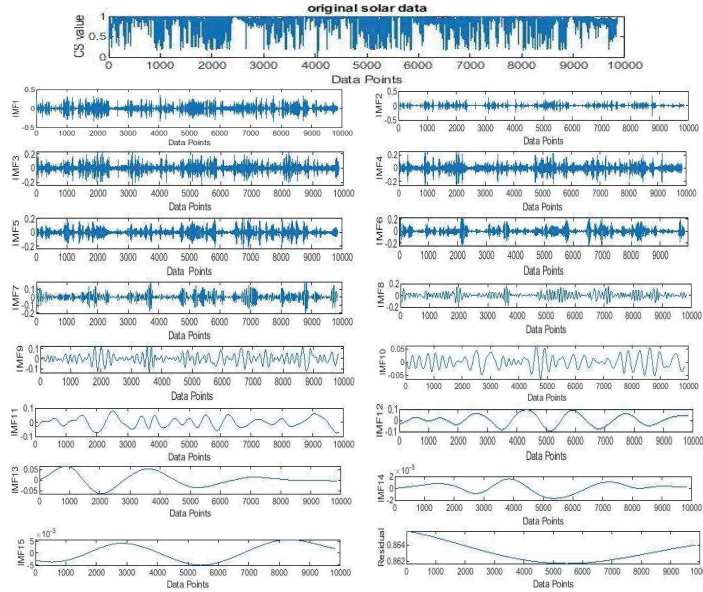


Figure 4 CEEMDAN Decomposition results.

it can be written as:

$$k = \frac{GHI_i(t)}{GHI_{CS}} \quad (20)$$

Where, GHI_{CS} is clear sky GHI values.

After refining and cleaning of time series data and render the data in stationary form using clear sky index calculation. Now, the CEEMDAN is applied on the prepared time series data in which decompose the data into fifteen IMFs and one residue. Figure 4 represents the CEEMDAN decomposition results

3.2 Hyper Parameters Selection Using Grid Search Algorithm

In literature, no any type of rule and regulation is present to select the hyperparameters. However, the study uses grid search algorithm to find the best hyperparameter value. The selection of hyperparameter is possible by changing the parameter's value with in a particular range. Table 1 showing selection of parameters with in a particular range and Figure 5 represent the specific flow graph used for selecting optimum parameters. The followings are the rules for choosing hyperparameters are mentioned below:

- (i) Assign default value to the initial hyperparameters

Table 1 Selection of hyper parameters

Hyperparameters	Search Bounds	Selected Value
Learning algorithm	Adam, sgd, RMSprop	Adam
Concealed units	60–125	100
Epoch	100–800	500
Drop factor	default	0.2
Learning rate	0.0001–0.1	0.0007
Gradient threshold	default	1
Drop period	50–175	125

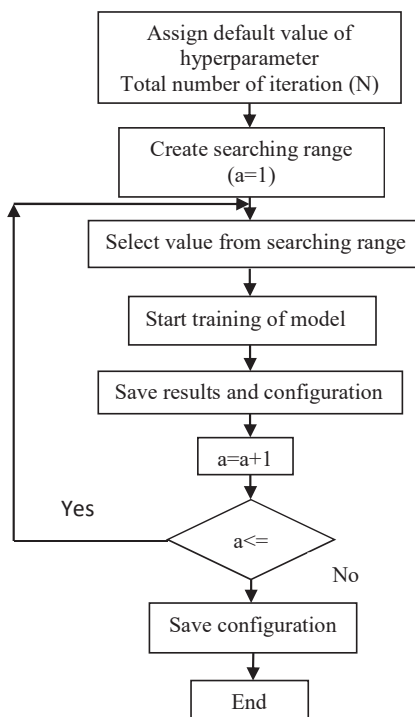


Figure 5 Hyperparameter selection flowchart.

- (ii) Select best learning rate
- (iii) Select the appropriate learning algorithm
- (iv) Choose appropriate number of concealed layers
- (v) Select relevant activation function
- (vi) Choose the optimum batch size & epoch value

3.3 Forecasting Process

In this stage, CSI value is decomposed using CEEMDAN in which fifteen IMF's & residue are obtained. As a result, the fifteen IMF's & residue with sufficient time legs are used as input features of the forecasted model. This research conducted a largescale experiment to determine the best GHI value using a various combination of decomposed components as shown in Table 3. The testing data is divided on seasonal basis: winter, spring, summer, monsoon and autumn as given in Delhi Tourism website [54]. The developed model performs short term forecasting for each season. The prediction value is in the form of CSI sequence. Using below equation, the CSI sequence is converted into real Global Solar Irradiation

$$GHI(t) = CS(t) \times CSI \quad (21)$$

CSI = clear sky index; $CS(t)$ = clear sky GHI.

3.4 Performance Evaluation

In this study, five statistical metrics are used to evaluate the performance of proposed model which are MAPE (mean absolute percentage error), RMSE (root mean square error), R^2 (correlation coefficient), FS (forecast skills) and percentage improvement test.

Mean Absolute Percentage Error (MAPE): – It is mostly used forecasting error to determine the performance of forecasted model in which measure the uniform forecasting error in percentage [1]

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{I(t) - \hat{I}(t)}{I(t)} \right| \times 100 \quad (22)$$

$I(t)$ = measured solar irradiance, $\hat{I}(t)$ = forecasted solar irradiance and n represent the total number of measured values.

Root Mean Square Error (RMSE): – It is a metric which is more sensitive to measure deviation between forecasted and observed value [2]

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (I(t) - \hat{I}(t))^2} \quad (23)$$

Correlation Coefficient (R^2): – This metric measure correlation between observed and forecasted value and its value is range from 0 to 1 [1, 2]

$$R^2 = 1 - \frac{\sum_{i=1}^n |I(t) - \hat{I}(t)|^2}{\sum_{i=1}^n I(t)} \times 100 \quad (24)$$

Forecast Skills: – The improvement in the proposed model with respect to reference model which is irrespective of prediction horizon, method and location

$$FS = 1 - \frac{indicator_{proposed\ model}}{indicator_{comparison\ model}} \quad (25)$$

The following expression are used to measure the percentage improvement between developed models

$$P_{MAPE} = \frac{|MAPE_1 - MAPE_2|}{MAPE_1} \quad (26)$$

$$P_{RMSE} = \frac{|RMSE_1 - RMSE_2|}{RMSE_1} \quad (27)$$

Where $MAPE_1/RMSE_1$ is the error of reference model and $MAPE_2/RMSE_2$ is the error of considered model.

4 Simulation Results

This study uses a combination of CEEMDAN & BiLSTM to improve forecasting accuracy. The developed model performance has been compared with standalone models: persistence model, GRU, LSTM, BiLSTM and other CEEMDAN based models. The all experiments are evaluate using MATLAB 2019a and numerous models scenarios are analyzed. The following simulations are run in order to propose an accurate model:

- (i) Persistence model, standalone GRU, LSTM & BiLSTM network.
- (ii) Standard CEEMDAN preprocessing strategy with BiLSTM model (CEEMDAN-BiLSTM (Standard)).
- (iii) Modified CEEMDAN preprocessing strategy with BiLSTMmodel (CEEMDAN-BiLSTM(modified)).
- (iv) Developed a CEEMDAN decomposition forecasting strategy which combine selected decomposition component (CEEMDAN-BiLSTM (SCF)). Where SCF represent the selected component forecast.

- (v) Developed model Performance is evaluated using MAPE, RMSE, R^2 and forecast skills and compared with persistence model, standalone GRU, LSTM, Bi-LSTM model, CEEMDAN-BiLSTM (Standard), CEEMDAN-BiLSTM(modified) models.

4.1 Data Description

The Indian location dataset is used for the forecasting because of the substantial improvement in the infrastructure of renewable sector and the ever growing scope in India. In the study, the dataset of New Delhi is used to evaluate the proposed model due to its mixed climate characteristics of the targeted location. According to the koopen climate classification system, the New Delhi has climate characteristics of 'cwa' and 'bsh'. Its mixed characteristics of climate provide the model to perform on the different weather conditions. Three year hourly data (2012–14) of New Delhi (Capital of India) location collected from National Solar Radiation Database (NSRDB) for training, validation and testing purpose [55]. Many academics have used NSRDB data in their research due to various advantages (1) free and easily access (2) extensive temporal and spatial coverage (3) no missing value in the data. NSRDB provides satellite based data which acquired using a satellite to irradiance model created by State University of Newyork. The collected data from NSRDB containing hourly GHI values and several other meteorological variables. Two year data used for training and one year data used for testing the developed model on seasonal basis: winter (December to January), spring (February to March), summer (April to June), monsoon (July to mid September) and autumn (September end to November). Table 2 gives a geographical coordinates, climatic condition and clear sky hours details of the selected location.

4.2 Experimental Setup

The performance of all generated models including proposed framework are assessed for error (MAPE, RMSE), correlation coefficient (R^2) and forecast skills. Various type of test has been undertaken in order to implement the upgraded and accurate forecasting model.

Table 2 Geographical details of Delhi location

Location	Rainfall (mm)	Clear-sky Hours	Climate	Altitude (m)	Longitude	Latitude	Region
Delhi	714	2809	Cwa, Bsh	225	77.1025°E	28.7041°N	North

Cwa = Humid Subtropical; Bsh = Hot semi-arid.

Case (1) forecasting using persistence model, unidirectional LSTM, GRU, BiLSTM model

The goal of this scenario is to create an experimental study on benchmark model, unidirectional deep learning models (LSTM, GRU, BiLSTM). This experiment utilized ten time legs as an input features of deep learning model; whereas, GHI is forecasted as the output value. Short term solar irradiation forecasting is performed on seasonal basis where selection of deep learning hyper parameters is one of the significant tasks to obtain improved forecasting accuracy. The developed model performance is judged using MAPE, RMSE and R^2 respectively. The MAPE value obtained by persistence, unidirectional LSTM, GRU and BiLSTM model for short term solar GHI forecasting varies between 15.64–35.45%, 8.34–15.39%, 7.15–13.54% and 5.43–10.68% respectively for 1-hr ahead, 16.64–37.69%, 9.54–17.82%, 8.21–14.92% and 6.34–12.76% respectively for 2-hr ahead, 26.82–42.29%, 14.65–21.76%, 12.05–18.89% and 8.24–14.43% respectively for 3-hr ahead and RMSE value ranges from 48.87–87.75 W/m^2 , 38.23–66.32 W/m^2 , 35.89–63.36 W/m^2 and 33.78–60.96 W/m^2 respectively for 1-hr ahead, 52.79–89.13 W/m^2 , 40.21–69.13 W/m^2 , 38.12–66.02 W/m^2 and 36.23–63.41 W/m^2 respectively for 2-hr ahead, 58.76–94.13 W/m^2 , 44.76–73.98 W/m^2 , 41.35–69.21 W/m^2 and 38.91–65.81 W/m^2 respectively for 3-hr ahead. Moreover, the R^2 varies from 0.76–0.85%, 0.87–0.92%, 0.90–0.94% and 0.91–0.95% for persistence, unidirectional LSTM, GRU and BiLSTM model respectively for 1-hr ahead, 0.74–0.83%, 0.85–0.91%, 0.88–0.93% and 0.90–0.94% respectively for 2-hr ahead, 0.68–0.80%, 0.83–0.89%, 0.86–0.91% and 0.88–0.93% respectively for 3-hr ahead. Figure 6 depicts a comparison of this case in terms of annual average MAPE and RMSE respectively and the performance of BiLSTM is superior to benchmark model, GRU and LSTM model respectively.

Case (2) Standard CEEMDAN based BiLSTM forecasting (CEEMDAN-BiLSTM(Standard))

This scenario uses CEEMDAN preprocessing technique to decompose the global horizontal irradiance data in which generate Fifteen IMF's and one residue. This experiment utilized all IMF's and ten timeleg as input of the BiLSTM model and forecast the global horizontal irradiance. Tables 4–6 shows the result of Standard CEEMDAN based BiLSTM model in terms of MAPE, RMSE and R^2 . The Standard CEEMDAN based Bi-LSTM model achieve lower RMSE, MAPE and enhanced R^2 in comparison to unidirectional LSTM, GRU and Bi-LSTM model. This model reported RMSE value

varies from 26.21–50.28 W/m² for 1-hr ahead, 28.56–54.29 W/m² for 2-hr ahead and 31.18–56.53 W/m² for 3-hr ahead respectively; whereas, MAPE value ranges from 3.05–8.34% for 1-hr ahead, 4.39–10.65% for 2-hr ahead and 6.52–12.56% for 3-hr ahead respectively and R² value ranges from 0.92–0.96 for 1-hr ahead, 0.91–0.95 for 2-hr ahead and 0.89–0.94 for 3-hr ahead respectively. Furthermore, the annual average RMSE value determined for the same is 40.71 W/m² for 1-hr ahead, 42.24 W/m² for 2-hr ahead and 44.34 W/m² for 3-hr ahead respectively; whereas, the annual average MAPE value is 6% for 1-hr ahead, 7.61% for 2-hr ahead and 9.66% for 3-hr ahead respectively. The annual average R² obtained for the same is 0.94, 0.93 and 0.91 for 1-hr, 2-hr and 3-hr ahead respectively.

Case (3) Modified CEEMDAN based Bi-LSTM forecasting (CEEMDAN-BiLSTM(modified))

This scenario present modified CEEMDAN based BiLSTM forecasting model in which use different-2 combinations of IMF's components with time leg input while in Standard CEEMDAN based BiLSTM model utilized all decomposed component as input to the BiLSTM model. The influence of different combination of deconstructed component as an input feature on the performance of the model has been thoroughly investigated and from investigation it is observed that combination of sum(IMF1–IMF14) resultant single subseries, IMF15 and residual gives a best result for all seasons with respect to Performance criterion. Table 3 illustrates the result of the observation made for various composition of deconstructed component as input features for monsoon season for 1-hr ahead only. The same composition of decomposed component is applied for spring, summer, autumn and winter seasons. Finally, residual, IMF15 and sum (IMF1–IMF14) are used as input features to the forecasted model. This model reported RMSE value varies from 21.36–43.76 W/m², 23.54–47.08 W/m² and 25.54–49.71 W/m² for 1-hr, 2-hr and 3-hr ahead respectively. MAPE ranges from 2.51–7.3% for 1-hr ahead, 3.51–9.86% for 2-hr ahead and 5.12–11.32% for 3-hr ahead respectively. The R² varies from 0.93–0.97, 0.92–0.96 and 0.90–0.94 for 1-hr, 2-hr and 3-hr ahead respectively. Moreover, the annual average RMSE value determined for the same is 34.34 W/m², 36.23 W/m² and 38.07 W/m² for 1-hr, 2-hr and 3-hr ahead respectively; whereas, the MAPE value for 1-hr, 2-hr and 3-hr ahead forecasting is 4.98%, 6.59%, 8.43% respectively. The annual average value of R² is 0.95, 0.94 and 0.92 for 1-hr, 2-hr and 3-hr ahead respectively. The performance of this model is better in all perspectives

Table 3 Investigation of CEEMDAN decomposed component combination

Decomposed Component Combinations	MAPE (%)	RMSE (W/m ²)	Decomposed Component Combinations	MAPE (%)	RMSE (W/m ²)
IMF1, IMF2, IMF3, IMF4, IMF5, IMF6, IMF7, IMF8, IMF9, IMF10, IMF11, IMF12, IMF13, IMF14, IMF15, Residual	12.94	48.92	IMF3, IMF4, IMF5, IMF6, IMF7, IMF8, IMF9, IMF10, IMF11, IMF12, IMF13, IMF14, IMF15, Sum (IMF1–IMF2), Residual	11.92	47.84
IMF1, IMF2, IMF3, IMF4, IMF5, IMF6, IMF7, IMF8, IMF9, IMF10, IMF11, IMF12, IMF13, IMF14, Residual	12.14	48.54	IMF4, IMF5, IMF6, IMF7, IMF8, IMF9, IMF10, IMF11, IMF12, IMF13, IMF14, IMF15, Sum (IMF1–IMF3), Residual	11.11	47.07
IMF1, IMF2, IMF3, IMF4, IMF5, IMF6, IMF7, IMF8, IMF9, IMF10, IMF11, IMF12, IMF13, Residual	11.98	47.41	IMF5, IMF6, IMF7, IMF8, IMF9, IMF10, IMF11, IMF12, IMF13, IMF14, IMF15, Sum (IMF1–IMF4), Residual	10.89	46.14
IMF1, IMF2, IMF3, IMF4, IMF5, IMF6, IMF7, IMF8, IMF9, IMF10, IMF11, IMF12, Residual	13.42	48.02	IMF6, IMF7, IMF8, IMF9, IMF10, IMF11, IMF12, IMF13, IMF14, IMF15, Sum (IMF1–IMF5), Residual	12.05	47.92
IMF1, IMF2, IMF3, IMF4, IMF5, IMF6, IMF7, IMF8, IMF9, IMF10, IMF11, Residual	11.37	47.17	IMF7, IMF8, IMF9, IMF10, IMF11, IMF12, IMF13, IMF14, IMF15, Sum (IMF1–IMF6), Residual	10.73	46.02
IMF1, IMF2, IMF3, IMF4, IMF5, IMF6, IMF7, IMF8, IMF9, IMF10, Residual	10.48	46.98	IMF8, IMF9, IMF10, IMF11, IMF12, IMF13, IMF14, IMF15, Sum (IMF1–IMF7), Residual	9.84	45.94

(Continued)

Table 3 Continued

Decomposed Component Combinations	MAPE (%)	RMSE (W/m ²)	Decomposed Component Combinations	MAPE (%)	RMSE (W/m ²)
IMF1, IMF2, IMF3, IMF4, IMF5, IMF6, IMF7, IMF8, IMF9, Residual	10.51	46.45	IMF9, IMF10, IMF11, IMF12, IMF13, IMF14, IMF15, Sum (IMF1–IMF8), Residual	9.15	45.37
IMF1, IMF2, IMF3, IMF4, IMF5, IMF6, IMF7, IMF8, Residual	9.50	45.90	IMF10, IMF11, IMF12, IMF13, IMF14, IMF15, Sum (IMF1–IMF9), Residual	8.34	44.82
IMF1, IMF2, IMF3, IMF4, IMF5, IMF6, IMF7, Residual	10.32	46.82	IMF1, IMF2, IMF3, IMF4, IMF5, IMF6, IMF7, IMF8, IMF9, IMF10, IMF11, IMF12, IMF13, Sum (IMF1–IMF10), Residual	9.72	45.72
IMF1, IMF2, IMF3, IMF4, IMF5, IMF6, Residual	10.84	46.34	IMF12, IMF13, IMF14, IMF15, Sum (IMF1–IMF11), Residual	9.52	45.52
IMF1, IMF2, IMF3, IMF4, IMF5, Residual	9.39	45.01	IMF13, IMF14, IMF15, Sum (IMF1–IMF12), Residual	8.22	44.22
IMF1, IMF2, IMF3, IMF4, Residual	9.42	45.23	IMF14, IMF15, Sum (IMF1–IMF13), Residual	8.10	44.01
IMF1, IMF2, IMF3, Residual	8.59	44.10	IMF15, Sum (IMF1–IMF14), Residual	7.30	43.76
IMF1, IMF2, Residual	9.20	44.29	Sum (IMF1–IMF15), Residual	8.15	43.94
IMF1, Residual	8.19	43.98			

as comparison to persistence model, standalone deep learning models and CEEMDAN-BiLSTM(Standard).

Case (4) Forecasting using Proposed CEEMDAN-BiLSTM Model (CEEMDAN-BiLSTM (SCF))

In this case, the proposed model is used with the vision of attaining credible prediction accuracy gains. Unlike the other investigation, this one assigned a separate BiLSTM model to every component i.e., sum (IMF1–IMF14), IMF15, Residual to forecast the solar irradiance. The ten time legs of every component used as input features to their Bi-LSTM network. The output is obtained by summing the forecasted value of every BiLSTM model to produce final prediction. From Tables 4 and 5, it is observed that the proposed model MAPE ranges from 1.12–3.11% for 1-hr. ahead, 2.54–6.94% for 2-hr. ahead and 4.39–9.25% for 3-hr. ahead respectively; whereas, RMSE varies from 10.98–28.11 W/m² for 1-hr. ahead, 12.31–32.65 W/m² for 2-hr ahead and 17.28–37.14 W/m² for 3-hr. ahead respectively. Table 6 showing the R² value varies from 0.94–0.98, 0.93–0.97, and 0.91–0.95 for 1-hr, 2-hr and 3-hr ahead respectively. The proposed model achieves lowest annual average MAPE which is 2.19%, 4.81% and 6.77% for 1-hr, 2-hr and 3-hr ahead respectively and annual average RMSE value is 18.86 W/m², 22.24 W/m², 26.25 W/m² for 1-hr, 2-hr and 3-hr ahead respectively. The improved correlation coefficient value for 1-hr, 2-hr and 3-hr ahead is 0.96, 0.95 and 0.93 respectively. Figures 6–8 represents the annual average RMSE (W/m²) and MAPE (%) of all developed models.

5 Result Analysis

This research performs short term solar irradiance forecasting for the location of Delhi, India. Various experimental analyses are performed to obtain precise model with improved forecasting accuracy. The prediction performance of the proposed model was compared with persistence, standalone deep learning network (LSTM, GRU, Bi-LSTM) and CEEMDAN based Bi-LSTM models. Firstly, discuss the standpoint of overall outcomes than season wise perspectives are discussed.

- (1) Persistence model provide the lowest forecasting accuracy among all developed models and its performance is checked for 1-hr, 2-hr and 3-hr ahead solar GHI forecasting. The annual average RMSE obtained by the same model is (70.49 W/m², 73.64 W/m², 77.17 W/m²) and annual average MAPE value is (24.96%, 26.41%, 33.25%) for 1-hr, 2-hr and

Table 4 Comparison of proposed model and reported model on RMSE, MAPE and R^2 for 1-hr ahead solar irradiance forecasting

		RMSE(W/m ²)						
Models		Winter	Spring	Summer	Monsoon	Autumn	Annual	
1-hr ahead solar GHI forecasting	Persistence	81.54	74.15	60.18	87.75	48.87	70.49	
	LSTM	57.02	54.89	49.12	66.32	38.23	53.11	
	GRU	53.89	51.02	46.52	63.36	35.89	50.13	
	BiLSTM	51.61	49.34	39	60.96	33.78	46.93	
	CEEMDAN- BiLSTM(Standard)	45.88	44.41	36.81	50.28	26.21	40.71	
	CEEMDAN- BiLSTM(modified)	39.29	38.01	29.3	43.76	21.36	34.34	
	Proposed Model	22.51	20.64	12.06	28.11	10.98	18.86	
			MAPE (%)					
		Persistence	29.22	23.03	21.47	35.45	15.64	24.96
	LSTM	13.89	12.41	10.71	15.39	8.34	12.14	
	GRU	11.73	10.05	9.21	13.54	7.15	10.33	
	BiLSTM	9.86	8.49	6.35	10.68	5.43	8.16	
	CEEMDAN- BiLSTM(Standard)	7.11	6.71	4.83	8.34	3.05	6	
	CEEMDAN- BiLSTM(modified)	6.76	5.32	3.01	7.3	2.51	4.98	
	Proposed Model	2.35	2.19	2.21	3.11	1.12	2.19	
		Correlation Coefficient(R^2)						
	Persistence	78.03	81.68	83.25	76.82	85.37	81.03	
	LSTM	89.78	90.68	91.68	87.26	92.88	90.85	
	GRU	91.89	93.24	94.56	90.2	94.89	92.95	
	BiLSTM	92.31	94.12	95.02	91.02	95.11	93.51	
	CEEMDAN- BiLSTM(Standard)	93.45	95.45	95.86	92.26	96.98	94.8	
	CEEMDAN- BiLSTM(modified)	94.98	96.02	96.91	93.89	97.79	95.91	
	Proposed Model	95.55	96.32	97.11	94.9	98.13	96.4	

3-hr ahead respectively due to the weakest performance in the monsoon and winter season. The value of R^2 is also lowest which represent the poorest performance of persistence model among all.

- (2) In unidirectional deep learning models (LSTM, GRU, BiLSTM), BiLSTM outperforms the LSTM and GRU in terms of forecasting performance. The annual average RMSE provided by the BiLSTM is 46.93 W/m², 50.58W/m² and 51.54 W/m²; whereas, annual average

Table 5 Comparison of proposed model and reported model on RMSE, MAPE and R² for 2-hr ahead solar irradiance forecasting

		RMSE(W/m ²)						
Models		Winter	Spring	Summer	Monsoon	Autumn	Annual	
2-hr ahead Solar GHI forecasting	Persistence	84.72	77.79	63.81	89.13	52.79	73.64	
	LSTM	59.72	55.08	51.81	69.13	40.21	55.19	
	GRU	55.81	53.04	49.19	66.02	38.12	52.43	
	BiLSTM	53.81	52.35	47.12	63.41	36.23	50.58	
	CEEMDAN- BiLSTM(Standard)	47.81	45.15	35.41	54.29	28.56	42.24	
	CEEMDAN- BiLSTM(modified)	41.25	39.05	30.26	47.08	23.54	36.23	
	Proposed Model	25.53	23.86	16.89	32.65	12.31	22.24	
			MAPE (%)					
		Persistence	30.24	24.85	22.67	37.69	16.64	26.41
	LSTM	15.81	13.32	11.19	17.82	9.54	13.53	
	GRU	13.56	11.51	10.34	14.92	8.21	11.7	
	BiLSTM	10.98	9.41	8.54	12.76	6.34	9.6	
	CEEMDAN- BiLSTM(Standard)	9.23	7.68	6.12	10.65	4.39	7.61	
	CEEMDAN- BiLSTM(modified)	8.76	6.34	4.51	9.86	3.51	6.59	
	Proposed Model	6.41	4.86	3.31	6.94	2.54	4.81	
		Correlation Coefficient(R ²)						
	Persistence	76.81	80.78	82.92	74.53	83.67	79.74	
	LSTM	88.68	89.39	91.46	85.88	91.22	89.32	
	GRU	89.99	91.12	92.89	88.66	93.89	91.31	
	BiLSTM	91.39	92.49	93.22	90.89	94.86	92.57	
	CEEMDAN- BiLSTM(Standard)	92.99	94.11	94.89	91.21	95.45	93.73	
	CEEMDAN- BiLSTM(modified)	93.11	94.98	95	92.89	96.02	94.4	
	Proposed Model	94.58	95.12	96.21	93.9	97.23	95.4	

MAPE obtained by the same is 8.16%, 9.6% and 11.38% for 1-hr, 2-hr and 3-hr ahead respectively. The annual average value of R² obtained between forecasted and observed value is 0.93, 0.92 and 0.90 for 1-hr, 2-hr and 3-hr ahead respectively. Among GRU, LSTM and persistence model, GRU perform better than the other two models with average annual RMSE of 50.13 W/m², 52.43 W/m² and 53.74 W/m² and average annual MAPE of 10.33%, 11.7%, 14.31% for 1-hr, 2-hr and 3-hr ahead

Table 6 Comparison of proposed model and reported model on RMSE, MAPE and R² for 3-hr ahead solar irradiance forecasting

		RMSE(W/m ²)					
Models		Winter	Spring	Summer	Monsoon	Autumn	Annual
3-hr ahead Solar GHI forecasting	Persistence	88.89	79.34	67.43	94.13	58.76	77.71
	LSTM	62.89	57.34	53.43	73.98	44.76	57.08
	GRU	56.03	54.86	50.26	69.21	41.35	53.74
	BiLSTM	54.45	52.16	46.41	65.81	38.91	51.54
	CEEMDAN-BiLSTM(Standard)	49.12	46.29	38.61	56.53	31.18	44.34
	CEEMDAN-BiLSTM(modified)	43.31	40.56	31.26	49.71	25.54	38.07
	Proposed Model	31.68	26.29	18.86	37.14	17.28	26.25
		MAPE (%)					
Persistence		36.04	32.89	28.23	42.29	26.82	33.25
LSTM		18.98	16.49	13.35	21.76	14.65	17.04
GRU		15.31	13.51	11.81	18.89	12.05	14.31
BiLSTM		12.35	11.43	10.49	14.43	8.24	11.38
CEEMDAN-BiLSTM(Standard)		11.23	9.81	8.21	12.56	6.52	9.66
CEEMDAN-BiLSTM(modified)		10.94	8.32	6.45	11.32	5.12	8.43
Proposed Model		8.81	6.21	5.23	9.25	4.39	6.77
		Correlation Coefficient(R ²)					
Persistence		70.04	75.89	78.89	68.28	80.24	74.66
LSTM		85.53	86.56	88.02	83.32	89.91	86.66
GRU		88.92	90.54	90.09	86.62	91.02	89.43
BiLSTM		89.09	91.89	92.02	88.02	93.89	90.98
CEEMDAN-BiLSTM(Standard)		91.99	92.29	93.99	89.99	94.11	92.47
CEEMDAN-BiLSTM(modified)		92.01	93.39	94.99	90.05	94.56	93
Proposed Model		92.18	94.12	95.28	91.9	95.23	93.74

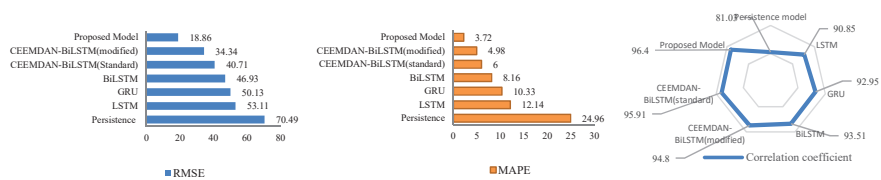


Figure 6 RMSE (W/m²), MAPE (%) and R²(%) of developed models on an annual average basis.

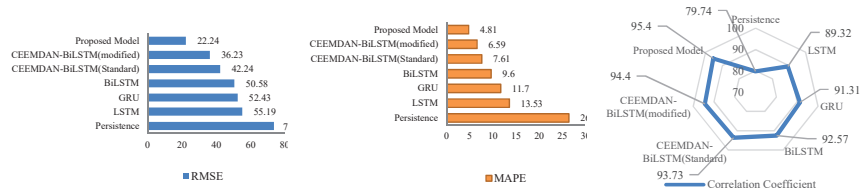


Figure 7 RMSE (W/m²), MAPE (%) and R²(%) of developed model on an annual average basis.

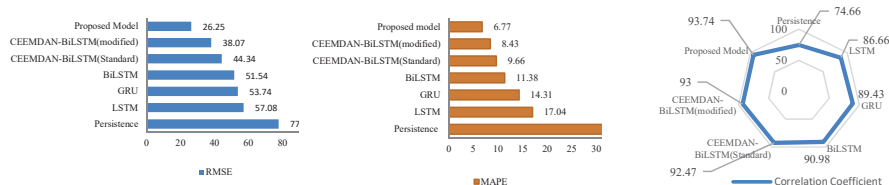


Figure 8 RMSE(W/m²), MAPE (%) and R²(%) of developed model on an annual average basis.

respectively. The obtained value of R² value is 0.92, 0.90 and 0.89 for 1-hr, 2-hr and 3-hr ahead respectively.

- (3) The standard CEEMDAN-BiLSTM (Standard) model accuracy is better than persistence and unidirectional deep learning models. It obtained 40.71 W/m², 42.24 W/m², 44.34 W/m² of annual average RMSE, 6%, 7.61%, 9.66% of annual average MAPE and 0.94, 0.93, 0.91 R² for 1-hr, 2-hr and 3-hr ahead respectively. On the other hand, modified CEEMDAN-BiLSTM (modified) achieved high accuracy as comparison to CEEMDAN-BiLSTM (Standard) in terms of RMSE (34.34 W/m², 36.23 W/m², 38.07W/m²), MAPE (4.98%, 6.59%, 8.43%) and R²(0.95, 0.94, 0.92) for 1-hr, 2-hr and 3-hr ahead respectively.
- (4) At last, the proposed model had the best annual average RMSE (18.86 W/m², 22.24 W/m², 26.25 W/m²) and annual average MAPE (2.19%, 4.81%, 6.77%) for 1-hr, 2-hr, 3-hr ahead respectively.

Now, discuss the results on seasonal basis

- (1) This paper evaluates the performance of all developed models on seasonal basis: winter (December to January), spring (February to March), summer (April to June), monsoon (July to mid-September) and autumn (September end to November). It is evident in Tables 4–6; persistence model accuracy is worst among all developed models. Among all seasons, performance of persistence model is good in autumn and summer

seasons due to less non-linearity in time series data in which easily modeled by the predictor while in spring, winter and monsoon seasons performance is not good due to overcast or rainy days in these seasons, which makes it very difficult to forecast.

- (2) As far as, comparison of unidirectional (LSTM, GRU & BiLSTM) models are concerned, the maximum accuracy is achieved by BiLSTM than single LSTM and GRU in all seasons for 1-hr ahead case from Table 4. The minimum RMSE (33.78 W/m²) and MAPE (5.43%) achieved by the BiLSTM model in autumn season and R² between real GHI and forecasted GHI is highest (0.95) for the same. The results indicates that the drastically improvement in forecasting accuracy of BiLSTM as comparison to persistence model, GRU and LSTM due to movement of information in both directions i.e., forward and backward direction
- (3) The standard CEEMDAN based BiLSTM model outperform unidirectional deep learning models and persistence model in terms of accuracy. Due to extraction of statistical information of input series, performance of CEEMDAN-BiLSTM (Standard) model is better in all seasons. It obtained minimum RMSE (26.21 W/m², 28.56 W/m², 31.18 W/m²) and MAPE (3.05%, 4.39%, 6.52%) to forecast 1-hr, 2-hr and 3-hr ahead solar irradiance forecasting for autumn seasons while in monsoon season, it achieved maximum RMSE (50.28 W/m², 54.29 W/m², 56.53 W/m²) and MAPE (8.34%, 10.65%, 12.56%) respectively to forecast 1-hr, 2-hr and 3-hr ahead solar irradiance forecasting.
- (4) It is indicated in results, using the modified CEEMDAN approach with the summation of deconstructed components IMF1–IMF14 improve the prediction accuracy over CEEMDAN-BiLSTM(Standard). The CEEMDAN-BiLSTM(modified) model achieved minimum RMSE (21.36 W/m², 23.54 W/m², 25.54 W/m²) and MAPE (2.51%, 3.51%, 5.12%) in the autumn season; whereas, it is (26.21 W/m², 28.56 W/m², 31.18 W/m²) and (3.05%, 4.39%, 6.52%) obtained by the CEEMDAN-BiLSTM(Standard) for the same season.

At last, the proposed model uses the separate Bi-LSTM model for each component: sum (IMF1–IMF14), IMF15, residual to forecast solar irradiance. The proposed model obtained minimum RMSE (10.98 W/m², 12.31 W/m², 17.28 W/m²) for the autumn season and maximum RMSE (28.11 W/m², 32.65 W/m², 37.14 W/m²) for the monsoon season; whereas, the MAPE produced by the proposed model is also lower than others i.e., minimum MAPE (1.12%, 2.54%, 4.39%) for the autumn season and maximum

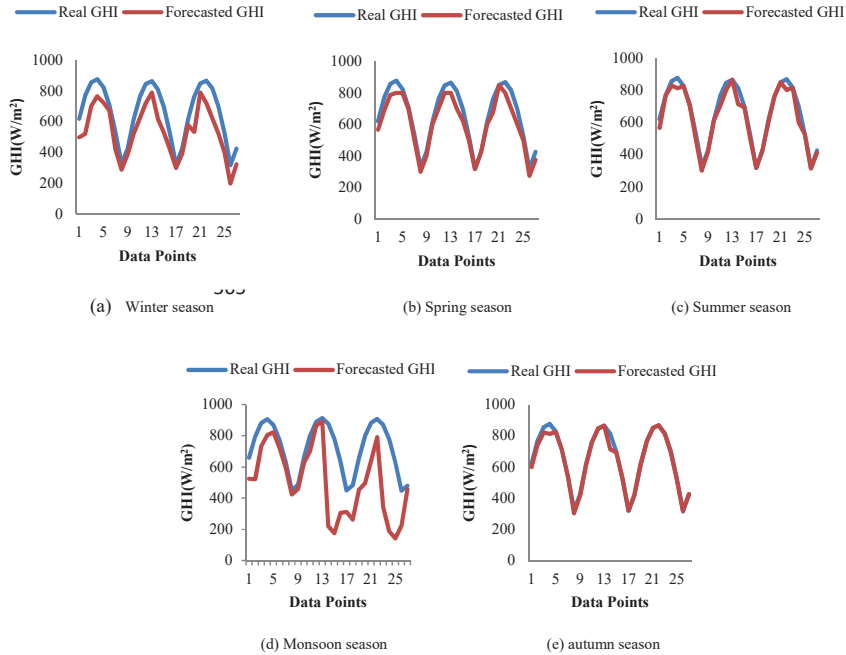


Figure 9 Proposed model performance for all seasons.

MAPE (3.11%, 6.94%, 9.25%) for the monsoon season. The correlation coefficient is drastically improved by the proposed model in the autumn season i.e., 98.13, 97.23, 95.23 while initially it was 85.37, 83.67, 80.24 for persistence model for one step, two step and three step ahead solar irradiance forecasting respectively.

For the deeper examination of the findings, Figure 9 provides a statistical representation of real and predicted GHI for the all seasons. These five seasons are used to compare the outcomes of the season with the lowest RMSE (autumn) and the season with the highest RMSE (monsoon). However, for clarity, only real and predicted GHI curve of suggested model is shown for selected seasons. From Figure 9, it is observed that substantial fluctuations in the real GHI generate a larger error in the results. For example, smooth curve of autumn season indicates the clear environmental circumstances in which easily traceable by the model. On the other hand, monsoon season shows substantial fluctuations in the real GHI due to existence of overcast or rainy days making it difficult to trace by the model resulting in maximum inaccuracies. From Figure 9, it can be deduced that if fluctuations in the real

GHI is higher, than similarity exist between real and predicted GHI is lower. Similarly, the resemblance between real and predicted GHI is higher when variance in the real GHI is fewer. However, with in a tolerated range of error, the suggested model also faces a number of ambiguities associated with a genuine GHI. As a result of these findings, the suggested model is a good forecasted model for stable as well as for unstable season.

6 Discussion

The previous section discusses about the developed forecasting model performance in terms of RMSE, MAPE and R^2 . This section provides a more comprehensive details of proposed model performance in terms of percentage improvement, hypothesis test and directional change of forecast. The details about the discussion are mentioned as follows:

6.1 Percentage Improvement

Percentage improvement is the main criterion to indicated the performance of proposed model against other developed models. Figure 10 represents that the proposed model offers a maximum percentage improvement in RMSE (49.07 W/m^2 , 49.46 W/m^2 , 45.3 W/m^2) and MAPE (75.87%, 61.24%, 50.14%) for 1-hr, 2-hr and 3-hr ahead respectively over single BiLSTM model. Likewise, CEEMDAN-BiLSTM(Standard) has also shows a significant improvement in RMSE (40.07 W/m^2 , 35.88 W/m^2 , 31.75 W/m^2) and MAPE (57.04%, 45.91%, 36.39%) for 1-hr, 2-hr and 3-hr ahead respectively. In addition, the proposed model exhibits a remarkable percentage improvement in terms of RMSE (24.7 W/m^2 , 22.21 W/m^2 , 15.45 W/m^2) and MAPE

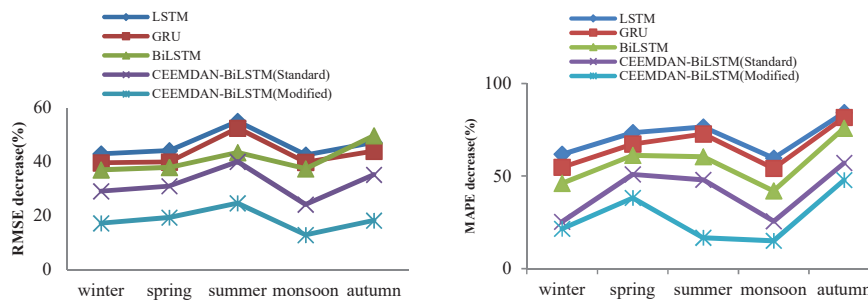


Figure 10 Proposed model outperform other developed models in term of percentage.

(47.8%, 27.63%, 25.36%) for 1-hr, 2-hr and 3-hr ahead respectively over CEEMDAN-BiLSTM(modified) model

6.2 DMH Test

Diebold-Mariano developed the DMH test to determine the performance differences between proposed and reference models. If z_i is the actual time sequence, z_i^1 is the first predicted sequence and z_i^2 is the next predicted sequence, then prediction error between these sequences represented as:

$$L[F_i^i] = z_i - z_i^1 \quad \text{and} \quad L[F_i^2] = z_i - z_i^2$$

$$i = 1, 2, 3 \dots n$$

then Diebold Mariano test can be calculated as:

$$DMH = \frac{\sum_{i=1}^n L[F_i^1] - L[F_i^2]/n}{\sqrt{\frac{u^2}{n}}} u^2 \tag{28}$$

Where u^2 is the deviation estimation.

Then, the null and alternative hypothesis can be calculated as:

$$\text{If } E_0 = F(L[F_i^1]) = F(L[F_i^2]) \tag{29}$$

$$\text{If } E_1 = F(L[F_i^1]) \neq F(L[F_i^2]) \tag{30}$$

According to the null hypothesis, there is no significant difference between two model’s performance while alternative hypothesis indicates the considerable difference in the forecasting ability of two models. The sorts of hypotheses are determined by comparing the output to a significance value of the standard error $H_{\beta/2}$. If the value of Diebold-Mariano test lie under the $[H_{\beta/2}, H_{-\beta/2}]$ is known as null hypothesis otherwise it is alternative hypothesis. Table 7 shows the DMH value for each season.

As shown in Table 7, using the suggested model, the majority of the DMH value are well over the crucial value of 1% level of significance. In other words, an alternate hypothesis is accepted at a 1% significance level or with a probability of 99 percent. The lower and upper values of DMH test for a 1-step ahead forecasting is 2.02 and 16.23 respectively. Furthermore, there are only three values lie below the crucial level of 1% of significance but over the 5% significance level. As a result, the suggested model accepts alternative hypothesis with a significance level of 5%.

Table 7 DMH test results for each season

	Step Ahead	Persistence Model					
			LSTM	GRU	BiLSTM	CEEMDAN-BiLSTM (Standard)	CEEMDAN-BiLSTM (Modified)
Winter	1	6.78*	5.43*	5.77*	6.23*	2.23*	3.45*
	2	7.82*	5.96*	6.45*	6.12*	5.12*	4.56*
	3	8.56*	6.11*	6.11*	7.21*	6.24*	5.34*
Spring	1	5.34*	4.23*	4.98*	8.23*	7.58*	6.45*
	2	4.98*	3.98*	3.12*	8.13*	7.34*	6.11*
	3	4.12*	3.12*	2.96*	6.71*	5.67*	4.56*
Summer	1	5.34*	4.12*	4.98*	4.56*	3.89*	2.19#
	2	5.76*	4.78*	3.89*	3.97*	2.17#	3.59*
	3	5.98*	4.98*	5.34*	5.12*	4.45*	3.45*
Monsoon	1	16.23*	15.89*	14.89*	13.28*	12.23*	11.23*
	2	16.57*	14.34*	13.56*	12.11*	11.67*	10.23*
	3	14.32*	13.24*	13.11*	12.21*	11.89*	10.12*
Autumn	1	4.35*	3.45*	3.98*	3.45*	2.02*	3.45*
	2	2.63*	2.12#	2.55*	2.48*	2.67*	3.76*
	3	4.29*	3.95*	4.12*	3.76*	3.21*	4.56*

*' indicate the 1% level of significance and '#' indicate the 5% level of significance.

Table 8 Crucial z value correspond to the significance level

Significance Level	10%	9%	8%	7%	6%	5%	4%	3%	2%	1%
Crucial z value	1.645	1.7	1.75	1.81	1.88	1.96	2.05	2.17	2.33	2.57

6.3 Directional Change in Forecasting (DC)

DC is used for evaluating the proposed model prediction ability and inform people on predicting direction of movement with defining moments. Higher DC value indicate the better forecasting model.

Figure 11 and Table 9 indicate the DC value of each model using below formula:

$$DC = \frac{100}{M} \sum_{i=1}^M G(t) \tag{31}$$

$$G(t) = \begin{cases} 1, & \text{if } (x_{f,i+1} - x_i)(x_{i+1} - x_i) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{32}$$

From below results, it is clear that DC score of proposed model is highest as comparison to other developed model. These findings can be viewed as a plausible turning points or a shift in the forecasted outcomes.

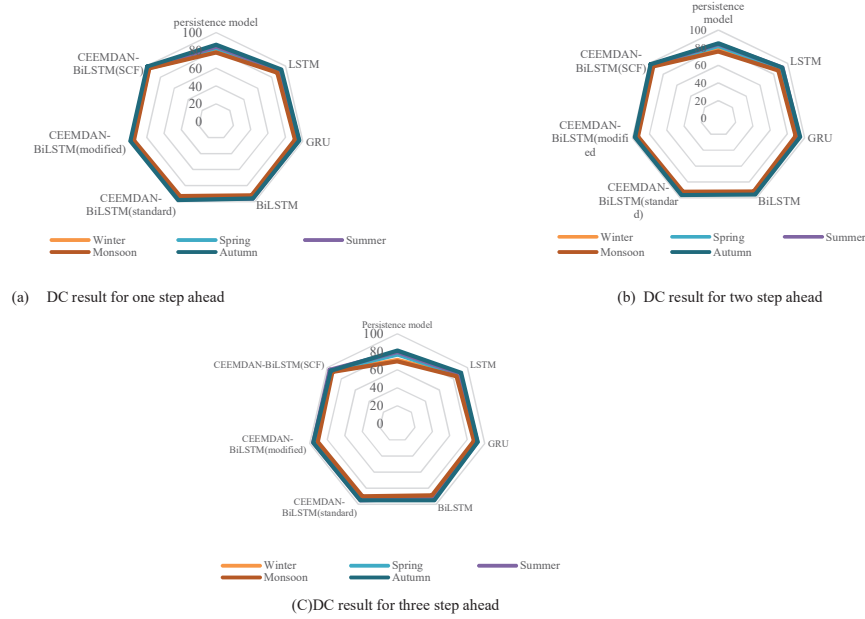


Figure 11 DC result for all seasons.

Table 9 DC Score of developed models

Seasons	Step Ahead	Persistence Model	LSTM	GRU	BiLSTM	CEEMDAN-BiLSTM (Standard)	CEEMDAN-BiLSTM (Modified)	CEEMDAN-BiLSTM (SCF)
Winter	1	79.03	90.78	92.89	93.31	94.45	95.98	96.55
	2	77.81	89.68	90.99	92.39	93.99	94.11	95.58
	3	71.04	86.53	89.92	90.09	92.99	93.01	93.18
Spring	1	82.68	91.68	94.24	95.12	96.45	97.02	97.32
	2	81.78	90.39	92.49	93.23	95.11	95.98	96.12
	3	76.89	87.56	91.54	92.89	93.29	94.39	95.12
Summer	1	84.25	92.68	95.56	96.02	96.86	97.91	98.11
	2	83.92	92.46	93.89	94.22	95.89	96	97.41
	3	79.89	89.02	91.09	93.02	94.99	95.99	96.28
Monsoon	1	77.82	88.26	91.20	92.02	93.26	94.89	95.9
	2	75.53	86.88	89.66	91.89	92.21	93.45	94.56
	3	69.28	84.32	87.62	89.02	90.12	91.05	92.45
Autumn	1	86.37	93.88	95.89	96.11	97.98	98.79	99.13
	2	84.67	92.22	94.89	95.86	96.45	97.02	98.23
	3	81.24	90.91	92.02	94.89	95.11	95.99	94.74

Table 10 Forecast skill (%) of proposed model on annual basis

Model	RMSE Forecast Skill			MAPE Forecast Skill		
	1-Step Ahead	2-Step Ahead	3-Step Ahead	1-Step Ahead	2-Step Ahead	3-Step Ahead
CEEMDAN-BiLSTM(Standard)	42%	43%	43%	76%	71%	71%
CEEMDAN-BiLSTM(modified)	51%	51%	51%	80%	75%	75%
Proposed Model	65%	63%	59%	89%	81%	79%

6.4 Forecast Skill (FS)

FS is another criterion for scoring a suggested model in comparison to the other developed model. Table 10 showing the FS of CEEMDAN-BiLSTM (standard), CEEMDAN-BiLSTM (modified) and CEEMDAN-BiLSTM(SCF) model with respect to persistence model. Greater value of FS indicates the better training capability over reference model.

Table 10 displays the forecasting skill performance of CEEMDAN-BiLSTM (standard), CEEMDAN-BiLSTM(modified)) and proposed model with respect to persistence model.

6.5 Validation Based on Prior Research

Table 11 shows the performance of the proposed model compared to previously published models in terms of MAPE, RMSE and forecast skill.

The proposed model performance is excellent over latest developed models. Table 11 represents that the proposed model offers a percentage improvement in RMSE (38.47 W/m^2), MAPE (42.59%) and FS (44.70%) respectively over singla, P., et al., 2022. Likewise, a significant improvement in RMSE (71.49 W/m^2) is showing by proposed model against Li et al., 2021. In addition, the proposed model exhibits a remarkable percentage improvement in terms of RMSE (45.35 W/m^2) and FS (52.70%) respectively over Toshniwal et al. model, 2021. The proposed model exhibits remarkable improvement over recent techniques in all prospective. Moreover, in literature various authors used different techniques to forecast solar GHI like as single, P., et al.; used a combination of WT-BiLSTM to forecast solar GHI. However, WT based model produced satisfactory results due to its superior localization features in both time and frequency domain. But it is unclear how to choose the appropriate wavelet function for a given dataset [61]. Similar problem occurred when using variational mode decomposition based method. Toshniwal et al, implement XGBF-DNN model to forecast solar GHI and measured developed model performance using RMSE (51.35 W/m^2) which

Table 11 Comparison of proposed work with previous developed models

Author and Year of Publication	Model	Place	Time Horizon	RMSE (W/m ²)	MAPE (%)	FS (%)
Zang et al., 2020 [47]	CNN-LSTM	Texas, USA	1-hr	69.26	–	–
Toshniwal et al., 2021 [52]	XGBF-DNN	NewDelhi, India	1-hr	51.35	–	40.2
Li et al., 2021 [42]	BiLSTM	United State	1-hr	98.44	–	–
Singla et al., 2021 [57]	WT-BiLSTM	Ahmadabad, India	24-hr	45.61	6.48	47
Gupta, A., et al. 2022 [59]	EEMD-GA-LSTM	New Delhi, India	1-hr	–	3.23	–
Gupta, A., et al. 2022 [60]	CEEMDAN-GA-BiLSTM	New Delhi, India	1-hr	–	2.23	59
This Work	CEEMDAN-BiLSTM (SCF)-Proposed Model	New Delhi, India	1-hr	18.86	2.19	89

is highly worst as comparison to current proposed work, because Extreme gradient boosting algorithm (XGBF) does not perform well on unstructured data and highly sensitive to outliers. So, in this research, the proposed work uses CEEMDAN (advance version of EEMD) & BiLSTM (advance type of LSTM) to predict solar GHI. CEEMDAN remove the Gaussian white noise added with the EEMD may not be cancelled after reconstruction while BiLSTM process the information in both direction (forward and backward) twice training of data is possible and prediction accuracy is better than single LSTM model. As a conclusion from the overall results, the proposed model appears to be a good alternative for forecasting solar GHI. However, the hyper parameters selection of deep learning models and longer simulation period are the main difficulties in creating the proposed model.

7 Conclusion

Solar energy is one of the most important and essential energy resource amongst all due to its great advantage over other renewable energy resources; yet its consistency and efficiency are critical for the smooth operation of

linked grid networks. So, this paper proposed a CEEMDAN-BiLSTM (SCF) model to forecast the solar GHI. CEEMDAN decompose the historical time series data into IMF's and BiLSTM used to forecast each subseries; whereas, grid search optimizes the learning parameters of deep learning model in a suitable search band. The systematic comparison has been done in the paper with other five models (LSTM, GRU, BiLSTM, CEEMDAN-BiLSTM(Standard), CEEMDAN-BiLSTM(modified)) under different time horizon (1–3hr ahead). The dataset of Indian location (New Delhi) is used for evaluating the performance of proposed model. The developed models performance is evaluate using DMH test, DC and common statistical metrics such as: MAPE, RMSE and R^2 . The result shows that proposed model achieves best performance among all developed models under all the time horizon. Finally, the conclusion can be summed up as follows:

- (1) Among all unidirectional models, BiLSTM outperform LSTM and GRU. The BiLSTM achieve lowest annual average RMSE (46.93 W/m^2) than those of LSTM (53.11 W/m^2) and GRU (50.13 W/m^2). These findings show that the BiLSTM has a superior data characterization training capability.
- (2) The CEEMDAN based hybrid models reduces the error and perform better as comparison to standalone models. For a 1-hr ahead forecast of solar irradiation, CEEMDAN-BiLSTM (standard) model obtained approximately average 65% of improvement in standalone BiLSTM. These results prove that CEEMDAN extract the hidden characteristics of time series data and improve the quality of input data.
- (3) Moreover, the CEEMDAN-BiLSTM (modified) model further improve the forecasting accuracy. This model uses different-2 combination of IMF's component with time lag input. From investigation, it is observed that combination of sum (IMF1–IMF14) resultant single sub series, IMF15 and residual gives a best result for all seasons. In the case of one step ahead forecast for Delhi location the RMSE (21.36 W/m^2 – 43.76 W/m^2) and MAPE (2.51%–7.3%) are significantly improved when compared to CEEMDAN-BiLSTM (Standard).
- (4) Furthermore, the selected component forecast model CEEMDAN-BiLSTM (SCF) provides superior results as compare to all other developed models. This one assigned a separate BiLSTM model to every selected component i.e., sum (IMF1–IMF14), IMF15, Residual to forecast the solar irradiance. The output is obtained by summing the forecasted value of every BiLSTM model to produce final prediction. For

a case of one step ahead forecasting, the CEEMDAN-BiLSTM (SCF) provides lowest RMSE (10.98–28.11 W/m²) and MAPE (1.12–3.11%) over CEEMDAN-BiLSTM (modified).

- (5) The Diebold-Mariano test represents that the score obtained by the proposed model is vastly different from the other developed models. The Proposed model accepts alternative hypothesis with a significance level of 5% for all seasons.
- (6) The results of DC demonstrate the proposed model forecasting ability and better defining moment of the suggested model. Higher value of DC indicates the better forecasting model.
- (7) The forecasting skill is another criterion for checking model performance. The proposed model achieves 89% FS against persistence model for one step ahead respectively.

From the overall results, it is proven that CEEMDAN-BiLSTM (SCF) model is best among all developed model and suggested to be good choice to forecast solar irradiation. However, while constructing the model, some challenges are face by the researcher such as higher computational time, accurate hyperparameter selection. As a result, by taking these issues into account in the future, more reliable and accurate results can be obtained in a shorter simulation period. Last but not least, it is meaningful to consider meteorological data and satellite images as input of the forecasted model and some other soft computing technique to automatically tune the hyper parameter with less running time.

Data Availability

The data that supports the finding of this study are available from the corresponding author upon reasonable request.

Code Availability

The codes of this study are available from the corresponding author upon reasonable request.

Declarations

Not Applicable.

Conflict of Interest

On the behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- [1] A. Gupta, K. Gupta and S. Saroha, A review and evaluation of solar forecasting technologies Materials Today: Proceedings, <https://doi.org/10.1016/j.matpr.2021.04.491>
- [2] Gupta, Anuj.; Gupta Kapil.; Saroha, Sumit.; Solar Irradiation Forecasting Technologies: A Review: Strategic planning for Energy and the Environemnt.2020: Vol 30, Iss 3–4, 2020. <https://doi.org/10.13052/spee1048-4236.391413>
- [3] M. Kocifaj, Sky luminance/radiance model with multiple scattering effect, Sol. Energy. 83 (2009) 1914–1922. <https://doi.org/10.1016/j.solener.2009.07.004>.
- [4] M. Kocifaj, M. Gangl, F. Kundracik, H. Horvath, G. Videen, Simulation of the optical properties of single composite aerosols, J. Aerosol Sci. 37 (2006) 1683–1695. <https://doi.org/10.1016/j.jaerosci.2006.08.002>.
- [5] Liang, Li.; Zhi Li.; Haiwei, Yu.; Medium load forecasting method with improved deep belief network for renewable. Distributed General and alternative energy journal. 2022: Vol 37, Iss 3, 2022. <https://doi.org/10.13052/10.13052/dgaej2156-3306.3735>
- [6] M. Kocifaj, Angular distribution of scattered radiation under broken cloud arrays: An approximation of successive orders of scattering. Sol. Energy. 86 (2012) 3575–3586. <https://doi.org/10.1016/j.solener.2012.06.022>.
- [7] Long, Fei.; Liu Fei.; Peng, Xiangli.; Yu, Zheng.; Power quality disturbance identification and optimization based on machine learning. Distributed General and alternative energy journal. 2022: Vol 37, Iss 2, 2022. <https://doi.org/10.13052/dgaej2156-3306.3723>
- [8] M. Q. Raza, M. Nadarajah, C. Ekanayake, On recent advances in PV output power forecast, Sol. Energy. 136 (2016) 125–144. <https://doi.org/10.1016/j.solener.2016.06.073>
- [9] C. Voyant, G. Notton, S. Kalogirou, M. L. Nivet, C. Paoli, F. Motte, A. Fouiloy, Machine learning methods for solar radiation forecasting: A review, Renew. Energy.

- [10] U. K. Das, K. S. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Y. I. Idris, W. VanDeventrer, B. Horan, A. Stojcevski, Forecasting of photovoltaic power generation and model optimization: A Review, *Renew. Sustain. Energy Rev.* 81 (2018) 912–928. <https://doi.org/10.1016/j.rser.2017.08.017>
- [11] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker, D. Renne, T. E. Hoff, Validation of short and medium term operational solar radiation forecasts in the US, *Sol. Energy.* 84 (2010) 2161–2172. <https://doi.org/10.1016/j.solener.2010.08.014>.
- [12] H. Yang, J. Kleissl. Preprocessing WRF initial conditions of coastal stratocumulus forecasting, *Sol. Energy.* 133 (2016) 180–193. <https://doi.org/10.1016/j.solener.2016.04.003>.
- [13] R. Perez, E. Lorenz, S. Pelland, Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe, *Sol. Energy.* 94 (2013) 305–326. <https://doi.org/10.1016/j.solener.2013.05.005>.
- [14] X. Mi, H. Liu, Y. Li, Wind speed prediction model using singular spectrum analysis, empirical mode decomposition and convolutional support vector machine, *Energy Convers. Manag.* 180 (2019) 196–205. <https://doi.org/10.1016/j.enconman.2018.11.006>.
- [15] M. Kocifaj, L. Komar, Modeling diffuse irradiance under arbitrary and homogenous skies: Comparison and validation, *Appl. Energy.* 166 (2016) 117–127. <https://doi.org/10.1016/j.apenergy.2016.01.024>.
- [16] M. Kocifaj, Unified model of radiance patterns under arbitrary sky conditions, *Sol. Energy.* 115 (2015) 40–51. <https://doi.org/10.1016/j.solener.2015.2015.02.019>.
- [17] G. Wnag, Y. Su, L. Shu, One-day-ahead daily power forecasting of photovoltaic systems based on partial functional linear regression models, *Renew. Energy.* 96 (2016) 469–478. <https://doi.org/10.1016/j.renene.2016.04.089>.
- [18] D. Yang, P. Jirutitjaroen, W. M. Walsh, Hourly solar irradiance time series forecasting using cloud cover index, *Sol. Energy.* 86 (2012) 3531–3543. <https://doi.org/10.1016/j.solener.2012.07.029>
- [19] X. Huang, J. Shi, B. Gao, Y. Taj, Z. Chen, J. Chen, J. Zhang, Forecasting Hourly Solar Irradiance Using Hybrid wavelet transformation and elman model in smart grid, *IEEE access.* 7 (2019) 139909–139923. <https://doi.org/10.1109/Access.2019.2943886>.
- [20] C. C. Turrado, M. del C. M. Lopez, F. S. Lasheras, B. A. R. Gomez, J. L. C. Rolle, F. J. de C. Juez, Missing data imputation of solar radiation

- data under different atmospheric conditions, *Sensors* (Switzerland). 14 (2014) 20382–20399. <https://doi.org/10.3390/s141120382>.
- [21] R. C. Deo, X. Wen, F. Qi, A. wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset, *Appl. Energy*. 168 (2016) 568–593. <https://doi.org/10.1016/j.apenergy.2016.01.130>.
- [22] F. Baser, H. Demirhan, A. fuzzy regression with support vector machine approach to the estimation of horizontal global solar radiation, *Energy*. 123 (2017) 229–240. <https://doi.org/10.1016/j.energy.2017.02.008>.
- [23] Z. Dong, D. Yang, T. Reindl, W. N. Walsh, A novel hybrid approach based on self-organizing maps, support vector regression and particle swarm optimization to forecast solar irradiance, *Energy*. 82 (2015) 507–577. <https://doi.org/10.1016/j.energy.2015.01.066>.
- [24] B. Amrouche, X. LePivert, Artificial neural network based daily local forecasting for global solar radiation, *Appl. Energy*. 130 (2014) 333–341. <https://doi.org/10.1016/j.apenergy.2014.05.055>.
- [25] H. Bouzgou, C. A. Gueymard, Fast short-term global solar irradiance forecasting with wrapper mutual information, *Renew. Energy*. 133 (2019) 1055–1065. <https://doi.org/10.1016/j.renene.2018.10.096>.
- [26] H. Liu, H. Q. Tian, X. F. Liang, Y. F. Li, Wind speed forecasting approach using secondary decomposition algorithms and Elman neural network, *Appl. Energy*. 157–194 (2015) 183–194. <https://doi.org/10.1016/j.apenergy.2015.08.014>.
- [27] S. Monjoly, M. Andre, R. Calif, T. Soubdhan, Hourly forecasting of global solar radiation based on multiscale decomposition methods: A hybrid approach, *Energy*. 119 (2017) 288–298. <https://doi.org/10.1016/j.energy.2016.11.061>.
- [28] S. Sun, S. Wang, G. Zhang, J. Zheng, A decomposition-clustering-ensemble learning approach for solar radiation forecasting, *Sol. Energy*. 163 (2018) 189–199. <https://doi.org/10.1016/j.solener.2018.02.006>.
- [29] H. Lan, H. Yin, Y. Y. Hong, S. Wen, D. C. Yu, P. Cheng, Day-ahead spatiotemporal forecasting of solar Irradiation along a navigation route, *Appl. Energy*, 211 (2018), 15–27 <https://doi.org/10.1016/j.apenergy.2017.11.014>.
- [30] K. Mohammadi, S. Shamshirband, C. W. Tong, M. Arif, D. Petkovic. A new hybrid support vector machine-wavelet transform approach for estimation of horizontal global solar radiation, *Energy Convers. Manag.* 92 (2015) 162–171. <https://doi.org/10.1016/j.enconman.2014.12.050>.

- [31] S. Hussain, A, Alalili, A hybrid solar radiation modeling approach using wavelet multiresolution analysis and artificial neural networks, *Appl. Energy*. (2017). <https://doi.org/10.1016/j.apenergy.2017.09.100>.
- [32] Qing X, Niu Y, Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy* 148: 461–468 (2018). <https://doi.org/10.1016/j.energy.2018.01.177>.
- [33] Kumari P, Toshniwal D, Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance. *Clean Prod* 279:123285 (2021) <https://doi.org/10.1016/j.jclepro.2020.123285>.
- [34] Zang H, Cheng L, Ding T, Cheung KW, Wei Z, Sun G, Day-ahead photovoltaic power forecasting approach based on deep convolution neural networks and meta learning. *Int J. Electr Power Energy Syst* 118:105790 (2020a), <https://doi.org/10.1016/j.ijepes.2019.105790>.
- [35] Feng C, Zhang J, SolarNet: a sky image-based deep convolution neural network for intra-hour solar forecasting. *Sol Energy* 204: 71–78 (2020). <https://doi.org/10.1016/j.solener.2020.03.083>
- [36] Liu D, Sun K, Random forest solar power forecast based on classification optimization. *Energy* 187 (2019). <https://doi.org/10.1016/j.energy.2019.115940>.
- [37] Mishra M, Byomakesha Dash P, Nayak J, Naik B, Kumar Swain S, Deep learning and wavelet transform integrated approach for short term solar PV power prediction. *Mesa J Int Meas Confed* 166:108250. <https://doi.org/10.1016/j.measurement.2020.108250>.
- [38] Gao. B, Huang X, Shi J, Tai Y, Xiao R, Prediction day-ahead solar irradiance through gate recurrent unit using weather forecasting data. *J. Renew Sustain Energy* 11(4):043705 (2019). <https://doi.org/10.1063/1.5110223>.
- [39] Ding M, Zhou H, Xie H, Wu M, Nakanishi Y, Yokoyama R, A gated recurrent unit neural networks based wind speed error correction model for short term wind power forecasting. *Neurpcomputing* 365: 54–61 (2019). <https://doi.org/10.1016/j.neucom.2019.07.058>
- [40] Kumar D, Mathur HD, Bhanot S, Bansal RC, Forecasting of solar and wind power using LSTM RNN for load frequency control in isolated microgrid. *Int Journal of Model Simulink* (2020). <https://doi.org/10.1080/00228203.2020.1767840>
- [41] Hu YL, Chen L, A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and differential evolutionary algorithms.

- Energy Convers Manag. 173:123–142. <https://doi.org/10.1016/j.enconman.2018.07.070>.
- [42] Li C, Zhang Y, Zhao G, Ren Y, Hourly solar irradiance prediction using deep BiLSTM network. *Earth Sci Informatics* 14:299–309. <https://doi.org/10.1007/s12145-020-005113>
- [43] Rai A, Shrivastava A, Jana KC, A CNN-BiLSTM based deep learning model for mid-term solar radiation prediction. *Int Trans Electr Energy Syst* 11(18):8613 (2021). <https://doi.org/10.3390/app11188613>.
- [44] Wu, Z.; Huang, N. E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* 2009, 1, 1–41.
- [45] K. M. Chang, Ensemble empirical model decomposition: A Noise assisted, *Biomed. Tech.*55 (2010) 193–201. <https://doi.org/10.1515/BMT.2010.030>.
- [46] P. Flandrin, E. Torres, M. A. Colominas, A complete ensemble empirical model decomposition with adaptive noise, in:2011 IEEE Int. Conf. Acoust. Speech Signal Process., IEEE, Prague, 2011:pp.4144–4147.
- [47] Zang H, Liu L, Sun L, Cheng L, Wei Z, Sun G, short term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations. *Renew Energy* 160:26–41 (2020b). <https://doi.org/10.1016/j.renene.2020.05.150>.
- [48] Hochreiter S, Schmidhuber J (1997) Long short term memory. *Neural Comput* 9(8):1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [49] Kulshrestha A, Krishnaswamy V, Sharma M (2020) Bayesian BiLSTM approach for tourism demand forecasting. *AnnTour Res* 83:102925 (2020). <https://doi.org/10.1016/j.annals.2020.102925>.
- [50] Fischer T, Krauss C, Deep learning with long short term memory networks for financial market prediction. *Eur J Oper Res* 270(2):654–669 (2018). <https://doi.org/10.1016/j.ejor.2017.11.054>.
- [51] Soreng, bineeta.; Zhi Li.; Haiwei, Yu.; An optimal islanding detection scheme for an inverter based distributed generation system. *Distributed General and alternative energy journal.* 2022: Vol 36, Iss 2, 2021. <https://doi.org/10.13052/10.13052/dgaej2156-3306.3735>
- [52] Yildirim O, A novel wavelet sequences based on deep bidirectional LSTM network model for EEG signal classification. *Comput Biol Med* 96:189–202 (2018). <https://doi.org/10.1016/j.combiomed.2018.03.016>
- [53] Bedi J, Toshniwal D, Deep learning framework to forecast electricity demand. *Appl Energy* 238:1312–1326 (2019). <https://doi.org/10.1016/j.renene.2018.08.044>

- [54] Yousif C, Quecedo GO, Santos JB, Comparison of solar radiation in Marsaxlokk, Malta and Valladolid, Spain. *Renew Energy* 49:203–206 (2013). <https://doi.org/10.1016/j.renene.2012.01.031>
- [55] Benali L, Notton G, Foulloy A, Voyant C, Dizene R, Solar radiation forecasting using artificial neural network and random forest methods: application tonormal beam, horizontal diffuse and global components. *Renew Energy* 132:871–884 (2019). <https://doi.org/10.1016/j.renene.2018.08.044>.
- [56] http://delhitourism.gov.in/delhitourism/aboutus/seasons_of_delhi.jsp
- [57] Gupta A., Gupta K., Saroha S, Solar Energy Radiation Forecasting Method. In: Agarwal P., Mittal M., Ahmed J., Idrees S. M. (eds) *Smart Technoloigoies for Energy and Environmental Sustainability. Green Energy and Technology*. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-80702-3_7
- [58] Singla, P., Duhan, M. & Saroha, S. An ensemble method to forecast 24-h ahead solar irradiance using Wavelet decomposition and BiLSTM deep learning network, *Earth Sci Inform* 15, 291–306 (2022). <https://doi.org/10.1007/s12145-021-00723-1>
- [59] Fan, J.; Wu, L.; Ma, X.; Zhou, H.; Zhang, F.: Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in air polluted regions. *Renew. Energy* 145, 2034–2045 (2020). <https://doi.org/10.1016/j.renene.2019.07.104>
- [60] Gupta, Anuj.; Gupta, Kapil.; Saroha, Sumit.; Short term solar irradiation prediction framework based on EEMD-GA-LSTM Method: Strategic Planning for Energy and the Environment, Vol. 41_3, 255–280. <https://doi.org/10.13052/spee1048-5236.4132>
- [61] Gupta, A., Gupta, K., & Saroha, S. (2022). Short Term Solar Irradiation Forecasting using CEEMDAN Decomposition Based BiLSTM Model Optimized by Genetic Algorithm Approach. *International Journal of Renewable Energy Development*, 11(3), 736–750. <https://doi.org/10.14710/ijred.2022.45314>
- [62] Gupta, A., Gupta, K., Saroha, S. (2023). Single-Step Ahead Solar Irradiance Forecasting Using Hybrid WT-PSO-Based Neural Network. In: Namrata, K., Priyadarshi, N., Bansal, R. C., Kumar, J. (eds) *Smart Energy and Advancement in Power Technologies. Lecture Notes in Electrical Engineering*, vol 927. Springer, Singapore. https://doi.org/10.1007/978-981-19-4975-3_31.

Biographies



Anuj Gupta received the B.Tech in Electronics and Communication Engineering from Kurukshetra University, M.Tech in Electronics and Communication Engineering from Kurukshetra University, Kurukshetra. Presently he is pursuing Ph.D. in the area of solar irradiance forecasting from Electronics and Communication Engineering Department, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, India. He participates in the many national and international conferences and seminars, organized at different levels. His research area is deregulated electricity market, solar irradiance forecasting. He has more than ten years teaching and research experience.



Sharad Sharma received his B.Tech in Electronics Engineering from Nagpur University, Nagpur, India in 1998 and M.Tech in Electronics and Communication Engineering from Thapar Institute of Engineering and Technology, Patiala, India in 2004. He has a teaching experience of more than 24 years. He has conducted many workshops on Soft Computing and its applications in engineering, Wireless Networks, Simulators etc. He has a keen interest in teaching and implementing the latest techniques related to wireless and mobile communications. He opened up a student chapter of IEEE as Branch Counselor. He has published 25 SCI/SCOPUS indexed research papers, 2 Books and 14 Book Chapters with international publishers. Presently, he

is working as Professor and Head-Electronics and Communication Engg. Deptt. MMEC, MMDU, Mullana, INDIA. His research interests are Internet of Things, Soft Computing, routing protocol design, performance evaluation and optimization for wireless mesh networks using nature inspired computing.



Sumit Saroha is currently working as Assistant Professor in the Department of Electrical Engineering, Guru Jambheshwar University of Science & Technology, Hisar, India.

He obtained Ph.D. in the area of forecasting issues in present day power systems. His research interests are Renewable Energy Forecasting, Transformer Design, Electricity Markets, Electricity Forecasting, Neural Networks, Wavelet Transform, Fractional Order Systems and Multi Agent Systems.

He is a Member of IEEE and he is an author and co-author of over 50 publications in various reputed journals including IEEE, Elsevier, Springer, Wiley Publication and many more. He participates in the many national and international conferences and seminars, organized at different levels. Presently, under his guidance and co-guidance more than 04 students are doing their Ph.D.

He is an author and co-author of 3 patents published at IPR India. Presently he is working on two projects entitled “**Cost Effective Multifunctional Prosthesis for Disabled Persons**” and “**Multifunctional Prosthesis Wheel Chairs for Disabled Persons**” under **RUSA 2.0, MHRD, Government of India** of amount **20 Lacs and 03 Lacs** Indian Currency respectively.

Further, Dr. Saroha is the **Startup Activity Coordinator** of **PDUIIC, RUSA 2.0, MHRD, Government of India** at **GJUS&T, Hisar Centre**. He is also **Faculty Co-Coordinator** of **AICTE-IDEA Lab** of worth **INR 1.10 Cr.** at the same.

