Short Term Solar Irradiation Forecasting using CEEMDAN Decomposition Based BiLSTM Model Optimized by Genetic Algorithm Approach

Anuj Gupta[1*], Kapil Gupta[1], Sumit Saroha[2]

[1]Maharishi Markandeshwar (Deemed to be University), Mullana-Ambala, India

[2] Guru Jambheshwar University of science and Technology, Hisar, India

Abstract. An accurate short-term solar irradiation forecasting is requiredregarding smart grid stability and to conduct bilateral contract negotiations between suppliers and customers. Traditional machine learning models are unable to acquire and to rectify nonlinear properties from solar datasets, which not only complicate model formation but also lower prediction accuracy. The present research paper develops a deep learningbased architecture with a predictive analytic technique to address these difficulties. Using a sophisticated signal decomposition technique, the original solar irradiation sequences are decomposed into multiple intrinsic mode functions to build a prospective feature set. Then, using an iteration strategy, a potential range of frequency associated to the deep learning model is generated. This method is developed utilizing a linked algorithm and a deep learning network. In comparison with conventional models, the suggested model utilizes sequences generated through preprocessing methods, significantly improving prediction accuracywhen confronted with a high resolution dataset created from a big dataset.On the other hand, the chosen dataset not only performs a massive data reduction, but also improves forecasting accuracy by up to 20.74 percent across a range of evaluation measures. The proposed model achieves lowest annual average RMSE (1.45W/m²), MAPE (2.23%) and MAE (1.34W/m²) among the other developed models for 1-hr ahead solar GHI, respectively, whereas forecast-skill obtained by the proposed model is 59% with respect to benchmark model. As a result, the proposed method might be used to predict short-term solar irradiation with greater accuracy using a solar dataset.

Keywords: Solar Irradiation, CEEMDAN, Genetic Algorithm, BiLSTM, Evaluation Metrics

1. Introduction

It is because of the greenhouse effect, pollution and the depletion of natural resources that now, more than ever, it is imperative to use renewable energy sources (RES) that do not pollute the environment and are free to be used to create electricity. Among RES, solar energy is one of the most popular energy sources for generating electricity with zero carbon emission and its market is growing significantly due to its long-term viability and support (Vasylieva *et al.*2019). Almost every year, the earth's surface receives around 1.5×10¹⁸ KWh/area of solar energy which is nearly ten times the current global usage. Among all Asian countries, China receives the highest annual average daily global solar radiation 20.2 MJ/ (m².d), while India receives just 18MJ/(m².d) solar radiation. 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 sources alone (Gupta *et al.* 2020). According to the International Energy Agency (IEA), the overall capacity of photovoltaic installations will reach 1700GW by 2030. However, according to the world energy state report, this power capacity has increased from 8.0 GW in 2007 to 402 GW in 2017 (Gupta *et al.* 2021) Furthermore, according to several studies, the power grid will be entirely functioning on the renewable energy source (RES) by the end of 2050 (Hajj *et al.* 2018). But, due to the variability in weather condition, the intensity of solar GHI is unstable which directly affects the output of photovoltaic power plant (Gupta *et al.*2022), resulting in poor reliability on photovoltaic power plant. So, a number of forecasting models have been developed to increase solar GHI forecasting accuracy (Singla *et al.* 2021). Solar irradiance forecasting technologies are classified into four categories: 1) Physical method 2) Machine learning method 3) Statistical method and 4) Hybrid method (Fang *et al.* 2019;

*Corresponding author:
Email: annu11gupta@gmail.com (A. Gupta)

A. Gupta et al

Int. J. Renew. Energy Dev. 2022, 11(3), 736-750

| 737

Gupta et al. 2022; Olatomiwa*et al.* 2015; Sharma *et al.* 2018; Yagli *et al.* 2019). The physical model uses meteorological and geographical parameters as an input to forecast solar irradiance and sets up a mathematical relation between meteorological data and forecasted GHI. Due to its complexity, less precision and high computational cost, this model is not popular among researchers (Hajj *et al.* 2021). European Centre for Medium Range Weather Forecast (ECMWF) and Weather Research Forecasting (WRF) are the two main methods in physical approach to forecast atmospheric and operational research (Richardson *et al.* 2020; Perez *et al.* 2012). Statistical methods such as Gaussian Progress Regression (GPR) (Shang *et al.* 2018; Piri *et al.* 2015), Autoregressive Integrated Moving Average (ARIMA) (Shadab *et al.* 2020) enhance forecasting accuracy and set up a mathematical relation between meteorological variables and GHI, but the poor correlation among input data and solar GHI leads to weak performance of these models. Machine learning models such as Artificial Neural Network (ANN) (Jahani *et al.* 2019), Elman Neural Network (ELMAN) (Dumitru *et al.* 2016) and Support Vector Machine (SVM) (Zeng *et al.* 2013) have a capability of self-correction and reduce the gap between forecast and measured data. Nevertheless, due to uncertain behaviour of GHI, single machine learning models get stuck in local minima and hence do not perform efficiently (Gupta *et al.* 2022). Therefore, hybrid models are discussed in literature to overcome these issues. Data decomposition based technique and machine learning model is one of the most used hybrid models. Several decomposition techniques such as wavelet Transform (WT), Empirical Mode Decomposition (EMD), Variational Mode Decomposition (VMD), etc. have been discussed in previous studies. The author (Monjoly *et al.* 2017) uses the EMD decomposition technique to decompose the input data and Auto Regression (AR) method and ANN model are used to make an estimate of GHI. The experimental result shows that the hybrid model achieves better result as compared to standalone AR and ANN models. Likewise, the EEMD and Self Organizing Map-Back Propagation (SOM-BP) network are combined to forecast solar irradiance (Zhang *et al.* 2018). EEMD decomposes input data and feeds to SOM-BP network to forecast solar GHI. The output of each SOM-BP network is aggregated to obtain output. In continuation, various studies have used WT as a decomposition technique with SVM and ANN model (Zendehboudi *et al.* 2018; Chen *et al.* 2021).

In addition, deep learning emerged as a powerful technique to forecast the solar GHI and obtained better performance than conventional models, in all aspects. Various researchers in their studies have suggested deep learning techniques with pre-processing strategy to enhance accuracy of the forecasting model. Many studies have employed Long Short Term Memory (LSTM) network to forecast HI, where weather data is used as an input to the LSTM network(Qing *et al.* 2018). The study proves the efficiency of LSTM network over BPNN linear regression in terms of RMSE. Hybrid model of LSTM and gradient boosting algorithm are implemented to prevent the situation of over fitting as compared to naïve predictor and SVM model (Kumari *et al.* 2021). The performance of ensemble approach shows that the proposed model significantly improves result in terms of RMSE. Similarly, scholars have developed a hybrid approach to forecast the solar irradiance using a combination of Convolution

Neural Network (CNN) and LSTM(Zang *et al.* 2020b). Historical properties of input data are acquired by using an LSTM network and geographical data is obtained using CNN. Further, short-term photovoltaic power forecasting is performed using a hybrid approach of residual network and CNNby author (Zang *et al.* 2020a). An ensemble approach of WT with LSTM model for irradiance forecasting for 1-h to 1-day ahead was successfully employed (Wang *et al.* 2018). The result shows that WT significantly improves forecasting accuracy. Gated Recurrent Unit (GRU) for the day ahead regarding solar GHI forecasting is proposed (Gao *et al.* 2019). The model utilizes meteorological and historical data as input into the proposed model and measured its performance through RMSE and forecast skill. The performance of GRU and Bi-LSTM, according to extant literature is equivalent. Bi-LSTM-based models appear to be more dependable because they have undergone more validations based on solar datasets from various places throughout the world (Fischer *et al.* 2018; Gao *et al.* 2020).All these studies exhibit satisfactory prediction results. In addition to deep learning network, various data decomposition techniques are used as pre-processing strategy to decompose irradiance data, clean up and define input data according to specifications. SOM, WT, EMD, EEMD, and Kalman filter are often used in solar irradiance forecasting. It is confirmed in a number of prior studies that WT-based models provide satisfactory results due to their outstanding localization properties in both temporal and sensitive attributes. However, it is not clear how to pick the right wavelet function for a set of data (Huimin *et al.* 2016). Same results are received when using VMD pre-processing approach, where the number of modes is an a priori value that must be specified at the start, but has a major impact on the decomposition results (see Liu *et al.* 2018). Given that by adding intrinsic mode functions, CEEMDAN technique displays its astonishing superiority in automatically responding to any irregular time-series (Wu *et al.* 2009). When confronted with such a challenge, it may be the best alternative. Prasad et al (2019) developed a CEEMDAN-RF model for multi-step ahead solar GHI forecasting. The author adds results of all sub-predictions from LSTM model, then rectifying summation using ant colony optimization technique. Qin *et al.* 2019) uses fuzzy classification technique with CEEMDAN-LSTM model. In this study, CEEMDAN divides the incoming data into many IMFs, fuzzy classification technique categorizes the IMFs into a number of groups and uses BiLSTM predictor for each group and summing all predictions to obtain final result. These studies have established a good base for applications that combine CEEMDAN and Bi-LSTM learning models.

However, the problem is that it is useless unless the number of sub-series is determined in advance using CEEMDAN-based models, since solar dataset's temporal resolution improves and recording period lengthens, dataset scale widens resulting in increased non-linearity and non-stability in the solar time series. The number of IMFs will increase dramatically as a result of using CEEMDAN approach on such a massive dataset. As a result, at least two barrier processes keep following the routines described in the preceding studies (Prasad *et al.* 2019; Wu et al. 2009). Firstly, more IMF components would result in more untrained data raising overall training cost. Secondly, if machine learning model is employed to predict components and forecasting error of

A. Gupta et al

Int. J. Renew. Energy Dev. 2022, 11(3), 736-750

| 738

each component adds up to the final error which affects the prediction accuracy of the model.

Therefore, with an aim to address this problem and increase its prediction accuracy; this paper proposes a new framework that combines CEEMDAN: a signal decomposition technique, Genetic Programming: a feature selection technique and Bi-LSTM: deep learning model. Unlike some prior work in this field (Prasad et al. 2019) all decomposed components from CEEMDAN method are no longer used for solar irradiation construction, but to provide a prospective feature set for BiLSTM model to learn from. Secondly, Genetic Algorithm decreases the size of the projected feature set collection and changes it into a subset with more useful data. As a result, the approach can significantly deal with the growing population of IMFs, allowing it to strongly employ the strong CEEMDAN approach for solar irradiation prediction.

Taking into account all of the preceding processes, the following are the primary contributions of this work:

- To deal with the increasing scale of datasets, a unique architecture of ensemble learning system incorporating CEEMDAN, GA, and Bi-LSTM for solar irradiation forecasting is presented. Rather than following the "decomposition—prediction—reconstruction" design used in previous investigations (Wu *et al.* 2009; Prasad *et al.* 2019) the suggested framework attempts to deliver a more compact and useful set of features out of a total prospective IMFs obtained by CEEMDAN technique.
- The approach proposed here is completely automatic and aposteriori. Unlike previous studies, this approach does not require the target system to have any pre-existing knowledge. wavelet functions for wavelet transform (Wang *et al.*2018) and fuzzy logical strategy for fuzzy based models (Qin *et al.* 2019). The entire model is based solely on the solar dataset that has been recorded.
- In this framework three-year data of Delhi location collected from NSRDB (National Solar Radiation Database), two-year data used for training and one-year data used for testing on seasonal basis have been used. The testing data is divided into seasons: winter, spring, summer, monsoon and autumn, as given in Delhi Tourism website.
- A detailed comparative evaluation of the results is undertaken in this work from a progressive multi-level. Unlike previous studies (Wu *et al.* 2009; Prasad *et al.* 2019) where comparisons with other models are made all at once, this study has focused on progressive features. First, a comparison between the present framework and non-CEEMDAN machine learning models is made, then it moves on to a comparison of models that use the CEEMDAN approach. In the proposed models we check the effectiveness of CEEMDAN decomposition, genetic algorithm feature selection technique and long short-term memory neural network model.

The performance of the proposed model is measured in terms of MAPE (%), RMSE (W/m²) and MAE (W/m²) evaluation metrics. The response of proposed model is better in all prospects with lesser annual average RMSE (1.45W/m²), MAPE (2.23%) and MAE (1.34W/m²), respectively for 1-hr ahead solar GHI forecasting.

2. Theoretical background of CEEMDAN and data driven model

2.1 CEEMDAN (Complete Ensemble EMD with adaptive noise)

The EMD was proposed by Huang in 1998. The basic idea is EMD to decompose the non-linear and non-stationary data into IMFs and its residue. However, research has revealed that EMD has mode mixing constraints (Huang *et al.* 2019). Mode mixing means similar elements exist in IMFs. The upgraded process EEMD is introduced to overcome the mode mixing problem in EMD. Even if mode mixing problem is addressed by the EEMD, the Gaussian white noise added with the EEMD may not be cancelled after reconstruction, resulting in an error (Bedi *et al.* 2019). Singla *et al.* 2022) suggest CEEMDAN technique which is more advanced form of EEMD to solve the aforementioned difficulty. CEEMDAN divide the original data sequence into fifteen IMFs and one residue which is shown in Fig.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

$$k^n(t) = k(t) + \varepsilon_o w^n(t) \qquad (1)$$

Where n=1, 2, 3, …..m.

2) The EMD decompose the data and the first IMF is calculated by averaging all the components of decomposition

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

The residual is calculated as

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

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:

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

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

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

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

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

$IMF_{x+1}(t)$ represent the $(x+1)_{th}$ IMF obtained by CEEMDAN.

Repeat equation (6) & (7) till the residual meets the requirement for stopping

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

Where Q represent the length of sequence K(t) &rx(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:

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

Where R represent the final residual value (c.f Huang *et al.* 1998)

A. Gupta et al

Int. J. Renew. Energy Dev. 2022, 11(3), 736-750

|739

2.2. Long Short Term Memory Neural Network (LSTM)

J.J.Hopfield developed a recurrent neural network (RNN) in 1982, where the RNN is related to the input via feedback acting like a dynamic memory. For short term forecasting this network works best, but for long term forecasting it becomes unstable. This inconsistency caused by gradient boosting i.e. substantial changes in training weights in a short period of time. This problem is 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. The basic configuration of LSTM network is shown in Fig.1. Each memory cell having a forget gate (f_t), input gate (i_t) and output gate (o_t) to accept/reject information. For a forward movement function, the previous cell state c_{t-1} is discarded by the LSTM network. At present, the LSTM network has three inputs $SI(t)$, previous memory cell output h_{t-1} and bias e_f. As a result, the activation value can be written as (Hochreiter *et al.* 1997)

$$f_t = sigmoid\big(z_f.\,[h_{t-1}, SI_i(t)] + e_f\big) \qquad (10)$$

The LSTM network uses the equation mentioned below to determine whether data information should be discarded or used

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

$$\tilde{c}_t = tanh(z_c.\,[h_{t-1}, SI_i(t)] + e_c) \qquad (12)$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \qquad (13)$$

Now the memory cell output represented as

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

$$h_t = o_t * tanh(c_t) \qquad (15)$$

Where $SI(t)$ is solar irradiation input value, e_f, e_i, e_c and e_o represents the bias voltage of LSTM and Z_f, Z_t, Z_C, Z_o are the weight factor of LSTM network. The value of activation function sigmoid and tanh lie from 0 to 1 and -1 to 1(Qin *et al.*2019)

2.3 Bi-directional Long ShortTermMemory Neural Network (Bi-LSTM)

Bi-LSTM neural network consists of two LSTM models, having capability to transfer information from past to future (forward direction) and future to past (backward direction). Due to processing of input in both directions, twice training of data is possible and prediction accuracy is better than single LSTM model (Singla *et al.* 2022). The basic architecture of Bi-LSTM is shown in Fig.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 Bi-LSTM is represented mathematically as follows:

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

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

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

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

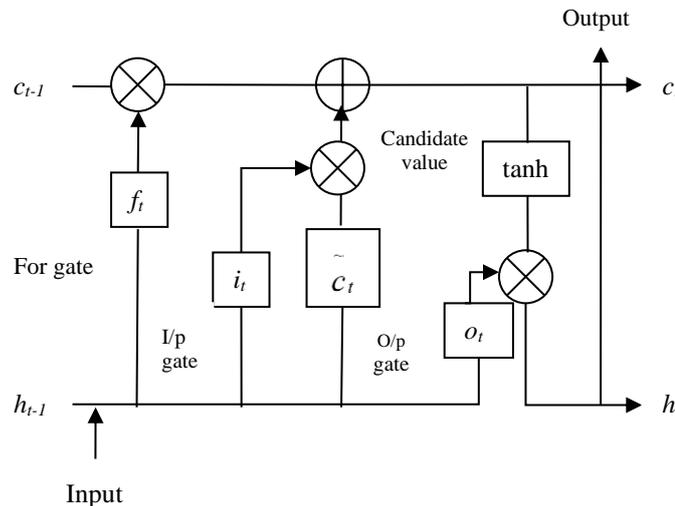**Fig.1** Basic configuration of LSTM network

A. Gupta et al

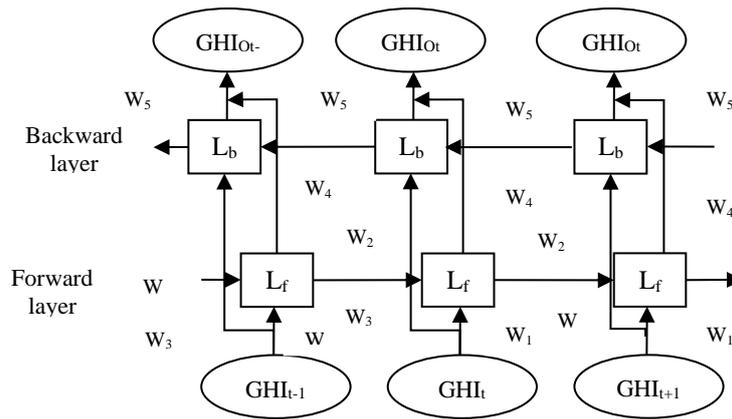*Int. J. Renew. Energy Dev. 2022, 11(3), 736-750*

| 740

Fig. 2 Bi-LSTM Structure

3. Structure of Proposed CEEMDAN-GA-BiLSTM framework

Purpose of present study is to find ways and means of increasing the accuracy of Solar GHI forecasting by employing a CEEMDAN based BiLSTM network with binary coded genetic algorithm. Figure 3 shows the schematic diagram of the developed model and its steps are duly discussed below:

3.1 Data description

The dataset of Indian location is used in the study to forecast solar GHI because of the substantial improvement in the infrastructure of renewable sector in India. In the study, the dataset of New Delhi is used to evaluate the proposed model due to its mixed climate characteristics vis-à-vis the targeted location. According to Koopen climate classification system, New Delhi has climate characteristics of 'cwa' and 'bsh'. Its mixed characteristics of climate provide the model to perform in different weather conditions. For this, three-year hourly data is used for training, validation and testing purposes. Many academics have used NSRDB data in their research because of various advantages 1) free and easy access 2) extensive temporal and spatial coverage 3) no missing value in the data (Gupta *et al.* 2022). NSRDB provides satellite-based data which was acquired by using a satellite for purposes of measuring of irradiance with the help of the model created by State University of NewYork. The data collected from NSRDB contains hourly GHI values along with several other meteorological variables. The study uses two-year data for training and one-year data for testing the developed model.

Table 1 provides geographical coordinates, climatic conditions and clear sky hourly details of the selected location.

3.2 Data quality assurance

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 model. Due to weak pyranometer reaction, there is a chance of finding incomplete and negative data recording (Zang *et al.* 2020b). Therefore, unsuitable recorded data must be deleted before feeding suitable data to the forecasting model. Furthermore, because of lack of solar radiation throughout the night, night hours data is omitted from the dataset. In addition to this, cosine error of sensor, the data pertaining to time just before and after sunset is also a perpetrator element in the model performance, hence it is also discarded. Moreover, to enhance the effectiveness of the forecasted model, the data is converted into stationary form before submitting it for analysis. Present paper, in order to enhance the quality of input data calculates normalized value of data for purposes of converting it into stationary form. Normalization is calculated as follows

$$X_{norm} = \frac{X_R - X_{min}}{X_{max} - X_{min}} \tag{19}$$

X_{norm} represents standardized value, X_R is the value to be normalized, X_{max} is the maximum value in all the values for related variables and X_{min} is the minimum value (Perez *et al.* 2012). After sifting of time series data and rendering it in stationary form using normalization calculation, CEEMDAN is applied to the prepared time series data, where in the data is decomposed into eight IMFs and one residue. Fig.4 represents the CEEMDAN decomposition results in the form of IMFs and residue.

Table1
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

A. Gupta *et al*
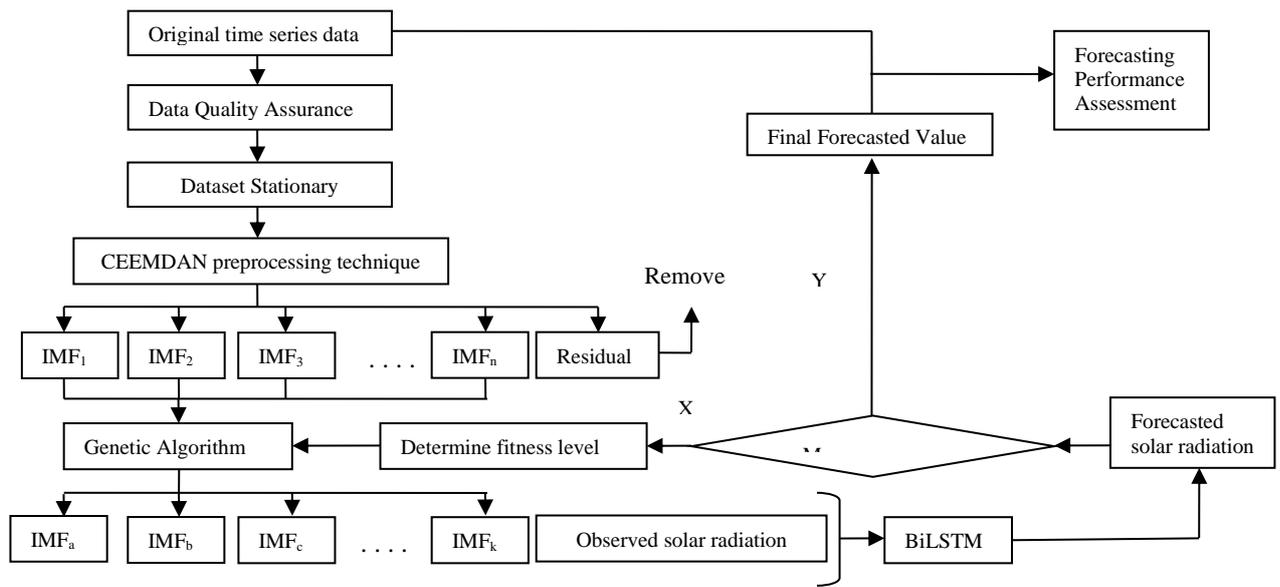
Int. J. Renew. Energy Dev. 2022, 11(3), 736-750

| 741

Fig.3 Schematic diagram of the proposed model
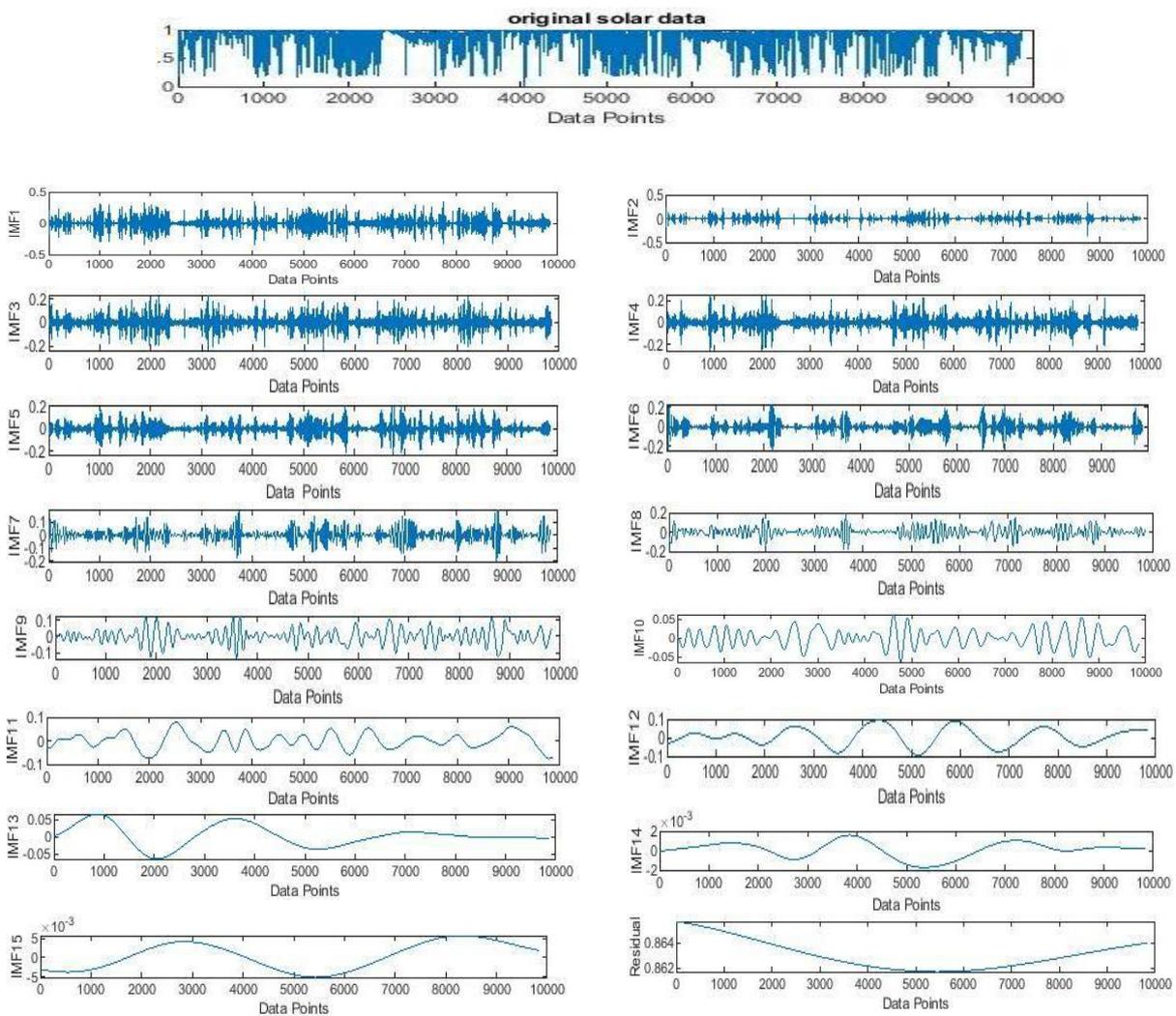
Fig. 4 CEEMDAN Decomposed IMF (Intrinsic Mode Function) components

A. Gupta et al

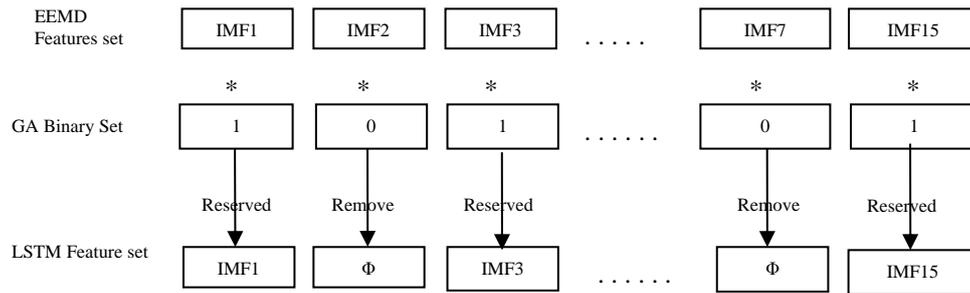Int. J. Renew. Energy Dev. 2022, 11(3), 736-750

| 742

Fig. 5 Genetic Feature selection in binary coding

3.3 Binary coded genetic algorithm

In feature extraction architecture and computational intelligence research the role of the wrapper is crucial. This study adopts binary code based GA to discover the best IMF as the feature set for training of LSTM in order to enhance the current solar irradiation predictor performance.

3.3.1 Binary coding

All fifteen IMFs are arranged from IMF1 to IMF15 to transfer in a small set of 1 and 0 (binary list) as shown in Fig.5. Combining these two lists via elemental multiplication yields the final selected list of IMFs. This allows us to decide whether IMF should be discarded or retained. The element under the binary list relevant index is set to 1 if an IMF is required; otherwise, it is set to 0.

3.3.2 Initial trails

For searching algorithms like GA, a proper initial condition is necessary because it can not only supply viable trails from the start but also disperse the searching spots globally. On the basis of these two considerations, the original population in this study is made up of binary sets as follows:
  i. All of the elements have been set to be one.
  ii. The value of all items has been set to zero.
  iii. The first half of the elements has a value of one, while the second half has a value of zero.
  iv. The first halves of the components are assumed to be 0, while the latter half is assumed to be 1.
  v. The items with the highest Correlation coefficient with the raw sequences are given a value of one while the others are given a value of zero.
  vi. The elements having the highest Pearson correlation with the raw sequence have their associated IMF's set to zero while the others are set to one

The Pearson correlation for a set of objective variables (P, Q) is given as

$$\rho_{P,Q} = \frac{E\big((P-E(P))(Q-E(Q))\big)}{\sigma_P \sigma_Q} \qquad (20)$$

where P denotes the unprocessed value and Q denotes the intrinsic mode functions; E (.) and σ (.) indicate the estimation and random deviation, respectively(Parsad *et al.* 2019).

3.3.3. Fitness

The best solution of this task is given as

$$F(\phi) = \frac{1}{min(MAE(\phi)_1, MAE(\phi)_2 \dots \dots MAE(\phi)_i)} \qquad (21)$$

Where $MAE(\phi)_i$ represents the mean absolute error between the forecasted and measured sequences on given data utilizing φ as the binary list for extracting features(Parsad *et al.* 2019).The purpose of finding the shortest path from numerous parallel tracks is to reduce the impact of arbitrary weight initialization and changing LSTM model parameter during training process which can lead to unanticipated localized end points even when the model structure is the same.

3.3.4 Evolutionary Procedure

Selection, crossover, and mutation all are parts of GA evolutionary process. These parts provide an overview of the main process.

3.3.4.1 Selection

There are two prerequisites to the selecting method. First, greatest individual generation sequence will sustain itself and bypass this filter. Furthermore, individuals with higher level of fitness will have greater chance of joining next generation; Probability is calculated by the equation given below

$$p(\phi_{g,k}) = \frac{F(\phi_{g,k})^2}{\sum_{m=1}^{M} F(\phi_{g,m})^2} \qquad (22)$$

Where $\phi_{g,k}$ denotes the k_{th} individual in the g_{th} generation, and m represents the population size (Qin et al. 2019).

The probability formula described above in equation no. 19 does not provide a pure elimination process. It does not only give promising individuals a better chance, but also allows those with low fitness to pass on to the next generation with a lower chance of survival. This is beneficial in terms of maintaining diversity of population and increasing algorithm reliability.

A. Gupta et al

Int. J. Renew. Energy Dev. 2022, 11(3), 736-750

|743

3.3.4.2 Crossover

Each DNA candidate will have the opportunity to recombine with another person from the same generation, which is known as crossover rate. DNA information from both parent sets will be inherited by the young individual. Participants and cross-points are picked at random for every cycle of crossover, as inspired by "the law of independent assortment" (Wu *et al.* 2009). It means every participant has equal opportunity to participate in the crossover process. As a result, there is no restriction on population diversity in this section.

3.3.4.3 Mutation

A mutation mechanism is implemented after crossover to minimize the pre-mature issue and to further broaden the seeking area. Should one of the DNA elements be changed, it will go from 0 to 1 or 1 to 0 but this operation does not have to be repeated in the case of all members of DNA collection, every moment as this may produce divergence issues and raise computation costs.

3.4 Forecasting Process

In this stage, normalized value is decomposed by CEEMDAN result in fifteen IMFs, and one residue is obtained. These IMFs and residue are used as input features in the forecast model. Large scale experiments were conducted in this research to determine the best GHI value. Data pertaining to two years was used for training and one year data was used for testing of the proposed model. The testing data has been divided on seasonal basis: winter, spring, summer, monsoon and autumn as given in Delhi Tourism website. The forecasting models are developed to predict one-hour ahead solar GHI for each season. Using the equation given below, normalized predicted sequence is converted into real solar GHI

$$X_{denorm} = X_{\min} + (X_{\max} - X_{\min}) \times X_{norm} \qquad (23)$$

X_{denorm} represents the denormalized value, X_{norm} is normalized value, X_{max} is the maximum value in all the values for related variables and X_{min} is the minimum value.

3.5 Performance Criteria

The training as well as testing set is separated from the raw dataset. Shuffling approach is not utilized so as to avoid information leakage, because this is a time series prediction issue. In this study, two years data set is utilized as the learning unit, whereas one year data is used as the testing dataset. It is in order to obtain the performance of developed model that testing data set is divided into five seasons: winter, spring, summer, monsoon and autumn. Assume $y = (y_1, y_2, \ldots, y_k, \ldots, y_n)$ is the solar irradiance time history and $\hat{y} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_k, \ldots, \hat{y}_n)$ is the predicted solar irradiance time series, used to calculate the performance of the proposed model (Gupta *et al.* 2021)

Mean Absolute Error (MAE):- It provides uniform forecasting error. This metric provides a difference between two set of data using Eq. (16) (Gupta *et al.* 2021)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| y_i - \hat{y}_i \right| \qquad (24)$$

Mean Absolute Percentage Error (MAPE): It provides uniform forecasting error in percentage using Eq. (17) (Gupta et al. 2021)

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \qquad (25)$$

Root Mean Square Error (RMSE): It is a statistic for assessing the largest expected error in the forecasted data Using Eq. (18) (Gupta et al. 2020)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)} \qquad (26))$$

Where n represent total number of points.

Forecast Skills: - The improvement in the proposed model with respect to reference model which is irrespective of prediction horizon, method and location (Singla et al. 2022)

$$FS = 1 - \frac{indicator_{proposed\,model}}{indicator_{comparison\,model}} \qquad (27))$$

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

$$P_{MAE} = \frac{|MAE_1 - MAE_2|}{MAE_1} \qquad (28)$$

$$P_{MAPE} = \frac{|MAPE_1 - MAPE_2|}{MAPE_1} \qquad (29)$$

$$P_{RMSE} = \frac{|RMSE_1 - RMSE_2|}{RMSE_1} \qquad (30)$$

Where $MAE_1/MAPE_1/RMSE_1$ is the error of reference model and $MAE_2/MAPE_2/RMSE_2$ is the error of considered model.

4 Result Analyses

This study uses a combination of CEEMDAN-GA-BiLSTM to improve forecasting accuracy. The developed model performance is compared with standalone models: Naïve Predictor, Gate Recurrent Unit (GRU), Recurrent Neural Network (RNN), Extreme Learning Machine (ELM), Back Propagation Neural Network (BPNN) and other CEEMDAN based models. All experiments are performed using MATLAB 2019a and numerous models' scenarios are analysed. Firstly, the results of the selected features from the GA are discussed. Secondly the proposed model performance is compared with naïve predictor, standalone GRU, BPNN, ELM and RNN model. Next, CEEMDAN method is applied to above mentioned standalone models and finally, evaluation of the selected features is studied.

4.1 Result of feature selection

In this study, the range of GA is set to 30 to balance the exploring ability and model cost. The synchronization lines of the mean and the optimum fitness among community are displayed in Fig.6, where average fitness increases steadily over the first 15 iterations and then gradually stabilizes after that despite minor oscillations. The best fitness grew slightly over the first five generations, but it

A. Gupta et al

Int. J. Renew. Energy Dev. 2022, 11(3), 736-750

| 744

remains practically unchanged after that indicating that the possible selected features have been discovered early.

Since average activities have not yet been converted to fitness values, an earlier halt is possible because the modification in fitness value is expected to be flat at this time. The DNA set given as:

$$[1\ 0\ 1\ 0\ 1\ 0\ 1\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ 0]$$

Above DNA set indicates that only six IMFs have been selected as the feature set out of the total fifteen IMFs. It means dropped 2/ 3 of the series have been fed into the BiLSTM, resulting in a substantially more compact model.

4.2 Analysis and assessment by comparison

Discussion herein is split into two sections: for the purposes of comparative study in the first half, various mainstream models will be considered. In the second half, it is on the basis of selected features evaluation that the result obtained by the proposed model is compared to the standalone BiLSTM model and the CEEMDAN-BiLSTM model which consist of all prospective features.

Case 1: Comparative research with standalone models

The goal hereinis to create an experimental study on benchmark model and non-CEEMDAN models: GRU, RNN, ELM, BPNN models. This experiment utilized ten time lag as an input feature of the non-CEEMDAN models, whereas solar GHI is forecast 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's performance is judged using MAPE (%), MAE (W/m²) and RMSE (W/m²) evaluation metrics. Data of two years has been used for training and one year data has been used for testing the developed models.

According to the statistical result of different developed models, as shown in Table 2 proposed model outperforms standalone models with a significant improvement regarding parameters in all seasons. However, forecasting performance for cloudy and rainy days produces high MAPE, RMSE and MAE in monsoon and winter seasons. The proposed model achieves lowest annual average result in respect of every statistical metric.
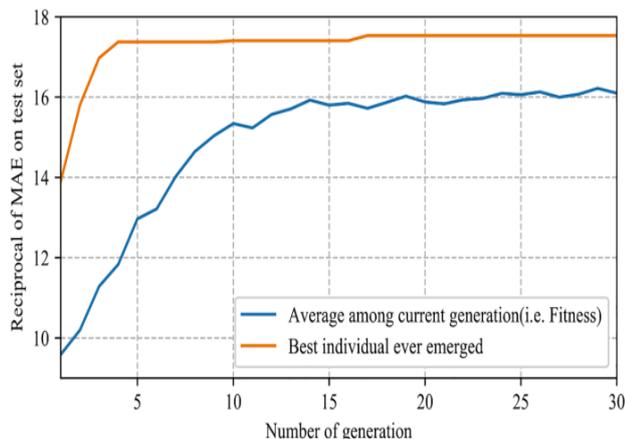

Fig.6 Generational changes in average and optimum fitness

The proposed model achieves lowest annual average MAPE (2.23 W/m²) in comparison to naïve predictor (5.36W/m²), BPNN (4.55W/m²), ELM (4.16W/m²), RNN (3.86W/m²), GRU (3.19W/m²), LSTM (3.07W/m²) for the year 2014. The credible percentage decrease in result of RMSE can be better observed in the proposed model in 2014 against naïve predictor (58.39%), BPNN (50.98%), ELM (46.39%), RNN (42.22%), GRU (30.09%) and LSTM (27.36%). Likewise for the year 2014, the minimum annual average RMSE is provided by the proposed model (1.45W/m²) in comparison to naïve predictor (4.85%), BPNN (4.11%), ELM (3.6%), RNN (3.18%), GRU (2.96%), and LSTM (2.73%). Proposed model exhibits decrease in value of the MAPE in percentile form against naïve predictor (70.10%), BPNN (64.72%), ELM (59.72%), RNN (54.40%), GRU (51.01%) and LSTM (46.88%). The proposed model out performs all other standalone models. Therefore, from the overall analysis, following observations are obtained:

(a) The performance of naïve predictor is the worst among all developed models because it does not use any knowledge or learning in order to make any prediction, while ELM and BPNN's performance is better as compared to Naïve, due to short training time because of random selection of hidden neuron, but unfortunately it faces fitness problem during training time, when its performance dips.

(b) Among standalone deep learning models RNN, GRU and LSTM, it is LSTM whose performance is better than that of RNN and GRU, since these models do not need fine adjustment of learning parameters, but it takes longer to train and process information, occurring only in one direction. These are the drawbacks of LSTM model. So, proposed model uses BiLSTM model to overcome the drawbacks of LSTM model. The result shows that the suggested model performs better than all other standalone models.

Case 2: Comparative research with CEEMDAN based models

The proposed method uses CEEMDAN preprocessing technique to decompose global horizontal irradiance data wherein are generated fifteen IMFs and one residue. This experiment utilizes all IMFs and one residue as input for CEEMDAN based model and forecast the global horizontal irradiance. From Table 3, it is observed that CEEMDAN significantly improved the performance of standalone models. CEEMDAN holds the input time series data and removes mode mixing constraints to improve the quality of input data. Among all CEEMDAN based models except the proposed model, CEEMDAN-LSTM outperforms other developed models. The CEEMDAN-LSTM obtained lowest annual average MAPE (2.54W/m²), RMSE (1.84%) and MAE (1.45W/m²) respectively. Proposed model also uses Genetic algorithm along with CEEMDAN to predict solar GHI. The proposed model achieves lowest annual average MAPE (2.23W/m²), RMSE (1.45%) and MAE (1.34W/m²), respectively for the year 2014. The credible percentage decrease in result of RMSE can be seen by the proposed model in 2014 against naïve predictor (58.39%), BPNN (50.98%), ELM (46.39%), RNN (42.22%), GRU (30.09%) and LSTM (27.36%), respectively. Similar is the

A. Gupta et al   Int. J. Renew. Energy Dev. 2022, 11(3), 736-750

| 745

percentage decrease in terms of MAPE by proposed model against naïve predictor (4.85%), BPNN (4.11%), ELM (3.6%), RNN (3.18%), GRU (2.96%), and LSTM (2.73%) for the year 2014. Moreover, proposed model decreases MAPE, RMSE and MAE in percentile form against CEEMDAN-BPNN, CEEMDAN-ELM, CEEMDAN-RNN, CEEMDAN-GRU and CEEMDAN-LSTM, respectively. Therefore, from the overall analysis, following remarks are obtained:

(a) CEEMDAN used as a preprocessing strategy with standalone models exhibits significant improvement vis-a-vis the performance of standalone models in terms of MAPE, RMSE and MAE, respectively.

(b) The performance of CEEMDAN based models is better in summer and autumn season because correlation between real and predicted GHI is better as compared to winter and monsoon season.

(c) Proposed model outperforms all CEEDMDAN based models with respect to performance criterion. The proposed model uses Genetic algorithm approach to select appropriate IMFs from the total number of data sub sets.

Table 2
Performance comparison between proposed model and non-CEEMDAN models

	Models	Winter	Spring	Summer	Monsoon	Autumn	Annual
	MAPE (%)						
	Naïve Predictor	5.10	6.39	3.71	6.81	4.81	5.36
	BPNN	4.21	5.20	3.10	6.16	4.11	4.55
	ELM	3.98	4.86	2.81	5.97	3.21	4.16
	RNN	3.73	4.14	2.65	5.61	3.19	3.86
	GRU	3.51	3.41	1.91	4.31	2.84	3.19
	LSTM	3.41	3.31	1.81	4.20	2.64	3.07
	Proposed Model	**2.20**	**2.96**	**0.91**	**3.91**	**1.21**	**2.23**
1-hr ahead solar GHI forecasting	**RMSE (W/m²)**						
	Naïve Predictor	4.42	5.91	3.30	6.31	4.33	4.85
	BPNN	3.56	4.63	2.91	5.91	3.54	4.11
	ELM	3.22	4.22	2.41	5.31	2.84	3.6
	RNN	3.13	4.10	2.23	4.12	2.35	3.18
	GRU	2.91	3.91	1.90	3.91	2.21	2.96
	LSTM	2.81	3.71	1.60	3.52	2.01	2.73
	Proposed Model	**1.51**	**2.31**	**0.21**	**2.41**	**0.83**	**1.45**
	MAE(W/m²)						
	Naïve Predictor	3.90	4.91	2.91	5.91	3.71	4.26
	BPNN	3.13	4.12	2.18	5.11	2.91	3.49
	ELM	2.61	3.81	1.81	4.61	2.13	2.99
	RNN	2.34	3.35	1.66	4.38	1.91	2.72
	GRU	2.10	2.10	0.41	3.10	1.61	1.86
	LSTM	2.01	1.99	0.39	3.01	1.51	1.78
	Proposed Model	**1.16**	**1.81**	**0.61**	**2.82**	**0.31**	**1.34**

Table 3
Performance comparison between proposed model and CEEMDAN models

	Models	Winter	Spring	Summer	Monsoon	Autumn	Annual
	MAPE (%)						
	CEEMDAN-BPNN	3.21	4.21	2.11	5.10	3.10	3.54
	CEEMDAN-ELM	2.91	3.81	1.81	4.91	2.21	3.13
	CEEMDAN-RNN	2.81	3.61	1.45	4.31	2.11	2.85
	CEEMDAN-GRU	2.71	3.41	1.21	4.12	1.81	2.65
	CEEMDAN-LSTM	2.51	3.26	1.10	4.02	1.76	2.54
	Proposed Model	**2.20**	**2.96**	**0.91**	**3.91**	**1.21**	**2.23**
1-hr ahead solar GHI forecasting	**RMSE (W/m²)**						
	CEEMDAN-BPNN	2.51	3.60	1.91	4.91	2.50	3.08
	CEEMDAN-ELM	2.20	3.21	1.41	4.32	1.81	2.59
	CEEMDAN-RNN	2.11	3.11	1.20	4.11	1.65	2.43
	CEEMDAN-GRU	1.97	2.91	0.91	3.20	1.21	2.04
	CEEMDAN-LSTM	1.81	2.71	0.61	3.02	1.09	1.84
	Proposed Model	**1.51**	**2.31**	**0.21**	**2.41**	**0.83**	**1.45**
	MAE(W/m²)						
	CEEMDAN-BPNN	2.10	3.10	1.19	4.10	1.99	2.49
	CEEMDAN-ELM	1.91	2.61	0.85	3.61	1.10	2.01
	CEEMDAN-RNN	1.84	2.21	0.67	3.31	0.91	1.78
	CEEMDAN-GRU	1.68	2.10	0.41	3.01	0.61	1.56
	CEEMDAN-LSTM	1.48	1.92	0.51	2.86	0.51	1.45
	Proposed Model	**1.16**	**1.81**	**0.61**	**2.82**	**0.31**	**1.34**

A. Gupta et al

Int. J. Renew. Energy Dev. 2022, 11(3), 736-750

|746

Case 3: Assessment of the chosen features

In this case, BiLSTM models: standalone Bi-LSTM, CEEMDAN based Bi-LSTM model include all statistical features and CEEMDAN-GA-BiLSTM model are compared as an evaluation of selected features. On the basis of loss function, the representative training process on three feature sets uses 200 epochs for the same Bi-LSTM network. When EEMD-GA was included the training loss decreased gradually from 0.00666 for standalone BiLSTM to 0.00444 for EEMD based BiLSTM model and 0.00262 when GA wrapper was included.

Table 4 indicates the results of standalone BiLSTM, CEEMDAN based BiLSTM and proposed model with respect to MAPE, RMSE and MAE performance criterion

Table 4
Performance comparison of Bi-LSTM models

	MAPE (%)						
	Models	**Winter**	**Spring**	**Summer**	**Monsoon**	**Autumn**	**Annual**
	BiLSTM	3.21	3.20	1.91	4.10	2.21	2.92
	CEEMDAN-BiLSTM	2.51	3.30	1.11	4.11	1.43	2.49
	Proposed Model	**2.20**	**2.96**	**0.91**	**3.91**	**1.21**	**2.23**
1-hr ahead solar GHI forecasting	**RMSE (W/m²)**						
	BiLSTM	2.57	3.31	1.61	3.61	1.85	2.59
	CEEMDAN-BiLSTM	1.90	2.72	0.90	3.02	0.94	1.89
	Proposed Model	**1.51**	**2.31**	**0.21**	**2.41**	**0.83**	**1.45**
	MAE(W/m²)						
	BiLSTM	1.82	2.09	0.31	3.12	1.10	1.68
	CEEMDAN-BiLSTM	1.43	2.11	0.10	2.91	0.21	1.35
	Proposed Model	**1.16**	**1.81**	**0.61**	**2.82**	**0.31**	**1.34**

Fig.7 Proposed model Performance for the summer and monsoon season
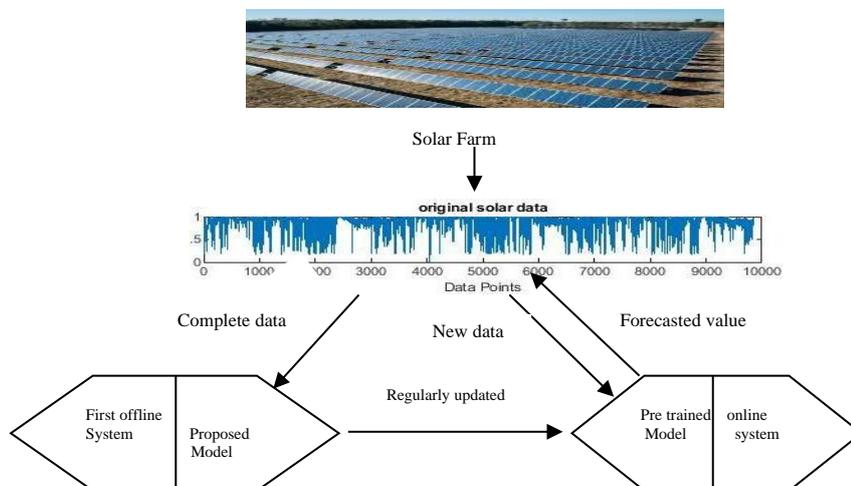
Fig.8 Steps of the distributed system in practical systems for the present framework

A. Gupta et al

Int. J. Renew. Energy Dev. 2022, 11(3), 736-750

| 747

5. Discussion

This research performs short term solar irradiance forecasting for the location of Delhi, India. Various experimental analyses are performed in this study to obtain precise model with improved forecasting accuracy. The prediction performance of the proposed model is compared with persistence model, standalone models (BPNN, ELM, GRU, LSTM and RNN) and CEEMDAN based models in order to demonstrate its superiority. Finally, based on features evaluation, the prediction results of proposed model are compared to standalone BiLSTM model and the CEEMDAN-BiLSTM model which consists all prospective features. From the results, it is clear that the CEEMDAN improves the forecasting accuracy of the standalone models. For a case of summer season, from the Table 2 to 3, it is observed that the CEEMDAN improved the RMSE (34.36% for BPNN, 41.49% for ELM, 46.18% for RNN and 52.10% for GRU and 61.87% for LSTM). However, in case of monsoon season, the accuracy decreased due to data instability of the season. But it is concluded that the CEEMDAN improved the forecasting performance of the standalone model. Similar observations can also be seen for MAPE and MAE. The proposed model uses GA as a feature extraction strategy over CEEMDAN based models. No doubt from the results, BiLSTM model outperforms all standalone models on all fronts. The lower RMSE, MAPE and MAE attained by BiLSTM prove its efficiency over other standalone models and encourages us to utilize this model for subsequent improvements. Therefore, GA with CEEMDAN process is applied on BiLSTM to prove the objective of the study. It is observed that GA with CEEMDAN credibly improves the forecasting performance of BiLSTM. The proposed model improves the results in terms of RMSE, MAPE and MAE as compared to all considered models. For a case of annual forecasting, the proposed model improves the RMSE (44.01% for BiLSTM; 23.28% for CEEMDA-BiLSTM), MAPE (23.63% for BiLSTM; 10.44% for CEEMDAN-BiLSTM) and MAE (20.23% for BiLSTM; 9% for CEEMDAN-BiLSTM) over BiLSTM and CEEMDAN-BiLSTM models. It is clear from this graphic analysis that the proposed model not only improves the RMSE, MAPE and MAE of standalone models but also improve the performance of CEEMDAN based models. These results prove its efficacy for real world application also, with minimum chances of error in the forecasting of solar irradiation.

Moreover, for a deeper examination of the findings, Fig.7 provides a graphic representation of real and predicted GHI for four consecutive days (2nd to 5th day) of summer and monsoon seasons. For clarity, only real and predicted GHI curve of suggested model is shown for selected seasons. It can be observed from fig 7 that those substantial fluctuations in real GHI generate a larger error in results. For example, smooth curve of summer season indicates clear environmental circumstances which are easily traceable by the model. On the other hand, monsoon season shows substantial fluctuations in real GHI due to existence of overcast or rainy days making it difficult for the model to trace resulting in maximum inaccuracies. From Figure 7, it can be deduced that if fluctuations in real GHI are higher than similarities existing between real and predicted GHI are lower. Similarly, resemblance between real and predicted GHI is higher when variance in real GHI is lower. However, with in a tolerated range of error, suggested model also faces a number of ambiguities associated withgenuine GHI. As a result of these findings, the suggested model is a good forecast model for stable as well as for unstable seasons.

Suggested short-term solar irradiation prediction may be implemented in practical systems utilizing a distributed system (Fig.8), with one system training offline models and the other making online forecasting. The impact of a single future data point on the CEEMDAN spectrum may be small for large scale solar history. As a result, the online system may generate accurate predictions in a short amount of time using pre-trained models. Furthermore, as new records are received, the offline system can update the model at the same time. The model will be transmitted back into the web application for better operations as the volume of data grows significantly.

For a detailed discussion of proposed model forecasting performance, Table 5 displays the forecasting skill (%) performance of BiLSTM, CEEMDAN-BiLSTM and proposed model for annual RMSE, MAPE and MAE with respect to persistence model. Forecast skill determines the association between predicted or prediction value to the observed value.

Table 5

Forecast skill (%) of proposed model on annual basis

Model	FS_{RMSE}	FS_{MAPE}	FS_{MAE}
BiLSTM	39%	38%	49%
CEEMDAN-BiLSTM	50%	45%	55%
Proposed Model	**58%**	**49%**	**59%**

Table 6

Performance comparison with previously published models

Author and year of publication	Model	Place	Time horizon	MAPE (%)	RMSE(W/m²)	MAE(W/m²)
Zang et al. 2020	CNN-LSTM	Texas, USA	1-hr	-	69.26	37.20
Huang et al. 2020	LSTM-MLP	Colorado, USA	1-hr	-	75.22	36.90
Kumari et al. 2021	XGBF-DNN	New Delhi, India	1-hr	-	51.35	-
Singla et al. 2021	WT-BiLSTM	Ahmadabad, India	24-hr	45.61	6.48	-
This work	CEEMDAN-GA-BiLSTM	New Delhi	1-hr	2.23	1.45	1.34

A. Gupta et al

Int. J. Renew. Energy Dev. 2022, 11(3), 736-750

| 748

Finally, Table 6 shows the performance of the proposed model compared to previously published models in terms of MAPE, RMSE and MAE. The proposed model performance is excellent over the latest developed models. Table 6 presents that the proposed model offers a percentage improvement in RMSE (77.62%), MAPE (95.11%) respectively over WT-BiLSTM model (Singla *et al.* 2021). Likewise significant improvement in RMSE (97.17%) is shown by the proposed model against XGBF-DNN (Kumari et al. 2021). In addition to this, the proposed model exhibits a remarkable percentage improvement in terms of RMSE (98.07%) and MAE (96.36%), respectively over LSTM-MLP (see Huang *et al.* 2020). Moreover, in literature various authors have used different techniques to forecast solar GHI using a combination of WT-BiLSTM to forecast solar GHI(see Singla *et al* 2021).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 data set. Similar problem occurred when using variational mode decomposition-based method. Implement XGBF-DNN model to forecast solar GHI and measure proposed model performance using RMSE ($51.35W/m^2$) which is the lowest as compared to current proposed work because Extreme gradient boosting algorithm (XGBF) does not perform well on unstructured data. So, the proposed work uses CEEMDAN (advance version of EEMD) and Bi-LSTM (advance type of LSTM) to predict solar GHI. CEEMDAN removes the Gaussian white noise added with the EEMD may not be cancelled after reconstruction, while BiLSTM process information in both directions (forward and backward), so 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 is a better alternative for forecasting solar GHI for practical solar power system.

4. Conclusions

In this study, an ensemble deep learning-based architecture is introduced as a method of predicting solar irradiation using a dataset of solar history. When time resolution and recording period of solar dataset increase, then there is expansion of non-linearity in time series data. The number of IMFs will increase dramatically as a result of using CEEMDAN approach on increased time series data. It means more IMF components would result in more untrained data resulting in rising of overall training cost. Deep learning model is employed to predict IMF components and forecasting error of each component which them adds up to get final error, affecting the prediction accuracy of the model. To address this problem and to improve forecasting accuracy three major algorithms make up the proposed model: CEEMDAN, GA and LSTM. In the First step, CEEMDAN is used as a pre-processing technique to rectify and extract the inherent characteristics of time series data to obtain intrinsic mode functions. The well tuned deep learning model and the intuitively picked feature set are synchronized through an optimization method using the paired GA-BiLSTM technique. The procedure is totally automated when using the suggested framework, and no preceding functions are required. Because of the process's simplicity, it is ideal for applications that require end-to-end functionality. The model's feasibility and effectiveness are thoroughly validated using analytic research. The suggested method demonstrates its amazing superiority over conventional models using assessment criteria such as MAE (W/m^2), RMSE(W/m^2) and MAPE(%). To begin with, the current model prediction accuracy is increased by 44.96 percent on an average as compared to non-CEEMDAN models. On the other hand, when comparing with CEEMDAN approach with other learning prototypes there is a substantial improvement in prediction accuracy of 26.2 percent on an average. Furthermore, when comparing the outcomes of the same teaching method with multiple feature sets, the suggested technique is even more powerful from two perspectives: First and foremost it should be possible to use the GA wrapper for selecting features. The length of the model input data is reduced to around a 2/3 of the total population set of features, making it more compact and robust to data fluctuation. Moreover, when compared to all feature and non-feature models, it exceeds them in terms of prediction precision across a wide range of evaluation criteria, with increases by 35.58 percent and 21.74 percent, respectively. From all results, it is confirmed that proposed framework is good forecasting model from all perspectives.

However, while constructing the proposed model, some challenges are faced by the researcher such as: selected features extraction and simulation time. So, in future there is scope of developing and enhancing present research by employing several searching algorithms such as: firefly algorithm, particle swarm optimization, etc. that investigate and compare as feature selection strategy for present framework. Next, use of meteorological parameters: temperature, humidity etc. with selected features set and to check its effect on model performance. Last but not the least, larger scaled solar dataset from other solar farms with different specifications can be considered in order to validate the present framework on a wider global scale.

Abbreviations

ANFIS: Adaptive Neuro Fuzzy Inference System, ANN: Artificial Neural Network, AR: Auto Regression, BPNN: Back Propagation Neural Network, CNN: Convolution Neural Network, EEMD: Ensemble Empirical Mode Decomposition, ELM: Extreme Learning Machine, EMD: Empirical Mode Decomposition, FT: Fourier Transform, GA: Genetic Algorithm, GRU: Gate Recurrent Unit, HHT: Hilbert Huang Transform, IMF: Intrinsic Mode Function, LSTM: Long Short Term Memory, MAPE: Mean Absolute Percentage Error, MAE: Mean Absolute Error, NWP: Numerical Weather Forecasting, RMSE: Root Mean Square Error, RNN: Recurrent Neural Network, SSA: Singular Spectrum Analysis, WT: Wavelet Transform

Declarations
Not Applicable

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

A. Gupta et al

Int. J. Renew. Energy Dev. 2022, 11(3), 736-750

| 749

References

Al-Hajj, R., Assi, A., Fouad, M. M.(2018) Forecasting Solar Radiation Strength Using Machine Learning Ensemble. *In Proceedings of the 7th IEEE International Conference on Renewable Energy Research and Applications (ICRERA),* Paris, France, 14–17 October 2018; pp. 184–188.

Al-Hajj, R., Assi, A., Fouad, M. (2021) Short-Term Prediction of Global Solar Radiation Energy Using Weather Data and Machine Learning Ensembles: A Comparative Study. *J. Sol. Energy Eng*. 8, 1–38.

Bedi, J., Toshniwal, D. (2019) Deep learning framework to forecast electricity demand. *Appl Energy* 238:1312–1326. https://doi.org/10. w1016/j.apenergy.2019.01.113

Chen, C.R., Ouedraogo, F.B., Chang, Y.M.; Larasati, D.A.; Tan,S.W. (2021) Hour Ahead Photovoltaic Output Forecasting Using Wavelet ANFIS. *Mathematics*,9,2438.https://doi.org/10.3390/math9192438.

Dumitru C-D, GligorA, Enachescu C 9(2016) Solar photovoltaic energy production forecast using neural networks. *Procedia Technol* 22: 808-815. https://doi.org/10.1016/j.protcy.2016.01.053

Fang, J., Wu, L., Zhang, F., Cai, H., Zeng, W., Wang, X., Zou, H.(2019) Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: A review and case study in China. *Renew. Sustain. Energy Rev*, 100, 186–212.

Fischer T, Krauss C (2018) Deep learning with long short-term memory networks for financial market predictions. *Eur J Oper Res* 270(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

Gao, B., Huang, X., Shi, J., Tai, Y., Xiao, R. (2019) Predicting day-ahead solar irradiance through gated recurrent unit using weather forecasting data. *J.Renew Sustain Energy* 11(4), 043705.https://doi.org/10.1063/1.5110223

Gao, B., Huang, X., Shi, J., Tai, Y., Zhang, J. (2020) Hourly forecasting ofsolar irradiance based on CEEMDAN and multi-strategy CNN-LSTM neural networks. *Renew Energy* 162:1665–1683. https://doi.org/10.1016/j.renene.2020.09.141

Gupta, A., Gupta, K., Saroha, S. (2022) A Comparative Analysis of Neural Network-Based Models for Forecasting of Solar Irradiation with Different Learning Algorithms. In: Khosla A., Aggarwal M. (eds) Smart Structures in Energy Infrastructure. Studies in Infrastructure and Control. Springer, Singapore. https://doi.org/10.1007/978-981-16-4744-4_2

Gupta, A., Gupta, K., Saroha, S. (2022) Single Step-Ahead Solar Irradiation Forecasting Based on Empirical Mode Decomposition with Back Propagation Neural Network. In: Gupta O.H., Sood V.K., Malik O.P. (eds) Recent Advances in Power Systems. Lecture Notes in Electrical Engineering, vol 812. Springer, Singapore. https://doi.org/10.1007/978-981-16-6970-5_10

Gupta, A., Gupta, K., Saroha, S. (2022) Solar Energy Radiation Forecasting Method. In: Agarwal P., Mittal M., Ahmed J., Idrees S.M. (eds) Smart Technologies for Energy and Environmental Sustainability. Green Energy and Technology. Springer, Cham. https://doi.org/10.1007/978-3-030-80702-3_7

Gupta, A., Gupta, K., Saroha, S.(2021) A Review and Evaluation of Solar Forecasting Technologies: Materials today proceedings 2021, Volume 47, Part 10, 2021, Pages 2420-2425. https://doi.org/10.1016/j.matpr.2021.04.491

Gupta, A., Gupta, K., Saroha, S.(2020) Solar Irradiation Forecasting Technologies: A Review: Strategic Planning for Energy and the Environemnt.2020: Vol 39 Iss 3-4 2020. https://doi.org/10.13052/spee1048-4236.391413

Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. *Neural Comput* 9(8),1735-80

http://delhitourism.gov.in/delhitourism/aboutus/seasons_of_delhi.jsp

Huang, C., Wang, L., Lai, L.L. (2019) Data-driven short-term solar irradiance forecasting based on information of neighboring sites. *IEEE Trans Ind Electron* 66(12), 9918–9927. https://doi.org/10.1109/TIE.2018.2856199

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q. (1998) The empirical mode decomposition and the Hilbert transform for nonlinear and non-stationary time series analysis. *Proc A* ,454(1971), 903-95.

Huimin, Z., Meng, S., Wu, D., Xinhua, Y.(2016) A new feature extraction method based oneemd and multi-scale fuzzy entropy for motor bearing. *Entropy* 19(1),14.

Jahani, B., Mohammadi, B. (2019) A comparison between the application of empirical and ANN methods for estimation of daily global radiation in Iran. *Theor Appl Climatol* 137(1-2), 1257-1269. https://doi.org/10.1007/s00704-018-2666-3

Kumari, P., Toshniwal, D. (2021). Extreme gradient boosting and deep neural network based ensemble learning approach to forecasts hourly solar irradiance. *J. Clean Prod* 279,123285.https://doi.org/10.1016/j.jclepro.2020.123285

Liu, H., Mi, X., Li, Y. (2018) Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, *Energy Convers Manage* 159, 54–64.

Monjoly, S., André, M., Calif, R., Soubdhan, T. (2017). Hourly forecasting of global solar radiation based on multiscale decomposition methods: A hybrid approach. *Energy*, 119(), 288–298. https://doi:10.1016/j.energy.2016.11.061

Olatomiwa, L., Mekhilef, S., Shamshirband, S., Mohammadi, K., Petkovic´, D., Sudheer, C. A. (2015) support vector machine–firefly algorithm-based model for global solar radiation prediction. *Sol. Energy*, 115, 632–644.

Perez, R., Kivalov, S., Schlemmer, J., Hemker, K., Hoff, T.E. (2012) Shortterm irradiance variability: Preliminary estimation of station pair correlation as a function of distance. *Solar Energy* 86(8), 2170–2176. https://doi.org/10.1016/j.solener.2012.02.027

Piri, J., Shamshirband, S., Petkovic´, D., Tong, C.W., Rehman, M.H.(2015) Prediction of the solar radiation on the earth using support vector regression technique. *Infrared Phys. Technol.*, 68, 179–185.

Prasad, R., Ali, M., Kwan, P., Khan, H. (2019). Designing a multi-stage multivariate empirical mode decomposition coupled with ant colony optimization and random forest model to forecast monthly solar radiation. *Applied Energy*, 236, 778–792. doi:10.1016/j.apenergy.2018.12.034

Qin, Q., Lai, X., Zou, J. (2019) direct multistep wind speed forecasting using LSTM neural network combining EEMD and fuzzy entropy. *Appl Sci*, 9(1).

Qing, X., Niu, Y. (2018) hourly day ahead solar irradiance predictions using weather forecasts by LSTM. *Energy* 148, 461–468.https://doi.org/10.1016/j.energy.2018.01.177

Richardson, D.S., Cloke, H.L., Pappenberger, F (2020) Evaluation of the consistency of ECMWF ensemble forecasts. *Geophys Res Lett* 47(11). https://doi.org/10.1029/2020GL087934

Shadab, A., Ahmad, S., Said, S. (2020) Spatial forecasting of solar radiation using ARIMA model. *Remote Sens Appl Soc Environ* 20, 100427. https://doi.org/10.1016/j.rsase.2020.100427

Shang, C., Wei, P. (2018). Enhanced support vector regression based forecast engine to predict solar power output. *Renew. Energy*, 127, 269–283.

Sharma, A., Kakkar, A.(2018) Forecasting daily global solar irradiance generation using machine learning. *Renew. Sustain. Energy Rev.* 2018, 82, 2254–2269.

Singla, P., Duhan, M., Saroha, S. (2021) A comprehensive review and analysis of solar forecasting techniques. *Front Energy*. https://doi.org/10.1007/s11708-021-0722-7

Singla, P., Duhan, M. & Saroha, S. (2022) An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and BiLSTM deep learning network. *Earth Sci Inform* 15, 291–306 . https://doi.org/10.1007/s12145-021-00723-1

Vasylieva,T., Lyulyov,O., Bilan, Y., Streimikiene, D.(2019) Sustainable economic development and greenhouse gas emissions: The dynamic impact of renewable energy consumption, GDP, and corruption. *Energies*, 12, 3289.

A. Gupta et al

Int. J. Renew. Energy Dev. 2022, 11(3), 736-750

| 750

Wang, F., Yu, Y., Zhang, Z., Li, J., Zhen, Z., Li, K. (2018) Wavelet decomposition and convolution LSTM networks based improved deep learning model for solar irradiance forecasting. *Appl Sci* 8(8), 1286. https://doi.org/10.3390/app8081286

Wu, Z., Huang, N.E. (2009) Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv Adapt Data Anal*, 1(01),1–41

Yagli, G.M., Yang, D., Srinivasan, D. (2019) Automatic hourly solar forecasting using machine learning models. *Renew. Sustain. Energy Rev.*, 105, 487–498.

Zang, H., Cheng, L., Ding, T., Cheung, K.W.,Wei, Z., Sun, G. (2020a) Day ahead photovoltaic power forecasting approach based on deep convolution neural networks and meta *Int J Electr power Energy Syst* 118, 105790. https://doi.org/10.1016/j.ijepes.2019.105790

Zang, H., Liu, L., Sun, L., Cheng, L., Wei, Z., Sun, G. (2020b) Short-termglobal horizontal irradiance forecasting based on a hybrid CNNLSTM model with spatiotemporal correlations. *Renew Energy* 160, 26–41. https://doi.org/10.1016/j.renene.2020.05.150

Zendehboudi, A., Baseer, M.A., Saidur, R. (2018). Application of support vector machine models for forecasting solar and wind energy resources: A review. *Journal of Cleaner Production,* 199(), 272–285. https://doi:10.1016/j.jclepro.2018.07.164

Zeng, J., Qiao, W. (2013) Short-term solar power prediction using a support vector machine. *Renew Energy* 52, 118–127. https://doi.org/10.1016/j.renene.2012.10.009

Zhang, X., Zhang, Q., Zhang, G., Nie, Z., Gui, Z.(2018) A Hybrid Model for Annual Runoff Time Series Forecasting Using Elman Neural Network with Ensemble Empirical Mode Decomposition. *Water* 10, 416. https://doi.org/10.3390/w10040416