Energy and AI 4 (2021) 100060

Contents lists available at ScienceDirect

Energy and AI

journal homepage: www.elsevier.com/locate/egyai

Review

# A review and taxonomy of wind and solar energy forecasting methods based on deep learning

Ghadah Alkhayat[a], Rashid Mehmood[b,*]

[a] Department of Computer Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[b] High Performance Computing Centre, King Abdulaziz University, Jeddah 21589, Saudi Arabia

## HIGHLIGHTS

- Taxonomy of wind and solar energy forecasting is rich with deep methods.
- Most works used datasets from locations in China (39), USA (31), and Australia (11).
- Hybrid forecasting and then Recurrent Neural Networks are the most used models.
- Convolutional neural networks are the third most used models in the field.
- Probabilistic and multistep ahead forecasting methods are gaining attention.

## ARTICLE INFO

*Article history:*
Received 5 January 2021
Received in revised form 16 February 2021
Accepted 24 February 2021
Available online 6 March 2021

*Keywords:*
Deep learning
Renewable energy forecasting
Solar energy
Wind energy
Taxonomy
Hybrid methods

## ABSTRACT

Renewable energy is essential for planet sustainability. Renewable energy output forecasting has a significant impact on making decisions related to operating and managing power systems. Accurate prediction of renewable energy output is vital to ensure grid reliability and permanency and reduce the risk and cost of the energy market and systems. Deep learning's recent success in many applications has attracted researchers to this field and its promising potential is manifested in the richness of the proposed methods and the increasing number of publications. To facilitate further research and development in this area, this paper provides a review of deep learning-based solar and wind energy forecasting research published during the last five years discussing extensively the data and datasets used in the reviewed works, the data pre-processing methods, deterministic and probabilistic methods, and evaluation and comparison methods. The core characteristics of all the reviewed works are summarised in tabular forms to enable methodological comparisons. The current challenges in the field and future research directions are given. The trends show that hybrid forecasting models are the most used in this field followed by Recurrent Neural Network models including Long Short-Term Memory and Gated Recurrent Unit, and in the third place Convolutional Neural Networks. We also find that probabilistic and multistep ahead forecasting methods are gaining more attention. Moreover, we devise a broad taxonomy of the research using the key insights gained from this extensive review, the taxonomy we believe will be vital in understanding the cutting-edge and accelerating innovation in this field.

## 1. Introduction and related work

The increase in international interest in renewable energy sources and the expansion of integrating such sources into the electrical grid around the globe has attracted many researchers to focus on this field [1–3]. Popular applications of smart energy systems include load forecasting, renewable energy output forecasting, energy pricing, power quality disturbances detection, and fault detection on power systems and equipment. Renewable energy output forecasting, especially wind and solar energies, has gained much attention recently because of its significant impact on making decisions related to operating and managing power systems. Accurate prediction of renewable energy output is vital to ensure grid reliability and permanency and reduce the risk and cost of the energy market and energy systems. Not only the power plants and the grid operators will benefit from such forecasting, but also energy traders and policymakers [4–6].

Due to the variable nature of weather, there will always be instability of the energy output from solar and wind energies. Thus, their output prediction is difficult and requires advanced methods. The techniques used for this task can be classified into four categories: physical methods, statistical models, artificial intelligence techniques, and their hybrid methods [1,6,7]. Physical methods or Numerical Weather Prediction (NWP) models are mathematical models that simulate the atmospheric dynamics according to physical and mechanical principles. Since they depend on computer simulation, they require extensive computer resources and thus are used for long term forecasting horizon [8]. On the other hand, statistical models are used to discover the mathematical

* Corresponding author.
*E-mail addresses:* grasheedalkhyat@stu.kau.edu.sa (G. Alkhayat), RMehmood@kau.edu.sa (R. Mehmood).

https://doi.org/10.1016/j.egyai.2021.100060
2666-5468/

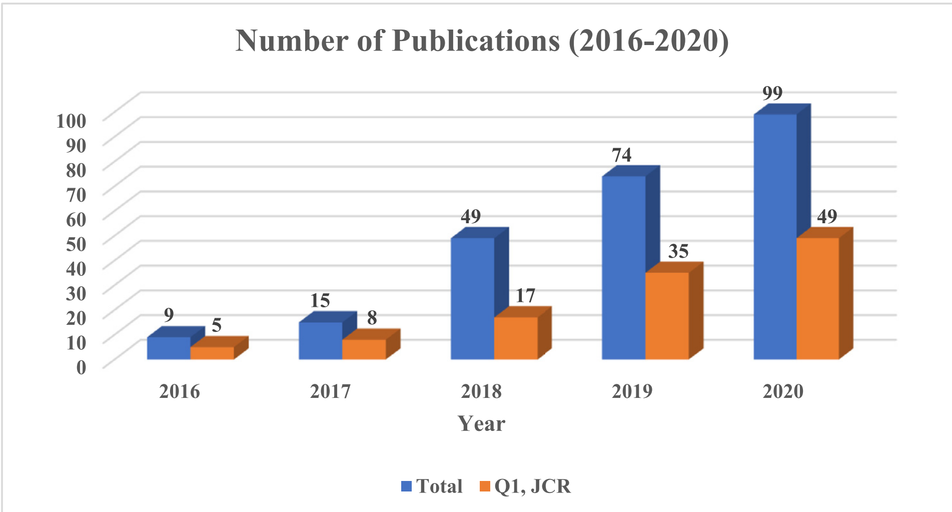


**Fig. 1.** Number of Publications in solar and wind energy forecasting using deep learning (Web of Science: 2016–2020).

**Table 1**
The survey papers in renewable energy forecasting.

| Review paper | ML | DL | All applications | Forecasting only | Wind energy | Solar energy | Published |
|---|---|---|---|---|---|---|---|
| Wang et al. [12] | √ | √ | | √ | | √ | April 2020 |
| Ahmad et al. [2] | √ | √ | | √ | | √ | February 2020 |
| Ozcanli et al. [4] | | √ | √ | | √ | √ | February 2020 |
| Ahmad et al. [3] | √ | | | √ | √ | √ | January 2020 |
| Mellit et al. [13] | √ | √ | | √ | | √ | January 2020 |
| Shamshirband et al. [5] | | √ | | √ | √ | √ | November 2019 |
| Wang et al. [1] | | √ | | √ | √ | √ | July 2019 |
| Liu et al. [14] | √ | √ | | √ | √ | | May 2019 |
| Mosavi et al. [6] | √ | | √ | | √ | √ | April 2019 |
| Akhter et al. [11] | √ | | | √ | | √ | March 2019 |
| Marugán et al. [15] | √ | | √ | | √ | | July 2018 |
| Voyant et al. [10] | √ | | | √ | | √ | January 2017 |

relationship between inputs and outputs assuming linear relationships. They are widely used in the literature, but their performance did not meet expectations because they are not suitable for discovering nonlinear relationships [1].

Artificial intelligence methods including Machine Learning (ML) algorithms and recently Deep Learning (DL) models have drawn researchers' attention due to their ability to discover nonlinear relationships and superior performance. Deep learning specifically has a promising future because of its generalization capability and unsupervised feature learning. It has already achieved huge success in many applications, such as image recognition, classification tasks, and natural language processing [1,5,9]. Solar and wind energy forecasting is not an exception where the number of publications reporting deep learning-based methods has rapidly increased in the last few years. This is evidenced in Fig. 1 that plots the number of papers on deep learning-based solar and wind energy research found in the Web of Science database for the period 2016–2020 (last five years). The figure shows that most of the papers in Web of Science are published in the last two years with 99 out of the total 246 papers published in 2020 alone (see the blue bars).

### *1.1. Aim, novelty and contributions*

This rapidly growing field necessitates survey papers that could provide insights into the existing research and facilitate further research and development in the area. It is no surprise that several literature review papers have been produced during the last few years on machine and deep learning methods in renewable energy forecasting and applications (see Table 1: these review papers were found through a Google Scholar search). One of these review papers has focussed on machine learning-based renewable energy applications (e.g., electricity pricing, energy system fault detection, economic dispatch, and energy forecasting) [6], another paper has focussed only on renewable energy forecasting application [3], while some other review papers have focussed solely on solar energy forecasting [2,10–13], or wind energy forecasting [14,15].

A total of three review papers have been published on renewable energy works that use deep learning exclusively. Ozcanli et al. [4] have reviewed works on all deep learning-based renewable energy applications, while two of these three papers have reviewed deep learning-based works on renewable forecasting application only [1,5]. Note that four of the review [2,12–14] are on machine learning including deep learning and these are excluded from the list of review papers on deep learning alone and were mentioned in the previous paragraph.

Table 1 shows that there are only five papers on machine learning or deep learning that were published in 2020, mostly published in January or February 2020, the latest among these was published in April 2020 and this paper focusses on solar energy alone. Considering the time required for review and publication, and based on our knowledge of these papers, none or very few 2020 works are reviewed in these papers. This clearly establishes the need for a review on deep learning-based works on renewable energy forecasting.

Motivated by this research gap, this paper provides a review of deep learning-based solar and wind energy forecasting research of the last five years discussing extensively the data and datasets used in the reviewed works, the data pre-processing methods, deterministic and probabilistic methods, and evaluation and comparison methods. Significant efforts are gone into summarizing the core characteristics of all the reviewed works and presenting these in tabular forms. We were aspired to add more information to the tables, however, due to the tables becoming unmanageable, we had to limit the information and present it in short forms. The existing survey papers mostly have given tabular summaries of a few selected papers. The tabular summaries in our paper would be useful, for instance, for the readers who wish to find specific information about a paper and compare it with other works. Moreover, we have carefully analyzed the literature to elicit current challenges in the field and future research directions.

This extensive review provides key insights into the state-of-the-art on the topic and has also allowed us to propose a broad taxonomy of deep learning-based solar and wind energy forecasting research. We have not seen such a broad taxonomy of deep learning-based solar and wind energy forecasting research and we believe that it will be vital in classifying and comparing works on the topic ultimately accelerating innovation in this field.

Since there are a total of 246 papers published during 2016–2020, to keep the length of the paper within limits, we have only included in this review the papers that were published in journals ranked in the first quartile (Q1) of Journal Citation Reports (see the orange bars in Fig. 1). Future work will look into incorporating other Web of Science papers in the review and the taxonomy.

The rest of the paper is organized as follows. Section 2 proposes a taxonomy of deep learning-based solar and wind energy forecasting research. Section 3 reviews the data and datasets used in forecasting while Section 4 examines the data preprocessing methods. Section 5 discusses deterministic forecasting models based on the used architectures including Convolutional Neural Network (CNN) based models (Section 5.1), Recurrent Neural Network (RNN) based models (Section 5.2), Stacked Autoencoder (SAE) based models (Section 5.3), Deep Belief Network (DBN) based models (Section 5.4), other deep models (Section 5.5), and hybrid and ensemble models (Section 5.6). Section 6 is devoted to reviewing probabilistic forecasting methods. Section 7 explores the evaluation and comparison methods used in the reviewed research. Section 8 provides a discussion of the research in this field of deep learning-based solar and wind energy forecasting and identifies the challenges and future research directions. Finally, Section 9 concludes the paper.

## 2. Taxonomy of wind and solar energy forecasting using deep learning

We propose in this paper a taxonomy of wind and solar energy forecasting methods using deep learning (see Fig. 2). This taxonomy covers the forecasting horizon, approach, data used, preprocessing methods, deep learning methods, hybrid methods, optimization methods, and evaluation methods (we call them "Level 1″ categories or taxonomy).

The forecast horizon includes four categories: ultra-short-term, short-term, medium-term, and long-term. The forecasting approach could be deterministic or probabilistic targeting the next time step or multi-steps. The data used for forecasting might be spatial, time series, or sky images. It could be the historical values of the wind speed or wind power for wind energy forecasting and solar power or solar irradiance for solar energy forecasting. This data could be used with or without other meteorological data. Data preprocessing methods include data normalization, data imputation, outlier treatment, changing data resolution, data transformation, data augmentation, correlation analysis, data clustering, data modeling as a graph or grid, data decomposition, and features selection.

The deep learning models that are frequently used are CNN, RNN, SAE, DBN, and other deep models. These deep learning methods might be combined with data decomposition methods, feature extraction methods, or data correction methods in hybrid models. Optimization techniques might be used for hyperparameters or parameters tuning, avoiding overfitting, or for accelerating training speeds. The forecasting model's evaluation methods involve using evaluation metrics, benchmarking models, computation time comparison, statistical testing, weather types comparison, input timesteps and data resolution comparison, data fusion, and decomposition methods comparison. Fig. 2 visualizes this taxonomy.

The taxonomy will be elaborated in the rest of the paper as we embark on discussing the reviewed works in detail. Each of the Level 1 categories are discussed in separate sections except horizon, approach and optimization that are discussed in the sections of forecasting methods.

## 3. Data and datasets

The data sets used to train and test the forecasting models included in this review were collected from different locations in 19 countries around the world. Table 2 shows the total number of references for each dataset location. In direct wind and solar energy output prediction, historical power output data collected from specific wind or solar farm is used. Almost half of the studies included in this paper are for direct power prediction. It is noted that in almost 60% of these studies, in addition to the historical power data, meteorological data, such as temperature, humidity, precipitation, air pressure, and wind direction, are used to improve the model accuracy. On the other hand, when the wind or solar farms are new and historical power data is not available or in case researchers want to prepare a feasibility study for a potential place of wind or solar farm, they predict the energy output indirectly by predicting wind speed or solar radiation. The power is then calculated according to specific formulas. This indirect method is more flexible. For example, since the generated electricity depends on wind turbine specifications, such as the hub height and the pitch angle of the blade, predicting wind speed can be adjusted accordingly. It is noted in [16] that wind speed is easier to predict than wind power because the latter depends on the wind turbine's specific features. Also, with indirect methods, other meteorological data in addition to wind speed and solar irradiation can be used as inputs to further improve the forecasting models. Using meteorological variables is especially helpful when the data is irregular because it helps to mitigate the irregularity effect on the model prediction performance. For example in [17], Zhang et al. used only the historical PV power data as input to the Autoregressive Integrated Moving Average (ARIMA) model to predict the regular component, while for the irregular component, the historical PV power data in addition to solar irradiance intensity and air temperature are fed into DBN-based model to generate the forecasting.

According to [2], the highly correlated variables with PV power are solar irradiance, air temperature, and dew point while relative humidity and cloud type are somewhat negatively correlated with it. On the other hand, blade pitch angle affects wind power prediction higher than wind speed and wind shear according to the results in [18]. It was noted also that ambient temperature, nacelle orientation, and yaw error have a small impact on the wind power prediction and can be removed from inputs. In [19], the four most important features for GHI prediction according to Pearson correlation values are temperature, clear sky index, relative humidity, and hour of the day while pressure, wind speed and direction are less important. However, the correlation between power output and meteorological data depends on the data location and its climate. Therefore, researchers should include a feature selection method to validate the importance of the variables as a preprocessing step before training their models. Also, this helps them to determine how many previous values (time steps) of each feature should be included in the inputs. For example, Zhu et al. [20] used the top-down relevant feature search algorithm in their model to determine the most important thirty features. According to the results, only average wind speed values for the previous four-time steps appear in the list while peak wind speed values for the previous six-time steps are included. Also, this method helps to identify less important features. For example, the accumulated precipitation value for the previous one-time step is ranked number 15 in the list and no more values of this feature are included.

Usually, the type of inputs used for renewable energy forecasting is time series data, but sometimes sky images are used for ultrashort


**Fig. 2.** Taxonomy of deep learning-based solar and wind energy forecasting.

**Table 2**
Number of articles for each dataset location.

| Location | Count | Location | Count | Location | Count | Location | Count | Location | Count |
|---|---|---|---|---|---|---|---|---|---|
| **China** | 39 | **UK** | 5 | **India** | 1 | **Philippines** | 1 | **France** | 1 |
| **USA** | 31 | **S. Korea** | 4 | **Belgium** | 1 | **Algeria** | 1 | **Brazil** | 1 |
| **Australia** | 11 | **Europe** | 4 | **Singapore** | 1 | **Cape Verde** | 1 | **Poland** | 1 |
| **Spain** | 7 | **Canada** | 3 | **Ireland** | 1 | **Belgium** | 1 | **Germany** | 1 |
| **Taiwan** | 5 | **Turkey** | 1 | **Egypt** | 1 | **Greece** | 1 | **Iran** | 1 |

prediction horizon only as in [21,22]. Spatial data is also helpful for predicting renewable power in a specific region or for multiple sites as proposed in [16,23–28].

## 4. Data preprocessing methods

### 4.1. Data normalization and denormalization

Normalization is an essential step when numerical values have different scales, which is the case with forecasting inputs. Ignoring this step, especially with gradient descent-based algorithms, hinders their learning process, and slow up their convergence speed towards the minima, thus, distort the performance results. Forty-four papers included in this review reported using normalization. It might have been used in the remaining papers, but not mentioned. The normalization technique used in almost all the papers is Min-Max scaling, which transfers the data into a range between 0 and 1. In [29], Gensler & Raabe used Min-Max scaling for all NWP data while they used the nominal output capacity of each PV facility to normalize the measured power output data. The reason is the power output data was collected from 21 facilities of different sizes and this normalization allows fair comparison. Moreover, in [30], Ju et al. used Min-Max scaling for power data and Zero-Mean normalization for other input variables that have positive and negative values, such as temperature and wind direction, to avoid losing the direction information. In [31], the sine and cosine values are used to normalize the wind direction while Min-Max method is used with other input variables. Transforming the values into a range between −1 and 1 is only reported in [32]. The logical postprocessing step after obtaining the model's prediction results is to transfer them back to their original scale to improve their readability. This process is called denormalization or anti-normalization.

### 4.2. Handling wrong or missing values and outliers

In most of the papers, it is mentioned that records contained wrong values or outliers were removed. However, replacing wrong values that are beyond the limit with the maximum value of that input variable is reported in [33]. Also, missing values in this paper were filled using the linear interpolation method, in which an estimation of the value is calculated using the previous and the next value. The same method is used in [34] as well. Filling missing values with the mean value of previous years is reported in [35]. In [36], negative solar power values were replaced by zeros. In [37], the equivalent probabilistic power curve method is used to correct the wrong wind power values. In [38], wind speed values were removed for times when wind turbines were not operational. In [39], missing values were replaced with the previous hour values.

Removing outliers is necessary to improve prediction. In [40], Sharadga et al. used the Hampel filter for this purpose and compared the prediction accuracy before and after this step to show its benefit. This filter replaces the outlier with the median value according to a given window size. For more about this method, refer to [41]. In [42], any value beyond three times the standard deviation is considered an outlier. Then, missing and outlier values were cleaned using quadratic spline interpolation. This method is explained in [43]. In [44], outliers were replaced with previous measurements. To determine which values are outliers, Lima et al. depended on ambient temperature and solar irradiance data taken in the same study location. In [18], outliers in wind power data were detected using Isolation Forest, which is a tree ensemble-based algorithm. Then, these outliers were automatically removed. The data before and after applying this method was visualized in the paper, which shows clearly the improvement of its quality. Moreover, in [45] two outlier detection methods were compared: Elliptic Envelope and Isolation Forest. Experiments show that the accuracy of the proposed model improved when Isolation Forest was used because wind power prediction data cannot be assumed to be Gaussian.

### 4.3. Changing the data resolution

When the available dataset is collected with high resolution, for example, every second or every minute and the model will be developed for a longer forecasting horizon, such as one hour ahead or 24 h ahead, then data resolution should be changed. One common method is averaging, which many authors claim it helps to smooth the data and reducing fluctuations. However, there is a concern that this method might let useful information to be lost [46]. To address this issue, Liu & Chen in [47] used two different methods to transform the resolution of data from 3 s to 1 min separately: the averaging method and the stacked sparse autoencoder. This integration of resolution transformation and feature extraction helps in reducing the resolution while keeping the representative information at the same time. The averaging method is used in [40] to transform data resolution from 15 min to 1 h. It is also used in [48] to change resolution from 2 min to 10 min, half an hour, and an hour, while in [49] it is used to change data resolution from one hour to one day. In [31], the Linear interpolation method is used to change the NWP data resolution from 1 h to 10 min.

Not only time-series data resolution affect forecasting, but also image resolution can have an impact as well. In [21], Sun et al. conducted many experiments to study the sensitivity of different image resolutions on PV output prediction.

### 4.4. Deleting inactive times records

Obviously, solar power reading during night hours would be zero, and the common practice is to delete these records to improve the forecasting accuracy. However, night hours change from one season to another and this should be considered before elimination. In [33], solar irradiance data is visualized before and after removing night hours. When the dataset is collected from several locations, variation in sunrise time might cause a problem. To deal with this issue and avoid deleting many records, zero values of solar irradiation were transformed to negative numbers in [50], then these negative numbers were converted back to zero after training. Similarly, wind turbines might stop working for maintenance or malfunctions. Related records should be removed from data.

### 4.5. Data augmentation

Data augmentation is a powerful method when researchers want to focus on a special weather condition and there is not enough data for such condition. For example in [51], Wang et al. divided the weather types into 10 categories and then used Generative Adversarial Network (GAN) to generate more data for some of these categories and improve the data balance. In [52] and [53], the inputs include wind direction only, but the authors calculated the wind direction sine and cosine and added them to inputs. In [54], Ospina et al. calculated five statistical measures of the PV power data: mean, standard deviation, variance, skewness, and kurtosis, then used these features to train the model. In [31], Zhang et al. extracted 12 characteristic features and 11statistical features from the original 6 NWP variables.

### 4.6. Correlation analysis

In [25], Mutual Information (MI), a measure of linear and non-liner correlation, is used to identify the spatial and temporal correlations in wind speed data. Using MI is also reported in [28,55–58].

To measure the similarity between two datasets, a cross-correlation method can be used. In [23], Hong et al. wanted to identify which wind farms have similar information to a target wind farm, therefore, they used a cross-correlation method called Spearman's rank-order correlation. In [18], Pearson Correlation Coefficients are calculated to show the linear relationships between input variables and the wind power output, then the results are visualized in a heat map. In [59], the correlation coefficients are calculated to show the correlation between multiple input variables, between multiple wind sites, and between variables and sites in the dataset. In [60], Principal Component Analysis (PCA) is used for feature selection. In [61], correlations between wind speed and seven meteorological variables are calculated. In [62] the Pearson product moment correlation coefficient is calculated to show the correlation between six weather variables and PV power.

### 4.7. Data clustering

Clustering might be involved in renewable energy data preprocessing for dividing the dataset into different seasons or weather conditions. According to [33], the fuzzy c-mean clustering algorithm is recommended for this task. In [63], historical PV power data was clustered into different groups to identify the daily pattern label. In [62], the k-means algorithm is used to cluster solar irradiance data into five clusters, which represent five types of sky conditions. Data clustering algorithms might be part of hybrid models as discussed in 5.6.2. under Hybrid Models with Features Extraction or Selection Method.

### 4.8. Graph or grid construction

When researchers want to add the spatial dimension to the wind or solar data and utilize the wind or solar farms' location information in their forecasting models, they construct a grid or graph for spatio-temporal representation. In [64], a specialized software was used to model spatio-temporal data as images. Graph data modelling technique is reported in [16,24,27,28], and [65] while grid technique is reported in [16,26,50], and [66]. In [59], the data was reconstructed into 3-dimensional matrix to represent the site, time, and feature dimensions while in [67], the data was reconstructed into 2-dimensional matrix to represent site and time dimensions only. In [66], PV power data collected from 56 PV plants are represented in 2-dimensional grid.

## 5. Deterministic forecasting methods

Deterministic forecasting models developed for solar and wind energies are divided into six categories according to the used architectures. These are discussed in the following subsections.

**Table 3**
Summary of the forecasting methods in which CNN based models are used.

| Ref # | [21] | [68] | [16] | [23] |
|---|---|---|---|---|
| Prediction objective | PV power | PV power | Spatiotemporal wind speed and power | Spatiotemporal wind speed |
| Forecast horizon | Ultra-short | Short (1D) | Short (10–60 m) | Short (1D) |
| Preprocessing | Changing image resolution, removing wrong data | Data normalization, reshaping the input data | Grid construction | Constructing 2D images, cross-correlation analysis |
| Optimization | ReLU, batch normalization, Adam, early stopping | SGD, SELU, ReLU | None | Taguchi, Adam, dropout |
| Dataset (inputs; resolution; period; samples; location) | Video of the sky, PV power; 1 m; 7 M; 39,000; USA | Module temperature, solar radiance, PV power; 1 h; 5D; –; Taiwan | Wind power; 10 m; 3Y; –; USA | Wind speed; 1 h; 1Y; 8760 × 7; Taiwan, China, S.Korea, Philippines |
| Evaluation and comparison | Weather conditions | Benchmarking models | Benchmarking models | Benchmarking models |

Note *Y*= year, *M*= month, *W*=week, *D*= day, *h*=hour, *m*=minute, *s*= second. In the dataset, semicolon (;) is used to separate labels and comma (,) is used to separate items under the same label. If information about a specific label is not available, double dashes are used (–) as a place holder.

### 5.1. Convolutional neural network-based models

CNN based forecasting models depend on convolutional layers to extract features from data, while the regression task is done in the last fully connected layer [4]. This section discusses studies that propose CNN-based forecasting models. Table 3 summarizes important information about these works including prediction objectives, forecast horizons, preprocessing methods, optimization methods, datasets, and evaluation and comparison methods. All the tables in this paper hereon follow the same structure and content convention.

CNN is the first option when inputs include images as in [21] where sky images were used to capture the useful information about the cloud coverage, which helps prediction models to achieve higher accuracy. Sun et al. conducted many experiments to study the sensitivity of several CNN structures, and input image resolution on PV output prediction. The developed models provide high accuracy in sunny condition and less accuracy in partly cloudy and overcast conditions. As mentioned earlier, CNN is usually used when data can be represented in two or three dimensions as with images and videos, however, it is possible to use it with one-dimensional inputs as in [68]. Huang and Kuo proposed a model based on one-dimensional CNN for 24-hour ahead PV power forecasting. They reshaped the data into (90,3) to represent the three variables: temperature, solar radiation, and PV system output power in 90 records. They evaluated their model against other machine learning models including Long Short-Term Memory (LSTM) and found that their model is more accurate on average and its training time is faster than LSTM.

Some researchers predict the wind power output in a specific area by developing forecasting models that use spatial and temporal features of wind turbines in that area. For example, Yu et al. [16] studied the effect of such features on wind speed and wind power prediction accuracy. The data collected from wind turbines were mapped to a grid space, which they called it scene. The scene time series is a multi-channel image, which represents the spatio-temporal feature of wind in a certain area and time. Therefore, they developed a prediction model based on CNN to extract features from these images. The results show that the proposed model achieves better accuracy than the existing methods and the wind speed is easier to predict than wind power because the latter depends on the wind turbine's specific characteristics. Hong et al. [23] also used 2-dimensional data that represents multiple temporal wind speeds at multiple sites as inputs for their day-ahead wind speed forecasting model. Their CNN model's structure and hyperparameters were optimized by the Taguchi method, which reflected in the model's high accuracy.

### 5.2. Recurrent neural network-based models

Time series data are sequential data recorded at a specific time interval, which usually ranges from few seconds to an hour in forecasting data. Processing sequential data usually is done using RNN based model and its advanced variations LSTM and GRU. For further details about RNN, LSTM, and GRU, refer to [69]. This section discusses studies that propose RNN, LSTM, and GRU for wind or solar energy forecasting, and these works are summarized in Table 4 (in a similar manner to Table 3).

In stacked LSTM, each LSTM layer output is the input to the next LSTM layer. This deep architecture enables a more complex representation of sequential data over time. Stacked LSTM-based models are proposed in reference [24] and [61] for wind speed prediction. In [24], the spatio-temporal information is modeled as a graph whose nodes are data-generating entities and its edges represent the interaction between these nodes. This approach allows information from neighboring stations to improve the forecasts of a target station. The proposed model provides forecasts of wind speed of all nodes in the graph at the same time. In [61], meteorological inputs are used to train the stacked LSTM-based model, which reflected on more accurate forecasting than a single LSTM. In [62], LSTM model with two hidden layers is used for PV power forecasting. Forecast and historical weather variables in addition to historical PV power data were used to train and test the model, which achieves higher accuracy than RNN, Extreme Learning Machine (ELM), and the generalized regression neural network.

Attention mechanism allows deep learning models to focus on the most important features in a sequence of inputs. Novel LSTM based models with the attention mechanism were developed in [70] and [71]. Niu et al. [70] proposed a sequence-to-sequence model for multi-step-ahead wind power forecasting. Their model is based on a multi-input and multi-output strategy, which unlike a recursive strategy, provides multiple outputs at different time steps using a vector of values as inputs in one simulation. The model architecture consists of two groups of GRU blocks, which work as encoder and decoder. The encoder transforms the sequence of inputs to a context vector and the decoder uses it to predict the output sequence. Their model performance on three-time steps was compared to another four models and the results prove its superiority in terms of accuracy. Moreover, the most important features for predicting wind power as selected by the attention mechanism are wind power and wind speed. On the other hand, Zhou et al. [71] proposed a model for short-term photovoltaic power forecasting. Their model consists of two LSTM networks for temperature and PV power data inputs. The outputs

**Table 4**
Summary of the forecasting methods in which RNN based models are used.

| Ref# | Prediction objective | Forecast horizon | Preprocessing | Deep learning | Optimization | Dataset | Evaluation & comparison |
|---|---|---|---|---|---|---|---|
| [24] | Spatiotemporal wind speed | Short (1 h) | Data normalization, graph construction | Stacked LSTM | RMSProp; cross validation, ReLU | Wind speed; 1 h; 1Y; –; USA | Benchmarking models |
| [61] | Wind speed | Short (5 m) | Data normalization, correlation analysis | Stacked LSTM | ReLU, dropout | Wind speed at 10 &2-meter, wind direction, temperature, humidity, pressure, dew point, solar radiation; 5 m; 1 M; 8353; USA | Benchmarking models |
| [62] | Multistep PV power | Short (6, 12, 24 h) | Data normalization, correlation analysis, data clustering | LSTM | None | Temperature, wind speed, humidity, sky type, solar irradiance, PV power; 1 h; 3Y; –; USA | Benchmarking models, weather types |
| [70] | Multistep wind power | Short | Data normalization, changing resolution | GRU with attention mechanism | Grid search | wind power, speed and direction, temperature, air pressure and density; 5 m, 1Y; 8760×3; USA | Benchmarking models |
| [71] | PV power | Short (7.5–60 m) | Data normalization | LSTM with attention mechanism | RMSProp | PV power, PV module temperature; 7.5 m; 4Y; –; China | Benchmarking models |
| [72] | Solar irradiation | Short | None | LSTM, GRU | None | GHI;1-h; 10Y; –; France | Benchmarking models |
| [33] | Solar irradiance | Short (1 h) | Data normalization | LSTM | None | GHI, GTI; –; 4D; –; Canada | Benchmarking models |
| [73] | Load and PV power | Short (1 h) | Data normalization | DRNN+ LSTM | Regularization, dropout, Adam Hyperopt library | Weather data, PV power;1 h; 2 M; –; Australia | Benchmarking models |
| [74] | PV power | Short (1 h) | None | LSTM | Adam | PV output; 1 h; 1Y; –; Egypt | Benchmarking models |
| [48] | Solar radiation | Short | Changing data resolution | LSTM | Levenberg-Marquardt | Air-dry-bulb temperature; 2 m; 8D; –; USA | Benchmarking models, data resolution |
| [32] | Solar irradiance | Short (1D) | Data normalization | LSTM | Adam | GHI, temperature, dew point, humidity, visibility, wind speed, weather type; 1 h; 18 M, 12 M; –; Cape Verde | Benchmarking models |
| [40] | PV power | Short (1, 2, 3 h) | Removing outliers | Bidirectional LSTM | Cross-validation, Adam | PV power; 15 m; 10 M; 3640; China | Benchmarking models |
| [44] | Solar irradiance | Short | Removing outliers | LSTM, Ensemble of MLP+ SVR+ RBF+ LSTM | Portfolio theory | Solar irradiance, temperature; 1 h, 10 m; 10Y, 3Y; 87,672+ 61,404; Spain, Brazil | Benchmarking models |
| [75] | Wind power | Short | Data normalization | LSTM with wavelet activation kernels | RMSProp | Wind speed and direction, zonal and meridional components; 1 h; 3Y; –; Europe | Benchmarking models |
| [78] | Solar power | Short (10 m) | Data normalization | LSTM in Autoencoder structure | None | Irradiation, module temperature, power; 10 m; 1Y; –; Poland | Benchmarking models |

of both networks are first processed with the attention mechanism, then combined using a fully connected layer to produce the forecasting output. To test the proposed model, the authors compared its performance with the other four models using four seasons datasets and with different time horizons. The experiments show that both the proposed model and traditional LSTM perform better than the other compared models in a time horizon of more than 15 min.

Comparing LSTM and GRU performance on solar irradiation forecasting is done in [72]. The performance of these deep models with one and two layers was compared to simple RNN and Naive models. The results show that LSTM and GRU models both provide good results, especially with two layers architectures.

LSTM-based forecasting models performance was compared with several machine learning models, such as Support Vector Machine Regression (SVR), Bagged Regression Trees (BRT), Feed Forward Neural Network (FFNN), the Linear Least Squares Regression (LLSR) method, Back Propagation Neural Network (BPNN), and Multilayer Perceptron network (MLP) for both solar radiation and PV power prediction in [33,73,74,48], and [32]. The results of all these studies confirm the superiority of the LSTM model. Also, a comparison between three statisti-

**Table 5**
Summary of the forecasting methods in which SAE based models are used for wind energy.

| Ref# | [79] | [80] | [81] | [55] |
|---|---|---|---|---|
| Prediction objective | Wind speed by Transfer Learning | Wind speed | Multistep ahead wind power | Multistep ahead wind power |
| Forecast horizon | Short (10 m, 30 m, 1 h, 2 h) | Ultra/short (10m-3 h) | Short (1–9 step) | Short (10m-2 h) |
| Preprocessing | Data normalization | None | Data normalization | MI, changing data resolution |
| Optimization | Sigmoid | Rough extension, regularization, random search, sigmoid | Particle Swarm Optimization | L2 regularization, orthogonal array testing, Adam |
| Dataset | Wind speed; 10 m; 1.5Y, 1Y; –; China. | Wind speed; 10 m; 2Y; 52,560; USA | Wind power; 15 m; 3 M; 6057; Ireland | Wind power, speed and direction, temperature, air pressure and density; 5 m; 6Y; 315,648; USA |
| Evaluation & comparison | Benchmarking models | Benchmarking models | Benchmarking models | Benchmarking models |

cal methods: Autoregressive Moving Average (ARMA), ARIMA, Seasonal Autoregressive Integrated Moving Average (SARIMA), and two LSTM-based models is provided in [40]. The findings indicate that bidirectional LSTM using historical PV power data only can achieve reliable prediction for one hour ahead and not longer. In [38], Zhanga et al. used an LSTM model for short-term wind turbine power forecasting and compared its performance with radial basis function, wavelet, DBN, BPNN, and Elman neural network (ENN). The results of the experiments show that LSTM has the best performance in one to five time-step forecasting. Furthermore, Lima et al. [44] compared the performance of the LSTM-based model with the ensemble of machine learning algorithms for solar irradiance forecasting. They found that although the LSTM model's accuracy is competitive, the ensemble model achieves higher accuracy.

Researchers have tried to improve LSTM-based forecasting models by changing their structure. For example in [75], Shahid et al. used four different wavelet activation kernels in LSTM layers instead of using the sigmoid activation function: Gaussian, Morelet, Ricker, and Shannon. Their proposed model for short-term wind power prediction provides more accurate predictions than Deep Neural Network (DNN), Support Vector Machine (SVM), ARIMA, and genetic programming-based ensemble methods. Besides, Li et al. [76] added a lower upper bound estimation method to the LSTM model for interval prediction of wind power. Their model consists of LSTM, fully connected layers, and the rank-ordered terminal, which produces the upper bound and the lower bound by ranking the outputs of the fully connected layers. The results prove the superiority of the proposed model over ELM, SVM, and ANN. Furthermore, Hu et al. [77] developed a density forecast model for wind speed and wind power prediction. It consists of two parts: a forecast network, which takes the data inputs and generates the parameters for the next part, and a distribution approximation network, which approximates the real cumulative density functions of the forecasting target. The structure of the model contains several LSTM layers and fully connected layers. The proposed model shows superior performance in terms of reliability, sharpness, and skill score over the other eleven models included in the comparison. Suresh et al. [78] developed a model based on LSTM and Autoencoder structure for PV power forecasting. The encoder part of the model consists of input layer and LSTM layer whereas the decoder has LSTM layer and FFNN.

### 5.3. Stacked autoencoder based models

Autoencoders networks are known for their ability to reduce dimensionality in data and to generate a close representation of the original data. In the case of denoising autoencoders, noise is added deliberately to data, which makes them more robust and more capable of handling noisy inputs and learning more features from the data than a standard autoencoder. More details about this model are found in [4]. Studies that suggest forecasting models based on both types are discussed in this section and their details are summarized in Table 5.

Stacked Denoising Autoencoder (SDAE) network in [79], is proposed for wind speed prediction with the transfer learning approach (TL). In the proposed model, the input and the two hidden layers are shared across many wind farms while the output layers are not shared. Each farm has its output layer since its data distribution differs from that of other farms. Using this architecture, the trained network transfers information from data-rich farms farm to a newly built farm. The experimental results show that the model's prediction accuracy is higher compared with SVR, two- hidden-layer DNN, and ELM. SDAE network is also proposed in [80] for wind speed forecasting, but with rough extension to handle uncertainties in data. The linear regression layer is used at the top of the network, which takes the extracted features for supervised learning. Experimental results for 10 min to 3 h ahead forecasting show that the proposed models outperform the Persistence model, FFNN model, time-delay neural network, and nonlinear autoregressive neural network model.

Stacked Autoencoder (SAE)-based model in [81], is proposed for multi-step ahead wind power forecasting, which consists of three autoencoders with sparse constraints and random noise to extract features from data, which are then fed into a BPNN to perform regression. The experiments' results show that the model achieves higher accuracy than BPNN and SVM even though its performance degrades when time-steps increase (from 1- step ahead to 9-step ahead) due to the error transmission phenomenon. Furthermore, another SAE-based model is developed in [55] for wind power sequence-to-sequence prediction. It is composed of SAEs, one feature fusion layer, and one prediction terminal layer. The features extracted by SAEs are unified in the fusion layer and then used by the prediction terminal layer for prediction generation. The results demonstrate that the proposed model outperforms Adaptive Neural Fuzzy Interference System (ANFIS), the autoregressive model with external inputs, MLP, SVR, Decision Tree (DT), and traditional SAE over 10-min to 2-hour ahead predictions.

### 5.4. Deep belief network-based models

DBN is a generative model in which neurons from different layers are connected while there is no connection among them in the same layer. DBN is constructed by stacking several layers of Restricted Boltzmann Machines (RBM). More details about this model are found in [4]. In this section, two forecasting models based on DBN are discussed and their details are provided in Table 6.

A DBN based model for photovoltaic power forecasting is proposed in [82]. According to the results, the optimal number of RBM in the DBN model was found to be two with 500 hidden nodes, and this structure achieves higher forecasting accuracy than ELM. On the other hand, DBN

**Table 6**
Summary of the forecasting methods in which DBN based models are used.

| Ref# | [82] | [83] |
|---|---|---|
| Objective | PV power prediction | Wind speed prediction |
| Horizon | Short | Short (1 h) |
| Optimization | Cross-validation | Genetic algorithm |
| Dataset (inputs; resolution; period; samples; location) | Panel temperature, ambience temperature, accumulated energy, irradiance; 15 m; 14D; 1339; Singapore | Wind speed, station and sea pressure, temperature, dew point, relative humidity, wind direction, max gust and its direction, precipitation amount and hours, sunshine hours; 1 h; 1Y+14D; –; Taiwan |
| Evaluation | Benchmarking models | Benchmarking models |

**Table 7**
Summary of the forecasting methods in which other deep models are used.

| Ref# | [84] | [18] | [45] | [85] | [86] | [87] |
|---|---|---|---|---|---|---|
| Prediction objective | Multistep wind speed | Impact of wind turbine's features on wind power | Wind power | Load and wind power | Solar irradiance | Solar & wind power |
| Forecast horizon | Ultra/short (5 s,10–60 m) | Ultra-short | Ultra-short | Long (monthly) | Short (1, 2, 3 h) | Short (1–48 h) |
| Pre-processing | None | Removing outliers, correlation analysis | Removing outliers, data normalization | Data normalization | Removing night hours data | Data normalization |
| Deep learning | SELM with Generalized Correntropy | 5-layer FFNN | 5-layer MLP | Deep echo state network | Deep echo state network | RBFNN |
| Optimization | Regularization, grid search | Xavier, ReLU | Xavier, ReLU | None | Ridge regression | Adam, cross-validation |
| Dataset (inputs; resolution; period; samples; location) | Wind speed; 1 s, 10 m; 3h+19 m, 5 M; 15,000; China | 4 wind speeds at different heights, 4 pitch angles, nacelle orientation, yaw error, temperature; 1 s; 1Y; –; UK | Wind speed and direction, nacelle orientation, yaw error, pitch angle, temperature, wind power; 1 s; 1Y; –; UK | Wind power, 1 M; 6Y+10 M; –; China | Solar irradiance; 1 h; –; –; USA | Short wave flux, cloud coverage, temperature, humidity; wind speed and direction, pressure, PV and wind power; 1 h; 2.5Y, 10 M, 1.5Y, 3Y; –; Greece, Europe |
| Evaluation & comparison | Benchmarking models | Data fusion | Benchmarking models | Benchmarking models | Benchmarking models | Benchmarking models |

with the Genetic Algorithm (GA) is used for wind speed forecasting in [83]. The genetic algorithm was employed to select the DBN parameters. Results demonstrate that the proposed model is more accurate than the SARIMA model, and the Least-Squares Support Vector Machine Regression (LSSVM) model with GA.

### 5.5. Other deep models

Other studies that suggest different deep learning architectures than what is covered in previous sections are examined in this section and their details are provided in Table 7.

Luo et al. [84] combined the advantage of deep neural networks and ELM algorithm, which is known for its fast training by proposing a stacked ELM model for wind speed prediction. This model uses generalized correntropy as the cost function instead of the mean square error to deal with outliers in the datasets. Experiments show that the proposed model achieves the best accuracy in both multi-step second-level and multi-step minute-level forecasting compared with other models, such as ANN, DNN, ELM, and Stacked ELM. For wind power prediction, Lin and Liu [18] used Five-layer FFNN trained by new input variables to study the effect of these features on the model accuracy. The inputs include four wind speed measurements at different heights, nacelle orientation, yaw error in addition to the wind turbine blade pitch angle features, and air temperature. The model was trained several times while replacing one feature at a time by its mean value. This helps to discover strong non-linear relationships between each feature and the output. Moreover, Lin et al. [45] used the MLP model for wind power prediction and Isolation Forest for outlier detection and elimination.

Echo state network, which consists of the input layer, dynamic reservoir, and output layer, is known for its fast-learning ability. Forecasting models based on the echo state network are suggested in [85] and [86] for both solar and wind energy. Hu et al. [85] stacked three layers of reservoirs in their model to combine the advantages of both deep learning and echo state networks. The proposed model achieves higher accuracy compared with the Persistence model, ARIMA, BPNN, and echo state network. Liet al. [41] also employed three layers of reservoirs in their model. They used time series analysis to preprocess inputs before training the model. Results demonstrate that the proposed model outperforms echo state network, BPNN, and ENN in 1, 2, and 3 h ahead prediction. Sideratos and Hatziargyriou [87] proposed a model based on Radial Basis Function Neural Network (RBFNN) with multiple layers for wind and solar power forecasting. The inputs were first grouped into clusters based on their importance, then RBFNN regressors generate the prediction. The final output is the sum of all the regressors' weighted outputs.

### 5.6. Hybrid and ensemble models

Combining several deep learning models or a deep learning model with other methods, such as data decomposition or feature selection, has improved the forecasting accuracy as will be discussed in this section and detailed in Tables 8–13. Tables 8–10 show studies of hybrid forecasting models for wind energy while Tables 11 and 12 show studies for solar energy. Table 13 presents studies that target both solar and wind energy together. All these tables present the reference number, prediction objective, forecast horizon, preprocessing methods, deep learning model used, optimization methods, information about the dataset (in-

**Table 8**
Summary of the forecasting methods in which hybrid models are used for wind energy (Part 1).

| Ref # | Prediction objective | Forecast horizon | Preprocessing | Deep learning | Optimization | Dataset | Evaluation & comparison |
|---|---|---|---|---|---|---|---|
| [88] | Wind speed | Ultra/short (10 m, 1 h) | None | Ensemble model of LSTM +SVR | Extremal optimization | Wind speed; 10 m, 1 h; 6D, 30D; 738+720; China. | Benchmarking models |
| [89] | Wind speed | Ultra/short (10 m, 1 h) | None | Hybrid model of LSTM+ ELM | Differential evolution, hysteretic function | Wind speed; 10 m, 1 h; –;720 × 2; China | Benchmarking models, statistical testing |
| [56] | Wind power by TL | Short | MI | Ensemble of 9 deep AE + DBN | Adding sparsity to AE | Wind power, Zonal and meridional component, wind speed and direction; 12 h; 3Y; –; Europe | Benchmarking models |
| [57] | Wind speed | Short | MI, WT | Hybrid model of WT+ FS+LSTM | Crow Search | Wind speed, 10 m, 1 h; 80D,1Y; 1200 × 2; Spain, Iran | Benchmarking models |
| [69] | Wind speed | Short (1 h) | WT | Hybrid model of WT+ RNN+SVM, WT+ LSTM+ SVM, WT+ GRU+ SVM | None | Wind speed; 1 h; 2 M, 5 M; –; China | Benchmarking models |
| [94] | Multistep wind speed | Ultra-short | WT | Hybrid model WT+ DBN +LGBM, Hybrid model WT+DBN+ RF | None | Wind speed; –; –;1000 × 4; China | Benchmarking models |
| [95] | Multistep wind speed | Short (10–12 h) | DWT | Hybrid model of DWT +LSTM | Dropout | Average and turbulence wind speed; 10 m; 1Y; 52,560; China | Benchmarking models |
| [34] | Multistep wind speed | Short | WSTD | Hybrid model of WSTD+ GRU | Cross-validated-grid search, dropout | Wind speed; 1 h; 42D; 1000 × 4; USA | Benchmarking models, decomposition comparison |
| [7] | Multistep wind speed | Short | WPD | Hybrid model of WPD+CNNLSTM+CNN | Adam, ReLU | Wind speed; 10 m; 1Y; 700 × 4; China | Benchmarking models |
| [98] | Wind speed | Short (10 m, 1 h) | Discrete WPD | Hybrid model of DWPD+ BiLSTM | Adam | Wind speed; 10 m, 1 h; 2Y; –; USA, Canada | Benchmarking models, computation time |
| [101] | Wind speed | Short (12 h, 24 h) | EMD | Hybrid model of EMD+SAE+ ELM | None | Wind speed and direction; 15 m; 2Y; –; UK | Benchmarking models |
| [102] | Wind speed | Short | EEMD | Hybrid model of EEMD+ MLP+LSTM+ ARIMA+ MOPSO | Adam, RelU, MOPSO | Wind speed; 10 m; –; 3600 × 3; China | Benchmarking models, decomposition comparison |
| [103] | Multistep wind speed | Short | VMD, KLD+ EM+ SA | Hybrid model of VMD+ KLD+EM+ LSTM+ error correction strategy | None | Wind speed;1 h; –; 43,824 *3; China | Benchmarking models |
| [104] | Wind power | Short (1–3 h) | VMD | Hybrid model of VMD+ Recurrent autoencoder | Adam, ReLU | Wind speed, wind power; 1 h; 1Y; –; Belgium, Spain, USA | Benchmarking models |
| [105] | Wind power | Short (1–3 h) | VMD, data normalization | Hybrid model of VMD+ residual CNN | Leaky ReLU, SGD | Wind power, speed, and direction; 1 h; 1Y; 8760; Turkey | Benchmarking models, weather types |
| [106] | Wind power | Short | EWT | Hybrid model of EWT + SE+ KELM+ GRU | Slime mould algorithm | Wind power; 10 m; 22D; 1704; Spain | Benchmarking models |
| [107] | Multistep wind speed | Short | EWT | Hybrid model of EWT + LSTM+RELM+ IEWT | None | Wind speed; 10 m; –; 1000 × 4; China | Benchmarking models |
| [108] | Wind Speed | Short | EWT | Hybrid model of EWT+ New Cell Update LSTM | Adam, Mini batch | Wind speed; 15 m; 60D; –; China | Benchmarking models, decomposition comparison |
| [109] | Multistep wind speed | Short | EWT | Hybrid model of EWT, LSTM, and Elman NN | None | Wind speed; 1 h; –; 700 × 4; China | Benchmarking models, decomposition comparison |

puts, resolution, period, number of samples, location), and evaluation or comparison methods (in a similar manner to Table 3).

Some researchers proposed ensemble models that combined a deep learning model with powerful machine learning algorithms. For example, Chen et al. [88]. developed an ensemble model based on LSTMs and SVR for wind speed forecasting. First, a cluster of six LSTMs models with a different number of hidden layers and neurons is used to perform the prediction. The reason for using six instead of a single LSTM is to improve the generalization capability and robustness of the model. Then, the results of LSTMs prediction are used to train the SVR model. Also, Hu et al. [89] proposed a hybrid model for wind speed forecasting. Their model consists of three different ELM models improved using hysteretic

**Table 9**
Summary of the forecasting methods in which hybrid models are used for wind energy (Part 2).

| Ref # | Prediction objective | Forecast horizon | Preprocessing | Deep learning | Optimization | Dataset | Evaluation & comparison |
|---|---|---|---|---|---|---|---|
| [47] | Multistep wind speed | Short | EWT | Hybrid model of SAE+ BiL-STM + MOMVO + ORELM | Regularization, cross-validation, Adam, MOMVO | Wind speed; 3 s; 1D; 30,000; China | Benchmarking models |
| [111] | Wind speed | Short (15 m, 1 h) | CEEMDAN | Hybrid model of CEEMDAN+ PE+ GRU+ RBFNN | Improved Bat algorithm | Wind speed; 15 m, 1 h; 20D, 80D; 1920 × 2; China | Benchmarking models, decomposition comparison |
| [112] | Multistep wind speed | Short | SSA | Hybrid model of SSA +CNN+ GRU+SVR | Adam, grid search, ReLU | Wind speed; 15 m; 2 M, 2 M, 1 M; 2688 × 3; China. | Benchmarking models |
| [113] | Wind power | Short | SSA | Hybrid model of SSA+ LSSVM+DBN+ LSH | 10-fold cross-validation, LSH | Wind power; 10 m; –; 3400; China | Benchmarking models |
| [114] | Multistep wind speed | Ultra-short (10–30 m) | SSA | Hybrid model of SSA+ MADANET | Dropout | Wind speed; 10 m; –; 700 × 4; China | Benchmarking models, decomposition comparison |
| [115] | Multistep wind speed | Short (1–4 h) | ISSA | Hybrid model of ISSA+ LSTM+DBN | Grasshopper optimization | Wind speed; 1 h; 1 M; 720; China | Benchmarking models, decomposition comparison |
| [117] | Multistep wind speed | Short (10 m, 1 h) | TVF-EMD, FE, SSA, PSR | Hybrid model of TVF-EMD+ FE+ +SSA+ PSR KELM+ ConvLSTM | Adam, MHHOGWO | Wind speed; 10 m, 1 h; 73D; 1008, 744; Spain, Australia | Benchmarking models, decomposition comparison |
| [118] | Multistep wind speed | Short | SSA +EMD | Hybrid model of SSA+ EMD+ CNN+ SVM | ReLU, Adam, dropout, grid search regularization, | Wind speed; 10 m; –; 700 × 4; China. | Benchmarking models |
| [119] | Multistep wind speed | Short | VMD+ SSA | Hybrid model of VMD+SSA+ LSTM+ELM | None | Wind speed;1 h; 8 M; 700 × 4; China | Benchmarking models, decomposition comparison |
| [53] | Multistep wind power | Short | EMD +VMD | Hybrid model of EMD+VMD+CNN+LSTM | ReLU, Adam | Wind power & speed, wind direction sine and cosine; 1 h; –; 720; Spain | Benchmarking models |
| [121] | Wind speed | Short | CEEMDAN, VMD | Hybrid model CEEMDAN+ VMD + LSTM | None | Wind speed; 15 m, 1 h; 30D, 90D; 2880 × 2 + 2160×2; USA | Benchmarking models |
| [122] | Wind speed | Short | SSA+ CEEMDAN | Hybrid model of Conv-LSTM+BPNN | ReLU, Modified whale optimization | Wind speed; 15 m; –; 2000 × 3; China | Benchmarking models |
| [25] | Multistep spatiotemporal wind speed | Short (10m-3 h) | MI | Hybrid model of CNN and LSTM | RMSProp, early stopping | Wind speed; 5 m; 1Y; 52,704; USA | Benchmarking models |
| [64] | Wind power and load | Ultra-short (10s-1 m) | Data modeling by software | Hybrid model of CNN+ LSTM | SGD, Adam, early stopping | Wind speed; 10 s; 12 h; 22,000; software | Benchmarking models |
| [123] | Wind power | Ultra-short (10s-1 m) | Changing data resolution | Hybrid model of stacked Convolutional LSTM | None | Wind fields; 0.4 s; 12 h; 158,400; software | Benchmarking models |
| [30] | Wind power | Ultra-short | Combining adjacent wind turbines data | Ensemble model of CNN+ LightGBM | ReLU, dropout, early stopping | Fan status, temperature, wind power & speed, motor speed, wind direction, daily power, pitch angle; 5 m; 1Y; –; China | Benchmarking models |
| [59] | Wind speed | Short | Correlation analysis | Hybrid model of CNN+LSTM | ReLU, Adam | Wind speed & direction, temperature, dew point, gust, altimeter setting, humidity; 5 m; 6 M; 25,918 +25,920; USA | Benchmarking models |

neuron activation function, and three different LSTM models optimized using differential evolution algorithm. In addition, the predictions of all the six models are combined by another LSTM model. Other researchers used ensemble models for a certain reason, such as transfer learning. For example, Qureshi et al. [56] proposed an ensemble model for wind power prediction, which contains nine deep auto-encoder as the base regressors and DBN as the meta regressor. Because there are five different wind farm datasets, transfer learning was implemented, to reduce the computational time of training and improve the prediction performance. With this approach, each deep autoencoder was just pre-trained on the first wind farm dataset, then fine-tuned to work with the rest four wind farm datasets. For each wind farm dataset, a separate DBN was trained. Moreover, Kumari and Toshniwal [19] developed an ensemble model that consist of eXtreme Gradient Boosting Forest (XGBF) and DNN for hourly GHI forecasting. They tested their model using data collected from 3 locations that have different climate and the results show the model superior performance compared with SVM and RF.

#### *5.6.1. Hybrid with data decomposition methods*

Time series data usually consist of four components: level, trend, seasonality, and noise. Achieving accurate forecasting involves using one of the data decomposition methods to separate all the four components or at least separate the level and noise. In hybrid models, researchers use one of the decomposition methods as a data processing step to decompose time-series data into several subseries before they train a forecasting model for each one. The final forecasting result is the combination

**Table 10**
Summary of the forecasting methods in which hybrid models are used for wind energy (Part 3).

| Ref # | Prediction objective | Forecast horizon | Preprocessing | Deep learning | Optimization | Dataset | Evaluation & comparison |
|---|---|---|---|---|---|---|---|
| [49] | Wind power | Short (24 h) | Data normalization | Hybrid model of CNN + RBFNN with DGF | Dropout, Adam | Wind power; 1 h; 1Y; 8760; Taiwan | Benchmarking models, weather types |
| [27] | Multistep spatiotemporal wind speed | Short (10 m-3 h) | Graph Construction | Hybrid model of LSTM + Conv layers+ Rough layers | ReLU | Wind speed and wind direction; 5 m; 6Y; 105,120 × 6; USA | Benchmarking models |
| [65] | Wind power | Short (90 m) | 4 clustering methods | Hybrid model of Spectral Clustering + LSTM | Improving LSTM | Wind speed and power; 10 m; 3Y; –; USA | Benchmarking models |
| [124] | Wind power | Short (1D) | K-means clustering | Hybrid model of k-means clustering +DBN | GD, greedy algorithm | Wind speed and direction, wind power, air pressure, humidity, temperature; 10 m, 1 h; 1Y; –; Spain | Benchmarking models, weather types |
| [31] | Wind power | Short (1D) | Dimension reduction, K-means clustering | Hybrid model of k-means clustering + LSTM with attention mechanism | None | Wind speed and direction, temperature, humidity, pressure; 10 m, 1 h; 1Y; 50,688; Spain | Benchmarking models |
| [60] | Wind power | Short (24 h) | PCA | Hybrid model of PCA+ LSTM | None | Wind power, air density, pressure, temperature, wind speed and direction; 5 m; 14 M; –; UK | Benchmarking models |
| [125] | Wind speed | Short | Clustering | Ensemble model of Adversarial AE+ORELM | MODWPT, MOFEPSO | Wind speed; 1 m; –; 2000; China. | Benchmarking models |
| [126] | Wind speed and power | Medium (1 W) | K-means clustering | Hybrid model of K-means clustering+ WNN+ RKF | Cross-validation | Wind power and speed; 1 h; –; –; Canada | Benchmarking models |
| [127] | Wind power | Short (24–72 h) | None | Ensemble of SDAE | Min-batch SGD, MLR | Wind speed, sin and cos of wind direction, wind power; 10 m; 1Y; –; China | Benchmarking models |

**Table 11**
Summary of the forecasting methods in which hybrid models are used for solar energy (Part 1).

| Ref# | Prediction objective | Forecast horizon | Preprocessing | Deep learning | Optimization | Dataset | Evaluation & comparison |
|---|---|---|---|---|---|---|---|
| [19] | GHI | Short (1 h) | Correlation analysis | Hybrid model of XGBF+ DNN | Grid search, dropout, Adam, ReLU | GHI, temperature, wind speed and direction, humidity, pressure, clear sky index; 1 h; 10Y; –; India | Benchmarking models, weather types |
| [93] | Multistep PV power | (1–60D) | DWT | Hybrid model of DWT+LSTM | Dropout | Cloudy index, visibility, temperature, dew point, humidity, wind speed, pressure, altimeter, power; 1 h; 2Y; –; USA | Benchmarking models |
| [54] | PV power | Short (30 m) | SWT | Hybrid model of SWT+ LSTM+ DNN | Cross-validation | PV power, predicted temperature, irradiance, PV statistics; 30 m; 2Y; 35,089; USA | Benchmarking models |
| [97] | PV power | Short (1 h) | WPD | Hybrid model of WPD + LSTM | None | PV power, GHI, DHI, wind speed, temperature, humidity; 5 m; 2Y; –; Australia | Benchmarking models, weather types |
| [17] | PV power | Short (1D) | VMD | Hybrid model of VMD+ DBN+ ARIMA | Adaptive ant colony algorithm | PV power, solar irradiance, air temperature; 5 m; –; –; Australia | Benchmarking models, weather types |
| [110] | Solar irradiance | Short (1 h) | CEEMDAN | Hybrid model of CEEMDAN+ CNN+ LSTM | Grid search | Solar irradiance; 1 h; 6Y; 210,336; USA, Algeria | Benchmarking models, weather types |

of the results given by all the forecasting models. There are many methods used in the literature. All of the studies in this section agree on the effect of using a decomposition method on increasing hybrid prediction models' accuracy.

Wavelet-based decomposition methods can separate signals into several components with different frequencies and thus higher resolutions [46]. One wavelet-based technique is Wavelet Transform (WT), which is used in [90–93,57,69], and [94]. Wang et al. [90] proposed a hybrid wind speed forecasting framework based on WT, DBN, and Quantile Regression (QR). They used Mallat, a discrete WT algorithm that consists of four filters, to decompose wind speed data into one approximation and several details. The DBN model used with the approximation component has four layers while with details components, a 10-layer DBN model is used. Again, Wang et al. [91], and [92] used the Mallat algorithm in their hybrid model for wind power forecasting, and PV power forecasting, but with a deep CNN network this time. Mishra

**Table 12**
Summary of the forecasting methods in which hybrid models are used for solar energy (Part 2).

| Ref# | Prediction objective | Forecast horizon | Preprocessing | Deep learning | Optimization | Dataset | Evaluation & comparison |
|---|---|---|---|---|---|---|---|
| [36] | Solar power | Short (1D) | Data Normalization | Hybrid model of CNN+ LSTM | ReLU, early stopping, batch normalization | Power, irradiation, temperature, wind speed, humidity; 10 m, 1 h; 4Y; 18,620; S.Korea | Benchmarking models, data resolution |
| [128] | Multistep PV power | Short (15m-3 h) | Correlation analysis | Hybrid model of CNN+ LSTM | None | PV power; 15 m; 1Y; –; Belgium | Benchmarking models, weather types |
| [39] | PV power | Long (1Y) | Correlation analysis | Hybrid model of CNN+ LSTM | ReLU, batch normalization, regularization | PV power, GHI, DNI, DHI, wet bulb, dew point, temperature; 1 h; 25Y; 5271; Australia | Benchmarking models |
| [35] | Multistep solar radiation | 1D, 1 W, 2 W, 1–8 M | Data normalization | Hybrid model of CNN+LSTM | Grid search, Adam, ReLU, early stopping, dropout, regularization | GSI; 1 m; 13Y; 60,743; Australia | Benchmarking models |
| [67] | Spatiotemporal GHI | Short (1 h) | Reconstructing spatial and temporal features | Hybrid model of CNN-LSTM | None | GHI, dew point, solar zenith angle, wind speed and direction, precipitable water, relative humidity, temperature; 1 h; 7Y; –; USA | Benchmarking models, weather types |
| [52] | PV power | Short (1D) | Removing outliers, Data augmentation | Hybrid of CNN+LSTM | Dropout | Current phase average, active power, wind speed, temperature, relative humidity, GHR, DHR, wind direction sine & cosine; 5 m; 4Y; –; Australia | Benchmarking models, input timesteps |
| [129] | Thermal power | Short (30 m) | None | Hybrid of CNN+LSTM+MLP | ReLU, dropout | GHI, DHI, GTI, DNI, solar altitude angle, solar azimuth angle, pressure, humidity, temperature, wind speed, power; 30 m; 1Y;–; China | Benchmarking models, weather types |
| [66] | Spatiotemporal PV power | Short (15–60 m) | Changing data resolution | Hybrid of Convo LSTM | Grid search, Nadam, Adagrad | PV power; 15 m; 1Y; 16,060; USA | Benchmarking models, data contaminations |
| [29] | Solar power | Short (24h- 48 h) | Data normalization | Hybrid model of AE+ LSTM | ReLU, early stopping | NWP data, solar power; 3 h; 990 D; –; Germany | Benchmarking models |
| [131] | PV power | Short | Removing outliers | Hybrid model of LSTM + CNN | Dropout | Current phase average, active power, wind speed, temperature, relative humidity, GHI, DHI, wind direction; 5 m; 6 M; 53,280; Australia. | Benchmarking models |
| [22] | PV power | Ultra-short | K-means clustering | Hybrid model of CNN +LSTM+ ANN | None | Sky images, irradiance; 15 m; 1Y; 25,000; USA. | Benchmarking models |
| [42] | Global solar radiation (GSR) | Short (1 h, 1–3D) | Removing outlier | Hybrid model of embedding clustering + functional DBN | Backward elimination algorithm | max, min, mean of dry-bulb, temperature, mean wind speed, mean relative humidity, sunshine duration, GSR; 1D; 22Y; –; China | Benchmarking models |
| [132] | PV power | short (24 h) | grey theory-preprocessing using ago | Hybrid model AGO + DBN+ FFNN | None | PV power; 1 m; 1Y; 5760; Taiwan | Benchmarking models, weather types |
| [63] | PV power | Short (1D) | Data Clustering for daily pattern label | Ensemble model of LSTM + time correlation modification + daily pattern classification | None | PV power; 15 m; 6Y; –; USA | Benchmarking models |
| [133] | PV power | Short (1D) | Changing data resolution | Hybrid model of LSTM+ DGM | Genetic algorithm, Adam, dropout, batch normalization | PV power, solar irradiance, air temperature, relative humidity, wind speed, cloud, air pressure, weather type; 15 m; 1Y; –; China. | Benchmarking models, weather types |
| [51] | Weather classification for PV power | Short (1D) | Data augmentation | Hybrid model of GAN + CNN for classification | Adam, ReLU | Solar irradiance; 1 m; 1Y; –; USA | Benchmarking models, weather types |

**Table 13**
Summary of the forecasting methods in which hybrid models are used for solar & wind energies together.

| Ref# | Prediction objective | Forecast horizon | Preprocessing | Deep learning | Optimization | Dataset | Evaluation & comparison |
|---|---|---|---|---|---|---|---|
| [99] | Wind & solar power | Short (6 h) | EMD | Hybrid model of EMD +LSTM, EMD +GRU, and EMD + DNN | Adam, RelU | Past 24 h renewable electricity supply; 1 h; 1000 h; –; S.Korea | Benchmarking models |
| [100] | Wind & solar power | Mid (7D) | EMD | Hybrid model of EMD +LSTM, EMD +GRU, and EMD + DNN | ReLU, dropout | Past 21D demand and supply; 1D; 5Y; –; S.Korea | Benchmarking models |
| [130] | Wind speed and solar irradiance | Short (1D) | None | Hybrid model of CNN+GRU | ReLU, Adam, dropout | Wind speed, solar irradiance; 30 m; 1Y; –; UK | Benchmarking models |

et al. [93] used Discrete WT to decompose solar power data and an LSTM model for PV power prediction. Also, WT is used in [57,69], and [94] to decompose wind speed data into subseries. In [57], an ensemble of four LSTM models are trained to generate the forecasting results of each subseries while in [69] Yu et al. compared RNN, LSTM, and GRU networks' performance, which are used for feature extraction of low frequency sub-series only and the prediction is performed by SVM. For high-frequency sub-series, data were fed directly to SVM skipping the feature extraction step. In [94], along with WT, DBN is used to extract the features. However, the new data samples provided by DBN are used for prediction differently. The first model predicts wind speed using Light Gradient Boosting Machine while the second model utilizes Random Forest (RF). Both methods are kinds of ensemble learning algorithms. In [95], Fei Li et al. used Discrete WT to decompose the wind speed data into the mean wind speed and the turbulent intensity. Turbulence intensity can be used as a measure of the uncertainty of wind speed. Then, LSTM is trained and tested with two ensemble strategies to fuse the final prediction result. Results show that the turbulence intensity feature has improved the multi-step prediction, especially with higher data resolution. A modification of Discrete WT called Stationary Wavelet Transform (SWT) is utilized in [54] to avoid shift-variant results associated with using the sub-sampling operation in the DWT method. After applying SWT to decompose the PV power signals data into three detailed and one approximate component, an ensemble of four LSTMs networks is trained to recognize the non-linear features, and the outputs are used to reconstruct the forecasted PV power signal using the inverse SWT (ISWT) process. Then, statistical features extracted from the data, such as the mean and standard deviation as well as the results obtained from ISWT are fed as inputs to the DNN model, which consists of three hidden layers, to produce the final prediction. Moreover, a new Wavelet Soft Threshold Denoising (WSTD) method is proposed in [34] for wind speed data processing, which is a signal processing technique that can filter noise signals without distorting the information. In addition to this method, the GRU model is employed to extract the features and make the prediction.

Wavelet Packet Decomposition (WPD) is a version of WT in which not only the approximation component is filtered, but also the detail component. Thus, it provides a better decomposition than WT [46]. Liu et al. used WPD in [7] and [96] as part of hybrid wind speed prediction models. In [7] a CNN model with 1-demintional convolution operator is used with the high-frequency subseries while another CNN model with an LSTM layer is used with the low-frequency subseries. In [96] the resulted subseries are reshaped into two-dimension tensors and fed into a CNN model to perform the forecasting. Furthermore, Li et al. [97] used the WPD method along with an ensemble of four LSTM networks in their model for PV power forecasting. They also utilized a linear weighting method to aggregate and produce the final prediction. Similarly, Dolatabadi et al. used discrete WPD in [98] for wind speed data decomposition. The resulted subseries were reconstructed using the theory of dynamic reconstruction and then fed into Bidirectional LSTM network to perform wind speed prediction.

Empirical Mode Decomposition (EMD) method can decompose data into many intrinsic mode functions. Unlike WT, this method is adaptive and thus there is no need to choose a function, structure, or decomposition level to optimize its performance [46]. EMD is part of hybrid forecasting models in [99–101]. In both references [99] and [100], EMD is used along with three models: LSTM, GRU, and DNN for renewable energy generation forecasting. Experiments show that the hybrid model with DNN has the lowest errors in [99] while the hybrid model with GRU achieves the best prediction performance in [100]. Also in [101], a hybrid wind speed forecasting model is proposed, which utilizes EMD to process data, and SAE to extract the features in addition to ELM for prediction. The results show that the proposed model performs the best in 12 h ahead forecasting. He et al. in [102] used Ensemble Empirical Mode Decomposition (EEMD) to analyze wind speed data, then SSA to reconstruct data. Forecasting is generated by an ensemble of three models, MLP, LSTM, and ARIMA. To obtain the weight of each model, Multi-objective particle swarm optimization (MOPSO) algorithm is used.

Variational Mode Decomposition (VMD) is a more recent decomposition method than EMD, but its performance depends on defining some parameters, such as the number of modes [46]. Zhang et al. [17] developed a hybrid model for PV power forecasting, which includes VMD to separate the data into regular and irregular components. The forecasting of the regular component is handled by the ARIMA model while the forecasting of the irregular component is done by the DBN model. Also, Wang & Li [103] used VMD to decompose wind speed data in their hybrid model while an LSTM network is trained to generate the prediction. Wang et al. [104] and Yildiz et al. [105] also utilized VMD to decompose wind power data before they are fed into their proposed hybrid model for wind power forecasting.

Self-adaptive decomposition methods are a category of decomposition methods that do not need parameters set in advance and they can adapt their parameters [46]. Empirical Wavelet Transform (EWT) belongs to this category, which is used in [106–109], and [47] as part of hybrid models for wind energy forecasting. In [106], Yan and Wu first used the EWT method to decompose wind power data into subseries. Then, either a simplified version of GRU model or a kernel ELM model is applied to generate wind power prediction based on the value of Sample entropy (SE) for each subseries. The final output is the sum of both models' outputs. In [107], subseries generated by EWT are grouped into two training sets, then LSTM network is trained with the first set to generate the prediction and tested with the second set to calculate the forecasting error series. Afterward, a regularized ELM network is trained on the resulted error. Both networks are combined and tested with the testing set. Finally, the combined forecasting results are reconstructed using Inverse EWT to form the final forecasting result. In [108], Pei et al. proposed a new LSTM model called New Cell Update LSTM, in which they combined the input gate and the forget gate into the New update gate. An ensemble of the new LSTM models is trained using subseries generated from applying the EWT method to produce the prediction. LSTM model is also used in [109] to predict the low-frequency sub-layer, while the ENN is employed to predict the high-frequency sub-layers. In [47], EWT is used to decompose wind speed data and the residual error series. In [110], Gao et al. utilized Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) method to decompose the solar irradiance data. This method is an advanced version of Ensemble Empirical Mode Decomposition (EEMD) method, which separates the data into high and low frequency signals. They also compared five different structures of CNN and LSTM hybrid model for hourly solar irradiance forecasting. Liang et al. [111] also used CEEMDAN method in their hybrid model for wind speed forecasting. Permutation entropy (PE) values were calculated for each wind speed data component to measure the complexity, then combine similar components accordingly. These components were fed into a cluster of five GRUs followed by RBFNN to combine the prediction of all components and generate the final output. The parameters of RBFNN were optimized by the improved Bat algorithm.

Singular Spectrum Analysis (SSA) is a nonparametric method for time series analysis [46]. In reference [112–114], SSA was used to analyze wind speed data in combination with other deep models to perform wind speed prediction. In [112] after using SSA to decompose the time series into the main trend component and some detail components, A CNN-GRU network is designed to predict the main trend component, and extract the deep features while SVR is adopted to predict the detail components. In [113], predicting the trend is done by LSSVM algorithm whereas predicting the detail component is done by the DBN model. In [114] prediction of all the wind data components is performed by a novel model called Modified Adaptive Structural Learning of Neural Network (MADANET), which does not have a fixed structure, rather it adapts its structure automatically according to the inputs. Moreover, Yan et al. [115] proposed a hybrid model for multistep wind speed forecasting, which utilizes the improved SSA method. Their model consists of LSTM model for low-frequency subseries forecasting, and DBN model for high-frequency subseries forecasting. DBN model's hyperparameters were optimized using Grasshopper optimization algorithm. The final forecasting result is obtained combining both models' outputs.

In some studies, researchers combined more than one decomposition method to reach more detailed components of the original timeseries and thus more accurate prediction. For example, Zang et al. [116] decomposed historical PV power series into seasonal, trend, and random components using VMD. Then, EWT is used to further decompose the random component into several stationary ones. The resulted data after decomposition were added to weather data in the last two weeks to reconstruct 2-dimensional feature maps. The weather type (sunny, partially cloudy, overcast/ rainy) was encoded in the forms of 2-dimensional maps as well. The 2-dimensional maps were used as inputs to train two CNN models: Residual Neural Network (ResNet), and Dense Convolutional Network (DenseNet) for day-ahead PV power forecasting. Moreover, Fu et al. [117] in their framework first used Time-Varying Filter-based Empirical Mode Decomposition (TVF-EMD) method to decompose the wind speed data into several intrinsic mode functions with different frequencies. Then, the fuzzy entropy (FE) values are calculated for each intrinsic mode function. The resulted values go through SSA to further divide them into high-frequency and low-frequency components. Then, for high-frequency series, the kernel-based extreme learning machine (KELM) is used as the predictor. For low-frequency series, CNN and LSTM- based model predicts the wind speed. In addition, Mi et al. [118] in their hybrid wind speed prediction model used SSA first to reduce the noise in the wind speed data. Then, they used EMD to decompose the data into sub-layers to extract features. Prediction is done by a model that combines CNN and SVM. Liu et al. [119] used VMD first to decompose wind speed data. Then, trend information was extracted using SSA. For low-frequency subseries, an LSTM network was trained to perform the forecasting while ELM was used for high-frequency subseries forecasting. On the other hand, Xiang et al. in [120] used SSA first, then applied VMD on wind speed subseries. Prediction in their hybrid model is done by an ensemble of bidirectional GRU networks. Yin et al..[53]. employed EMD to decompose the wind speed series into several intrinsic mode functions and then applied VMD to decompose the first intrinsic mode function only. In their hybrid model. CNN model is used for feature extraction while LSTM is used for prediction. Ma et al. [121] first applied CEEMDAN to decompose wind speed time series, and then applied VMD to decompose error time series. The resulted data are used to train an LSTM model and get the prediction. Experiments' results prove that this double decomposition strategy improved the accuracy of the LSTM model. On the other hand, Chen et al. [122] used SSA first to separate data into training and testing sets then used CEEMDAN to decompose the training data only. A combination of CNN and LSTM network is used to extract the feature and perform prediction.

### 5.6.2. *Hybrid with features extraction or selection methods*

Many researchers utilize the advantages of deep learning models for non-linear feature extraction from data, such as CNN or LSTM, and other researchers prefer to use different feature selection methods in their hybrid forecasting approach as will be discussed in this section.

CNN model is widely used for deep features extraction in renewable energy forecasting models, especially when data contain spatial information. Lee et al. [36] used a doubled CNN with different size filters to extract short time features from data. The output feature of CNNs is then fed into an LSTM network to produce the solar power prediction. On the other hand, in [128], PV power prediction result is the combination of both CNN output and LSTM output. Moreover, Ray et al. [39]. proposed a hybrid model of CNN and LSTM for yearly PV power prediction .Ghimire et al. [35], Zang et al. [67], Wu et al. [37], Zhu et al. [25], and Wang et al. [52] also combined CNN for feature extraction and LSTM for the following prediction tasks: GHI prediction, wind power prediction, concurrent wind speed prediction of multiple sites, and PV power prediction. Similarly, Qin et al. [64] used CNN to depict the regional construction (spatial trend) of wind current time series and LSTM to extracts the dynamic features of the wind time series. Also, Wang et al. [129] developed a hybrid model for thermal power prediction. It consists of CNN for spatial feature extraction, LSTM for temporal feature extraction, and MLP network for forecasting. Besides, Woo et al. [123]. used four convolutional LSTM networks to capture the spatio-temporal features of wind fields and predict the wind turbine response. Their model is built based on the theory of turbulent dynamics. Convolutional LSTM model is also proposed in [66] for PV power forecasting of multiple locations simultaneously. The model is implemented based on encoder-decoder structure, and tested with several data contaminations cases including incompletion, Gaussian noise, mixed noise, and outliers. Additionally, Afrasiabi et al. [130] in their framework for wind turbine and photovoltaic power forecasting, converted the time series data into two-dimensional vectors, then used CNN block and GRU block to extract the features, followed by a dense block to produce the forecasting outputs. A hybrid model of CNN and GRU is also proposed in [26] by Liu et al. to extract the spatial-temporal features from the wind turbine grid matrix that was constructed using wind speed data and wind turbines' location information. Moreover, Ju et al. [30] in their model for wind power prediction, used CNN for feature extraction and the LightGBM algorithm to generate the forecasting results. LightGBM is a boosting algorithm that can significantly expediate forecasting and reduces memory utilization. Before training the model, they combined data from adjacent wind turbines to improve forecasting accuracy. This step proved to be useful in the results. Besides, Chen et al. [59] first structured the data into a 3-dimentional matrix to represent the three aspects of wind speed data correlations: between multiple sites, between multiple meteorological factors, and between sites and factors. Then, they used CNN to extract the spatial relationship between the meteorological factors of a target site and its adjacent site and LSTM to extract the temporal features. Also, Hong et al. [49] proposed a hybrid model of CNN and RBFNN with Double Gaussian Function (DGF) for day-ahead wind power forecasting. The CNN part of the model extracts the features from wind power data while the RBFNN part generates the prediction. The RBFNN with Double Gaussian Function (DGF) can filter the spikes in the data and improve the accuracy.

Autoencoder is a well-known feature extracting model in hybrid models. For example, Gensler & Raabe [29] used Autoencoder for feature extraction and LSTM network for forecasting the energy output of solar power plants. LSTM model could also be used for temporal feature extraction only and not to generate the prediction. For example, in Wang et al. hybrid model [131], LSTM is not used to generate prediction as usual, but rather CNN is used to extract the spatial features and generate the output. They noted that extracting the temporal features first then the spatial features improve the accuracy more than doing the opposite steps.

When the data are represented in a graph structure, deep models are favored for feature extraction as in [27] and [28]. Khodayar and Wang [27] modeled the wind farms' data in a graph structure. Then, they employed an LSTM model to extract the temporal features from the wind speed sequential data. They also applied the Rough Set theory in additional convolutional layers to handle noise and capture interval deep features which are then used to generate the prediction. On the other hand, Khodayar et al. [28] modeled the spatio-temporal data collected from solar sites as a graph where each node represents a solar site and each edge reflects the correlation between the corresponding sites. Then, they used a convolutional graph ANN model to extract the spatio-temporal features as well as an encoder and decoder ANN model to produce samples drawn from the probability densities learned at each node.

In some studies, a feature selection method is used after using a decomposition technique. For example, Memarzadeh and Keynia [57] used the MI method on the decomposed subseries before the reconstruction step, which ranks candidate inputs according to their information value for wind speed prediction. Moreover, Wang & Li [103] used Kullback–Leibler Divergence (KLD) and Energy Measure (EM) of signal strength to select features and eliminate noise in wind speed sub-signals. Afterward, a novel approximate entropy called Sample Entropy (SE) is utilized to integrate the selected features according to the SE values.

A features selection method could also be a data preprocessing step before training the forecaster model. For example, Yu et al. [65] tried four clustering methods to cluster similar wind turbines' data to the target turbine and filter out sequential correlation features: K-means, Agglomerative hierarchical clustering, Balanced Iterative Reducing Clustering using Hierarchies, and Spectral Clustering. They found that the latter method along with their improved structure of the LSTM model give higher accuracy and higher convergence speed than the traditional LSTM. A clustering method is also part of Wang et al. [124] model for wind power forecasting. Specifically, they used the k-means clustering algorithm to analyze the NWP data and identify the largest historical samples that have the greatest impact on forecasting accuracy. Then, they used a DBN model of five layers for prediction. Zhang et al. [31] also used k-means clustering to divide the samples into 3 clusters and then trained an encoder-decoder network based on LSTM with attention mechanism for wind power prediction. Zhen et al. [22]. used sky images as inputs for their PV power forecasting model. Therefore, they utilized a Convolutional Autoencoder model to convert the sky images into feature vectors. Then, they applied K-means clustering on the obtained vectors. They found that the images are optimally grouped into three classes. The classified data are fed into a hybrid model that consists of CNN, LSTM, and ANN to be mapped to irradiance values. On the other hand, Xiaoyun et al. [60] used the PCA method to select the most relevant features for wind power prediction, which are found to be wind speed and wind direction. Moreover, Khodayar et al. [58] used the MI method to calculate the correlation among historical wind speed data and select the most important time lags for prediction. Then, they used interval DBN to extract the nonlinear features. The top-down relevant feature search algorithm is used in Zhu et al. [20] model for probabilistic wind speed forecasting, which combines Gaussian Process Regression (GPR) and LSTM. In addition, Zang et al. [42] decided to use a novel clustering method called embedding clustering to group similar metrological stations' data into clusters, then train a model for each cluster. This method consists of an autoencoder and k-means clustering. After clustering, data are fed to a hybrid model of DBN and a functional network. The model structure consists of four layers: knowledge functional layer, polynomial functional layer, neuron elimination layer, and DBN. The neurons of the knowledge functional layer include ten metrological inputs and eleven well-known functions from weather physical (empirical) models. The polynomial functional layer combines neurons to constitute a 3rd-order polynomial family of functions. In the neuron elimination layer, the backward elimination algorithm is used to eliminate redundant functions. The result of this elimination process is 182 input nodes, which are used to train the DBN. Furthermore, Duan and Liu [125] developed a novel wind speed ensemble forecasting model. They first utilized deep Adversarial Auto-Encoder to convert data into two-dimensional Gaussian distribution feature space, then clustered this feature space using bat algorithm. For each cluster, an Outlier Robust Extreme Learning Machine model (ORELM) is trained to perform the prediction, which is further improved by Maximal Overlap Discrete Wavelet Packet Transform (MODWPT). To obtain the best ensemble weights, The Multi-Objective Feasible Enhanced Particle Swarm Optimization (MOFEPSO) is used. Also, Chang and Lu [132] developed a DBN model with a grey-theory-based data preprocessor and FFNN for day-ahead PV power output forecasting. The grey theory is applied using Accumulated Generating Operation (AGO) in the input layer of the DBN to reduce irregularity in data series while the FFNN is used in the output layer of the DBN to calculate the prediction output. In [126], Aly developed several hybrid models for wind power and speed forecasting to compare their performance. He used k-means clustering to group data into four clusters to improve the forecasting accuracy. The experiments show that Wavelet Neural Network (WNN) combined with Recurrent Kalman Filter (RKF) model achieved the best performance among other 11 models.

### *5.6.3. Hybrid with data correction methods*

In some hybrid forecasting models, the error obtained from forecasting in addition to the forecasting results are fed into the model to produce the final forecasting output. For example, in [121], the error series is decomposed using VMD, then an LSTM model is trained to generate the error prediction. The results are used to correct wind speed prediction. Also In [96], the adaptive multiple error corrections method is employed to filter out the remaining predictable components in the forecasting residuals. ARIMA model is used to correct the residuals if Ljung-Box Q-test results show that there is autocorrelation. Correction shrinkage rate controls the correction process, which is optimized by blocked cross-validation. Moreover, in [103] an error correction strategy called Generalized Auto-Regressive Conditionally Heteroscedastic (GARCH) is employed. If there are correlation and heteroscedasticity in the error component, the hybrid LSTM-GARCH model is applied. If only heteroscedasticity exists, GARCH model alone is applied for correction. Furthermore, Liu & Chen [47] developed a hybrid multi-step wind speed forecasting model with error correction. First, they used two different methods to transform the resolution of data: the averaging method and the stacked sparse autoencoder. Then, two Bidirectional LSTM (BiLSTM) models are trained: one using the averaged data and the other using the output of the autoencoder. The BiLSTM models generate 10-step predictions, which can simultaneously learn forward and backward information. The Multi-Objective Multi-Verse Optimization (MOMVO) method is used to find the optimal combination weights of both predictions. Afterward, the residual error is calculated, which is further analyzed by EWT and ORELM methods to eliminate the outliers from the error subseries.

Researchers also have tried to use the weather type information to correct or improve the forecasting results. For example in [63], Wang et al. proposed a PV power ensemble forecasting model that consists of two sub-models: an LSTM model and a time correlation model, which are integrated through the partial daily pattern prediction framework. The inputs for the LSTM model are PV power data in the previous three days of the forecasting day and the output is PV power in the forecasting day. The results are modified using the time correlation model. Then, a partial daily pattern prediction framework is applied to predict the daily pattern of the forecasting day, which is used to optimize the time correlation model. Moreover in [133], Gao et al. developed several PV power prediction models for ideal and non-ideal weather conditions. Metrological data were used as inputs for the ideal weather conditions prediction models, which are based on LSTM. On the other hand, only power output data of similar days and adjacent days were used for non-ideal weather prediction conditions models, which are based on LSTM and Discrete grey Model (DGM). DGM is used to correct the hourly PV power data before they are fed into LSTM as input. Additionally, Wang et al. [51] studied the effect of weather classification on PV power forecasting accuracy. They developed a model based on GAN and CNN that can classify data into ten weather types. Since there are no sufficient samples in each weather type, they utilized GAN to generate new samples for training dataset augmentation. The CNN is used to perform the classification, which is trained by the augmented training dataset. To verify the effect of weather classification on PV power forecasting accuracy, they compared three forecasting models: a model based on ten weather types, a model based on four weather types, and a model that does not use weather type in forecasting. The results show that the model based on ten weather types can accurately predict the uncertainty and fluctuation of the solar irradiance curve. Furthermore, Yan et al. [127] proposed a Numerical Weather Prediction correction method using an ensemble of SDAE, which corrects the weather inputs of multiple sites before generating the wind power forecasting for a region.

## 6. Probabilistic forecasting methods

Probabilistic forecasting aims to find the prediction intervals in which the actual values fall or assign a probability to a prediction result. Quantifying the uncertainty associated with renewable energy power forecasting is essential for assisting in planning and managing the electric systems. This has attracted researchers to develop probabilistic forecasting models. Twenty-two papers, which proposed probabilistic forecasting models for renewable energy were published in the Web of Science database between 2016 and 2020. Half of them were published in 2020 alone. Various parametric and nonparametric methods are covered in these studies. For example, Zhang et al. [38] used the results of their LSTM model for wind turbine power forecasting to conduct uncertainty analysis, in which the Mixed Gaussian model is applied. This model provides probabilistic forecasting as confidence intervals. The results show that the Mixed Gaussian model is more accurate than two uncertainty analysis methods: Mixture Density Neural Networks and Relevance Vector Machine. In [134], Zhang et al. developed a model for wind power direct probabilistic forecasting using deep Gaussian Mixture Density Network. To improve the model, Beta distribution is assumed and modified ReLU function is used in the output layer. In addition, Zhu et al. [20] developed a model for probabilistic wind speed forecasting, which combines GPR and LSTM. Also, Li et al. [76] used the Lower upper bound estimation method and LSTM model for wind power interval prediction while Wang et al. [91] used the same method with ensemble technique, in which the results of point forecasting are used to construct the prediction intervals. The lower upper bound estimation method is also utilized in [120]. In this hybrid model, wind speed data after being decomposed by VMD, go through the Fuzzy Information Granulation (FIG) process, where they are further divided into three components: minimum, middle, and maximum. Afterward, these three components are fed into an ensemble of three Phase Space Reconstruction (PSR)- Bidirectional GRU networks to generate the forecasting interval. Khodayar et al. [58] developed a model for wind speed probabilistic forecasting based on interval probability distribution learning. Their model utilizes interval DBN, to extract the temporal features from wind speed time-series in addition to a fuzzy type 2 inference system to generate the prediction values. Moreover, Hu et al. [77] proposed a density forecast model for wind speed and wind power prediction that consists of a distribution approximation network, which estimates the real cumulative density functions of the forecasting output, and a forecast network, which predicts the previous network's parameters. On the other hand, Wu et al. [37] analyzed the error of their hybrid model for wind power prediction and found that it does not follow the Beta or Gaussian distribution. Therefore, they decided to assume the empirical distribution and formulate conditions set. Then, they used the clustering method to split the conditions set and form the error empirical probability density function, which can be used to estimate the error. Furthermore, QR is used in [90] and [92] as part of the developed hybrid models. Also. Liu et al. [96] used Bivariate Dirichlet Process Mixture Model (BDPMM) as part of their hybrid model. The input to this model is the corrected deterministic forecasting residuals and the output is heteroscedasticity probabilistic forecasting results with non-parametric distributions. Additionally, Zang et al. [116] applied transfer learning in probabilistic forecasting when they initialized the CNN models' parameters with the same point forecasting parameters and retrain them according to quantile loss functions. Then, the results are merged using kernel density estimation. Furthermore, Khodayar et al. [28] developed a novel model for spatio-temporal solar irradiance probabilistic forecasting. The spatio-temporal data, collected from 75 solar sites, are modeled as a graph where each node represents a solar site, and each edge reflects the correlation between the corresponding sites. The model consists of a convolutional graph feature extraction ANN, which extracts spatio-temporal features from GHI observations as well as encoder and decoder ANN, which produces samples drawn from the probability densities learned at each node. Additionally, Liu et al. [50]. first constructed a grid to represent PV data and their locations information. Then, they developed GRU with a convolutional operator to extract the spatio-temporal features from this grid. Afterward, the Variational Bayesian inference is utilized to perform the probabilistic solar irradiation forecasting of an entire region. Again, they used Variational Bayesian as part of the proposed model in [26] to provide probabilistic forecasting using Kernel density estimation. More details about the aforementioned studies in this section are provided in Table 14.

## 7. Evaluation and comparison methods

### 7.1. Evaluation metrics

The common three metrics for deterministic forecasting are Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The coefficient of determination ($R^2$) and Standard Deviation of Error (SDE) are frequently used as well. On the other hand, probabilistic forecasting results are usually reported using the Average Coverage Error (ACE), Continuous Ranking Probability Score (CRPS), and interval sharpness (IS).

In [17], SDE is used to measure the stability of the forecasting model while MAE, MAPE, and RMSE are used to measure accuracy. In [116], Zang et al. used eight loss functions to update the convolutional layers' parameters: mean squared error, MAE, mean squared logarithmic error, mean absolute scaled error, negative index of agreement, Theil U-statistic 1, Theil U-statistic 2, and mean Huber error. All these metrics were used to report the results of deterministic forecasting. On the other hand, for probabilistic forecasting, they used three metrics: ACE, IS, and pinball loss. In [77] for the proposed probabilistic forecasting model, the average proportion deviation is used to measure the model reliability while the average width of prediction Intervals is used to measure its sharpness. Skill score is also used to show both measures in one value. Sometimes researchers use the promoting percentage of error criterion to compare the performances of two models and show the degree of improvements, such as $P_{MAE}$, $P_{MAPE}$, and $P_{RMSE.}$ These percentages are used in [74,108,112,114,117–119], and [121]. Also, Normalized MAE and normalized RMSE are used in [60, 41], and [67]. Furthermore, the direction accuracy is used to evaluate the sensitivity of the proposed model in [122] while the variance of prediction error is used to evalu-

**Table 14**
Summary of the forecasting methods in which hybrid models are used for probabilistic forecasting.

| Ref# | Prediction objective | Forecast horizon | Preprocessing | Deep learning | Optimization | Dataset | Evaluation & comparison |
|---|---|---|---|---|---|---|---|
| [38] | Wind power | Short | Data normalization | Hybrid model of LSTM+ Gaussian Mixture | Adam | Wind speed and power, NWP data; 15 m; 3 M; 3072; China | Benchmarking models |
| [134] | Wind power | Short | Data normalization | Deep Gaussian Mixture Density Network | Adam, SGD, batch normalization, gradient clipping | Wind speed, direction, vector, power,; 1 h; 4Y; 18,757; Europe | Benchmarking models, computation time |
| [20] | Multistep wind speed | Short (15m-60 m) | Data normalization | Hybrid model of top-down relevant feature search + Gaussian Process Regression +LSTM | Semi-stochastic alternating GD | Average & peak wind speed, temperature, dew point, relative & specific humidity, station & sea-level pressure, accumulated precipitation; 1 m; 1 M; 2880 +2976; USA | Benchmarking models |
| [76] | Wind power | Short (10 m) | | LSTM with fully connected layers and rank ordered terminal | RMSProp, RelU | Wind power; 10 m; 1Y; –; USA | Benchmarking models |
| [91] | Wind power | Short (15m- 8 h) | WT | Hybrid model of WT+CNN | SGD | Wind power; 5 m, 15 m; 2Y; –; China. | Benchmarking models |
| [120] | Multistep wind speed | Short | SSA+ VMD, FIG, PSR | Hybrid model of SSA+ VMD+ FIG+ PSR+ bidirectional GRU+ CSO | Chicken swarm optimization, grid search | Wind speed; 10 m; 2 M; 8784 × 2; China | Benchmarking models |
| [58] | Multistep wind speed | Short (1h- 24 h) | MI | Hybrid model of interval DBN with rough pattern recognition+ fuzzy type 2 inference | SGD, L2 regularization | Wind speed; 10 m; 3Y; 52,560 × 3; USA | Benchmarking models |
| [77] | Wind power and speed | Very short (5 m) | None | Cumulative density functions+ LSTM | Grid search, early stopping, batch normalization | Wind power, speed and direction; 5 m; 3Y, 12Y; –; Australia, USA | Benchmarking models |
| [37] | Wind power | Short (4 h) | Data normalization | Hybrid model of CNN +LSTM+ conditional wind power prediction error analysis | ReLU | Wind power and speed at 70 m, weather prediction; 15 m; 1Y; –; China | Benchmarking models |
| [90] | Wind speed | Short (15m-8 h) | WT algorithm | Hybrid model of WT+ DBN+QR | SGD | Wind speed; 1 h, 5 m; 2Y; –; China, Australia | Benchmarking models, weather types |
| [92] | PV power | Short (15m-2 h) | WT | Hybrid model of WT+ CNN+QR | SGD | PV power; 15 m; 1Y; –; Belgium. | Benchmarking models, weather types |
| [96] | Wind speed | Short | WPD | Hybrid model of WPD+ CNN+ Adaptive Multiple Error Correc- tions + BDPMM | Blocked cross validation | Wind speed; 1 m; –; 4500 × 3; China | Benchmarking models |
| [116] | PV power | Short (1D) | VMD + EWT | Hybrid model of VMD+ EWT+CNN+ QR | Sigmoid & ReLU, batch normalization, grid search, dropout | PV output, wind speed, temperature, humidity, GHI, DHI, rainfall, actual and predicted weather type; 1 h; 10Y+2 M; 89,000; Australia | Benchmarking models, data fusion |
| [28] | Spatiotemporal solar irradiance | Short (30m- 6 h) | MI | Convolutional Graph Autoencoder | SGD, ReLU | Solar sites coordinate, GHI; 30 m; 19Y; –; USA | Benchmarking models |
| [26] | Spatiotemporal wind speed | Short (3 h) | Grid construction | Hybrid model of 3D CNN+ GRU+ Variational Bayesian inference | ReLU, dropout, L2 regularization, Adam, grid search | Wind speed and direction, temperature, air pressure, relative humidity, dew point, precipitable water; 1 h; 2Y; –; USA | Benchmarking models |
| [50] | Spatiotemporal solar irradiation | Short (3 h) | Grid construction | Hybrid model of Convolutional GRU+ Variational Bayesian inference | Monte Carlo, Adam, ReLU, dropout, batch normalization, | GHI, Zenith Angle, temperature, dew point, relative humidity, precipitable water, wind speed and direction; 1 h; 2Y; 17,520; China | Benchmarking models |

ate the forecasting stability. In addition to using RMSE, $R^2$, and mean squared logarithmic error metrics in [18], the explained variance score is used for the model performance comparison. In [37], to compare probabilistic forecasting results, the mean value of interval skill scores of all prediction intervals is used in addition to ACE while in [96], prediction interval coverage probability, prediction interval normalized average width, and Coverage width-based criterion are used.

### 7.2. Benchmarking models

To validate the forecasting models superiority, researchers usually compare the proposed deep learning model performance with the Persistence model first because it is the baseline, then other machine learning model, such as SVR, RF, DT, and ELM, in addition to shallow neural networks, such as MLP and FFNN, or statistical methods, such as ARMA, ARIMA, and SARIMA. However, sometimes they include physical models in the comparison as done in [29,105], and [116]. In [42], the proposed model was compared to three empirical models, the SVR model, GPR model, ANFIS model, DBN model, and functional DBN models. The proposed model achieves the best performance on average for all locations and forecasting horizons.

### 7.3. Weather types (conditions and seasons)

In many studies, the accuracy of the forecasting model is compared under different weather conditions and seasons. In [21], the CNN based model achieves relatively higher accuracy in sunny conditions compared to partly cloudy and overcast conditions., In [17], The proposed hybrid model of VMD, DBN, and ARIMA performance was compared to the ARIMA model, LSSVM model, WNN, WT- particle swarm optimization -SVM model, GA-SVM model, and EMD-sine cosine algorithm-ELM model in terms of accuracy and stability. Results show that the proposed model outperforms them all on sunny days, cloudy days, rainy days, and under extreme weather conditions. In [33], the LSTM model outperforms SVR and FFNN in four weather conditions: few clouds, scattered clouds, overcast, and clear sky. In [51], the forecasting model performance when using ten different weather conditions achieves higher accuracy than using four weather classifications or without classification at all. This proves the importance of differentiating the weather conditions on the model accuracy. In [67] the proposed hybrid model of CNN-LSTM achieves higher accuracy than the Persistence model, SVM model, ANN model, CNN model, LSTM model, CNN-ANN model, and ANN-LSTM model in three different sky conditions: cloudy, partly cloudy, and cloudless for five locations. Also, the same models were compared in all four seasons for the same five locations. In [44], the proposed ensemble model of MLP, SVR, RBFNN, and LSTM performance was compared to the same models separately in all four seasons. Results show the proposed model's superiority in all seasons. In [54], the dataset was divided to represent three seasons only: spring, summer, and fall. The proposed hybrid model of SWT, LSTM, and DNN outperforms the Persistence model, WT-BPNN model, WT- RBFNN model, DNN model, SVR model, and LSTM model in almost every metric used in all three seasons. In [23], the proposed CNN-based model outperforms the FFNN model and LSTM model in all four seasons. In [124], the hybrid model of k-means clustering and DBN outperforms the BPNN model and the Morlet Wavelet Neural NetworkMWNN) model in all four seasons. In [49], the hybrid model of CNN, RBFNN with DGF outperforms the multi-FFNN-GA model, RBFNN-GA model, RBFNN-DGF-GA model, CNN-multi FFNN model, and CNN-RBFNN model in all the four seasons. In [97], the hybrid model of WPD and LSTM performance was compared with the LSTM model, GRU model, RNN model, and MLP model in all four seasons and using three sunny, cloudy, rainy days in each season. The results show the proposed model superiority in all the mentioned cases. In [132], the proposed hybrid model of AGO, DBN, and FFNN outperforms the DBN model, ARIMA model, BPNN model, RBFNN model, SVR model. The improvement over the DBN model is high on non-typical days while it is low on typical sunny days. Comparison between all the mentioned six models was conducted for all the 12 months in addition to the four seasons. In [71], the performance of the proposed model based on LSTM with attention mechanism was compared to single LSTM model, the Persistent model, MLP model, and ARIMA model with an exogenous variable in all 12 months for 7.5 min, 15 min, 30 min, and 60 min forecast horizons. The results show the superiority of both LSTM based models in a forecast horizon longer than 15 min. In [133], the LSTM model alone is proposed for prediction under sunny day condition while a hybrid model of LSTM and DGM is proposed for prediction under the other three types of weather: rainy, cloudy, and overcast. Both proposed models achieve the highest accuracy in all four seasons and conditions compared to the BP model, LSSVM model, and WNN model. In [90], the proposed hybrid model of WT, DBN, and QR exhibits the best deterministic and probabilistic prediction performances in all four seasons for both datasets (China, Australia) compared to the ARMA model, BPNN model, and MWNN model. In [92], the proposed hybrid model of WT, CNN, and QR achieves the highest deterministic forecasting accuracy in all four seasons and all 12 months compared to the BPNN model, SVM model, and WT+SVM model. Also, the probabilistic forecasting accuracy is higher than the WT+SVM +QR model. In [116], the proposed hybrid model with DenseNet outperforms Theta model, the exponential smoothing model, SVR model, random forest regression model, the physical model, MLP model, and CNN model in point forecasting in all-weather types (sunny, partially cloudy, and overcast/rainy) while the hybrid model with ResNet gives the best accuracy in probabilistic forecasting. It was also noted that the accuracy of DenseNet decreases when weather type is not used in the input. The daily pattern of the forecasting day is used in the proposed forecasting model in [63] to improve prediction accuracy. Simulations include comparing the forecasting results before and after this modification, which proved its effectiveness.

### 7.4. Input time-steps

The influence of different input time steps or time sequence on the forecasting accuracy is examined in a few studies only. For example, in [104], Wang et al. studied the effect of different input sequence on their PV power forecasting model accuracy. They trained their model with input sequences that range from 0.5 year up to 4 years. They found that the highest accuracy is obtained with 3 years of data. In [98], Dolatabadi et al. compared the performance of their hybrid model with different input vectors and found that the best result achieved with embedding dimension equals to 7 and embedding delay equals to18.

### 7.5. Data resolution

The effect of different data resolutions or sampling frequencies is studied in many papers. In [36], Lee et al. experimented with different data intervals ranging from 1 hour to 6 h. They found that the day ahead solar power forecasting model achieves the highest accuracy when dividing the input sequence of a day by 1-hour rather than using longer intervals. In [40], Sharadga et al. compared the effect of using the 15 min resolution data and the data after averaging to 1 hour resolution on the model accuracy. They found that their model prediction has improved with hourly data. Also in [48], Pang et al. studied the impact of different data resolution on the solar radiation prediction results (i.e., 10 min, half an hour, and an hour). They found that the accuracy of the model improves when data resolution increases. In [49], one-hour data resolution and one-day data resolution were used for the next 24-hour prediction. The results show that the performance of the proposed model with a one-day data interval has improved.

### 7.6. Data fusion

The impact of using meteorological data in addition to power data on the model accuracy is analyzed in some studies. In [36], Lee et al.

studied the effect of using weather information in addition to PV power data as inputs on their proposed model accuracy. Results show that this helped the model to achieve the highest accuracy compared to the other nine algorithms included in the experiments.

### 7.7. Decomposition comparison

In Section 5.6.1, hybrid forecasting models with data decompositions methods are discussed. Researchers sometimes compare the performance of several decomposition methods to choose the method that improves the model the most. In [34], WSTD denoising method performance was compared to the low-pass filter method, and the wavelet hard threshold denoising method in combination with 9 forecasting models. The results show that WSTD gives the best results when it is combined with all forecasting models. In [108], experimental results show that the EWT method improved the prediction accuracy of the ensemble model more than VMD and EMD. Also in [109], the EWT method performance was compared in this study with WPD and EMD. According to the results, EWT shows the best decomposition. In [114], SSA performance with MADANET is better than EMD, WPD, EWT, WPD with EMD, and SSA with EMD. In [119] the proposed hybrid model of VMD with SSA, LSTM, and ELM was compared to a hybrid model of VMD, LSTM, and ELM, a hybrid model of EMD with SSA, LSTM, and ELM, and a hybrid model of WPD, LSTM, and ELM. Their proposed model achieves the highest accuracy, which proves that using VMD with SSA for data decomposition is better than using VMD alone, WPD alone, or using EMD with SSA. In [117] the TVF-EMD decomposition performance in the proposed hybrid model was compared to EMD alone, and its modified version CEEMDAN. Results show that TVF with the EMD method helped the proposed model in achieving the highest accuracy. In [17], to validate the effectiveness of using the improved VMD method in the proposed hybrid models, simulations with other methods including WT, EMD, enhanced EMD, and VMD were conducted. Results prove the improved VMD advantage over the aforementioned methods. In [111], the hybrid model with CEEMDAN method was compared to WPD, EMD, and EEMD methods. The results show that the model with CEEMDAN achieved the least error. In [115], the performance of ISSA in decomposing wind speed data was compared to five other methods to prove its superiority: SSA, VMD, CEEMDAN, EWT, and ensemble intrinsic time-scale decomposition. The comparison included quantitative, qualitative, and computing time results.

### 7.8. Computation time

To help make a tradeoff between the model accuracy and its computation complexity, the running time of the proposed model is reported and compared with the running time of comparative models in many papers. This common practice is followed in [40,70,86], and [54] to show the feasibility of the proposed model. In [69], three hybrid models were compared in terms of computation time. In all the three models, WT is the decomposition technique and SVM is the regression model, but for feature extraction, either RNN, LSTM, or GRU is used. The order of the models in terms of calculation time, RNN in the first place, followed by GRU, then LSTM. However, all of them took a relatively long time for training ranges from 211 to 571 s. In [52], an LSTM model, CNN model, and a hybrid model of CNN and LSTM were compared in terms of training time. It is found that LSTM is the fastest while the hybrid model is the longest. Yu et al. in [65] compared the convergence speed of their enhanced LSTM model and the traditional LSTM model. In [34], it is reported that the CPU time of the hybrid model of WSTD and GRU is at least 20 times faster than that of the hybrid model of WSTD, LSTM, and GRU. In [27], the online training time of the proposed model is shown for different graph sizes. It is noted that the running time increases when the forecast horizon is extended because longer horizons mean larger graphs with more nodes. The computational time of the ANN model and RNN model were compared in [48]. Results show that the RNN model computational time is significantly longer than the ANN model in all the scenarios. In [122], the mean running time of the proposed hybrid model is around 153 s. In [18], the CPU time before and after reducing some input variables is compared. Total running time in minutes for all the comparative models is provided in [26]. According to the results, the hybrid model of CNN and GRU requires 83 min of running time. In [114], the running time of the proposed hybrid model with all four datasets is provided, which ranges from 194 s to 211 s. In [112] the running time of the proposed hybrid model is 155 s while in [42] is 382 s. In [123], the computation time was calculated for different input sizes. With 15 inputs (time-steps), the computation time was less than half a second, which shows the model feasibility for real time applications.

### 7.9. Statistical testing

In some papers, statistical tests were conducted to reject or accept the null hypothesis and show the significance of the results. For example in [49,89,102,111,112,117], and [122], Diebold and Mariano tests results are reported. In [88], three statistical tests were performed including Friedman, Aligned, and Quade. In [56], Pearson's correlation was calculated for the five wind farm datasets to show the strength of correlation between the actual and predicted wind power. Results of Pearson's correlation and the associated P-value are presented for each dataset. In [71], the independent two-sample *t*-test was used while in [75], one–way analysis of variance (ANOVA) test was conducted.

## 8. Discussion, challenges, and future research directions

A review of deep learning-based forecasting models for wind and solar energy has been presented in this paper, which covers the period from 2016 up to 2020. The papers included in this review are classified first according to deterministic and probabilistic methods, then, according to the used deep architecture into CNN based models, RNN based models including LSTM and GRU, SAE based models, DBN based models, others deep models, and hybrid models. Analysis of the works covered is summarized as follows:

- Forecasting data: in some of the reviewed studies, direct forecasting is carried out using historical power output data alone or with meteorological data while in other papers indirect forecasting is done by predicting the wind speed and solar radiation using their historical values or with meteorological data as well. Most of the researchers believe that including meteorological data improves the forecasting accuracy, however, the correlation between such features and the forecasting output differs from one location to another. Therefore, more comparative studies should be carried out to show the effect of including specific meteorological features on the models' performance to draw a conclusion in this matter.
- Forecast horizon: The forecast horizon is an important factor the leads the decision of the data resolution choice and the model structure as well. It can be ultra-short, short, medium, or long-term depending on the application or the decision-making process that the forecasting will support. There is no agreement on horizon classifications and the categories might overlap. However, a common classification found in many papers considers forecasting for few seconds to 30 min as very short, from 30 min up to 6 h as short, from 6 h up to 1 day ahead as medium-term, and anything beyond that is long-term [46]. Very short-term and short-term forecasting help energy traders and grid operators to make decisions about electricity pricing, economic electricity dispatching, and maintenance schedules. Therefore, short to medium forecasting supports any task related to the gird management and operation while the long-term horizon is beneficial for tasks related to energy systems planning [3]. Most of the studies included in this review proposed models for the next 24 h forecasting or less. Supporting decisions related to gird management

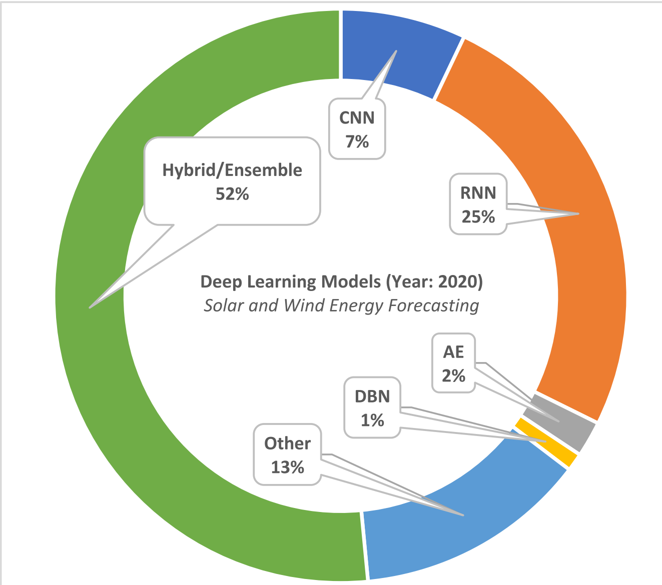


**Fig. 3.** Deep learning models for solar and wind energy forecasting used in 2020 publications in the web of science.

and operation is the main drive for this trend. Also, the drop of accuracy that accompanies longer forecast horizons due to the variable nature of weather conditions, such as the cloud cover, discourages researchers from targeting them. The shortest horizon in this review is 5 s [84] and the longest horizon is one year [39], then, 8 months [35].

- Model popularity ranking: Recently, there is more interest in hybrid models. Almost half of the papers published in the Web of Science in 2020 are for hybrid prediction models. The number of studies that proposed hybrid forecasting models included in this review paper is 83. This trend explained by the superiority of such models over single deep learning models as can be concluded from all the comparative experiments. Out of these 83 hybrid models, 58 models include LSTM or GRU as part of them. RNNs with all their newer versions, such as LSTM and GRU, are the second popular models after hybrid models, which is logical because the renewable energy data is time-series, and RNN models are known for their ability to extract temporal features. For this reason, researchers experiment with them alone or in combination with other methods. Fig. 3 displays the percentage of using each deep learning architecture in the papers published in the Web of Science in year 2020 (for wind and solar energy forecasting). The percentage is calculated based on the total of 99 published papers in that year. On the other hand, Fig. 4 shows the total number of papers published in the Web of Science from 2016 to 2020 for each deep model architecture proposed for wind and solar energy forecasting.

- Model complexity: Although the performance of hybrid models is superior to single models, their computation time is relatively high. Therefore, there is a tradeoff between the model accuracy and its computation time and complexity. The existence of big data frameworks and high-performance computers might affect this decision as well.
- Experimental cases and evaluation: To account for different weather conditions and seasons, researchers either divide the datasets and train the model several times, for example for each season, or they add the weather type as an input to train the model. Moreover, some simulations include testing the model using dataset collected from different locations to show the model generalization capability.

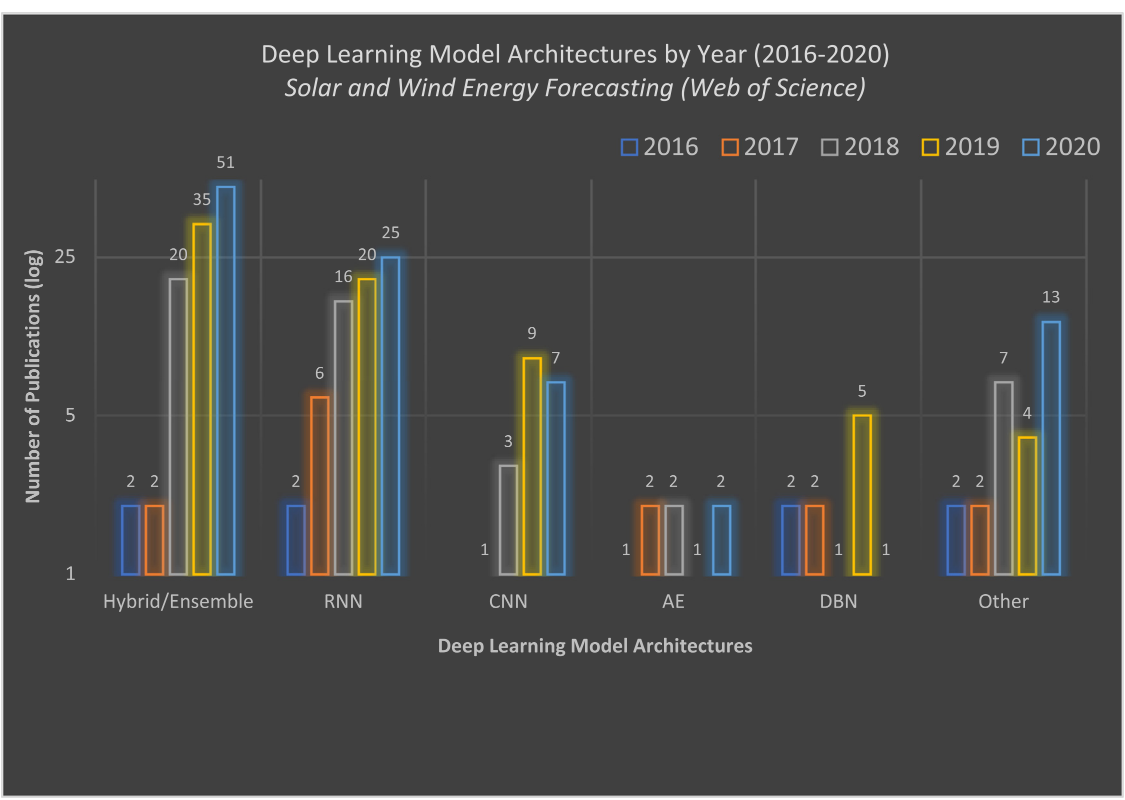


**Fig. 4.** Deep learning model architectures for solar and wind energy forecasting (by Year: 2016–2020) (Web of Science).

 

**Table 15**
Optimization methods used in the literature.

| Method | Use | Count | Method | Use | Count |
|---|---|---|---|---|---|
| Adam algorithm | Weights update | 31 | Slime mould algorithm | Parameters optimization | 1 |
| Dropout | Avoid overfitting | 21 | MHHOGWO | Parameters tuning | 1 |
| Gird search | Hyperparameter tuning | 13 | Adaptive ant colony | Parameters tuning | 1 |
| Regularization | Avoid overfitting | 12 | Crow search | Structure tuning | 1 |
| SGD | Weights update | 11 | Taguchi | Hyperparameters tuning | 1 |
| Batch normalization | Training acceleration | 10 | Genetic algorithm | Parameters tuning | 1 |
| Early stopping | Avoid overfitting | 8 | MOPSO | Weights combination | 1 |
| RMSProp | Weights update | 5 | Levenberg-Marquardt | Training | 1 |
| Xavier | Weights initialization | 2 | Extremal optimization | Parameters tuning | 1 |
| Random search | Hyperparameters tuning | 2 | Orthogonal array tuning | Hyperparameters tuning | 1 |
| Greedy learning | Pretraining | 2 | Differential evolution | Hyperparameters tuning | 1 |
| Bat algorithm | Parameters tuning | 2 | Chicken swarm optimization | Parameters combination | 1 |
| Hyperopt python library | Hyperparameters tuning | 1 | Practical swarm optimization | Hyperparameters tuning | 1 |
| MOFEPSO | Weights combination | 1 | Modified whale optimization | Parameters tuning | 1 |
| MOMVO | Weights combination | 1 | | | |

However, rarely researchers consider using datasets collected from different climates or comparing the model performance with and without using weather data. These two gaps should be considered in future studies.

- Multistep ahead forecasting: Twenty-eight studies in this review proposed multistep ahead forecasting models and nine of them were published in 2020. There are two approaches for multistep ahead forecasting: recursive approach and direct approach. In the recursive approach, the first predicted value is used as an input to the next step and so forth, which allows the error to accumulate [2]. On the other hand, the direct approach can be implemented in two ways: building separate models for each forecasting step or building a model that can take multiple inputs and generate multiple prediction outputs. The latter is called the sequence-to-sequence method, which requires less computation time [135]. Four papers published in 2020: [26,31,55], and [70] used the sequence-to-sequence method and more studies will use it in the future.
- Probabilistic forecasting: Recently, there is more interest in probabilistic forecasting since almost half of the papers that suggested probabilistic forecasting models in the Web of Science were published in 2020 alone. This trend will continue in the future driven by the advantages of probabilistic forecasting in risk assessment and decision-making support.
- General models: Building a forecasting model for each location is not feasible. Few studies proposed forecasting models for a whole region and other studies suggested transfer learning as a solution to save time where models developed previously can be quickly trained to make a prediction for new locations. Also, self-adaptive optimization methods can help in this matter. More studies in this direction will be carried out in the future.
- Implementation: Training deep learning models to find the optimal solution is still a difficult and time-consuming process, especially in the absence of guidance rules for models' structure and parameter selection. The choice is usually driven by the data nature and the researchers' prior knowledge of the field. In most cases, researchers depend on trial and error to find near-optimal solutions. However, in some studies, it has been mentioned using optimization algorithms and methods for hyperparameters tuning and avoiding problems, such as overfitting and underfitting, have been reported. Table 15 lists some of these methods and the number of studies in which they were used. It is worth saying that such methods might have been used in the included studies, but not mentioned specifically.

## 9. Conclusion

Renewable energy and deep learning are considered among the most essential and promising technologies for the future. The use of deep learning for renewable energy forecasting has shown great promise manifested for instance in the richness of the proposed methods and the increasing number of publications, with 99 publications alone in 2020 in the Web of Science collection. The current survey papers on the topic have not covered the research published during the year 2020. This paper provides a review of deep learning-based solar and wind energy forecasting research published during the last five years discussing extensively the data and datasets used in the reviewed works, the data pre-processing methods, deterministic and probabilistic methods, and evaluation and comparison methods. The core characteristics of all the reviewed works are summarized in tabular forms to enable methodological comparisons. A broad taxonomy of the research is proposed.

The most used architectures are the hybrid models followed by Recurrent Neural Network models including Long Short-Term Memory model and Gated Recurrent Unit and then, Convolutional Neural Networks in the third place. Most methods that are combined with deep learning models are, first, data decomposition techniques and, second, feature selection methods. According to the results of all the experiments in the included studies, deep learning-based forecasting models always achieve relatively higher accuracies and generalization ability compared to other machine learning models and statistical methods, especially when they are combined with other algorithms in hybrid models. However, a definite conclusion cannot be drawn about the forecaster that has the best performance unless extensive testing is done using datasets from different climates and topographies that contain data about all seasons and weather conditions.

The future trends and directions for research that we have identified include investigations into the effects of including specific meteorological features on the models' performance; generalizability or usability of proposed models across different locations, weather conditions, and seasons; multistep ahead forecasting; probabilistic forecasting; self-adaptive optimization methods, and others.

The fact that this paper has reviewed the 2020 research that has not been reviewed before, our methodological approach in structuring, reviewing, presenting, and comparing the information, and the proposed taxonomy, sets this paper apart from other survey papers in the field. We believe that our work in this paper will be vital in understanding, classifying, and comparing works on the topic ultimately accelerating innovation in this field.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix

Table 16

**Table 16**
Abbreviations used in the paper.

| NWP | Numerical weather prediction | ML | Machine learning |
|---|---|---|---|
| DL | Deep learning | CNN | Convolutional neural network |
| RNN | Recurrent neural network | DBN | Deep belief network |
| SAE | Stacked autoencoder | PV | Photovoltaic |
| ANN | Artificial neural network | AI | Artificial intelligence |
| ARIMA | Autoregressive integrated moving average | GAN | Generative adversarial network |
| MI | Mutual information | PCA | Principal component analysis |
| LSTM | Long short-term memory | GRU | Gated recurrent unit |
| SVR | Support vector machine regression | FFNN | Feed forward neural network |
| BRT | Bagged regression trees | LLSR | Linear least squares regression |
| MLP | Multilayer perceptron network | GHI | Global horizontal irradiation |
| ARMA | Autoregressive moving average | RF | Random forest |
| ENN | Elman neural network | DNN | Deep neural network |
| TVF-EMD | Time Varying Filter-based Empirical Mode Decomposition | ELM | Extreme learning machine |
| SDAE | Stacked denoising autoencoder | SAE | Stacked autoencoder |
| ANFIS | Adaptive neural fuzzy interference system | DT | Decision tree |
| RBM | Restricted Boltzmann machines | WT | Wavelet transform |
| GARCH | Generalized auto-regressive conditionally heteroscedastic | DWT | Discrete wavelet transform |
| LSSVM | Least squares support vector machine | SWT | Stationary wavelet transform |
| WSTD | Wavelet soft threshold denoising | ISWT | Inverse SWT |
| WPD | Wavelet packet decomposition | EMD | Empirical mode decomposition |
| MOPSO | Multi-objective particle swarm optimization | EEMD | Ensemble Empirical Mode Decomposition |
| VMD | Variational mode decomposition | EWT | Empirical wavelet transform |
| SSA | Singular spectrum analysis | ACE | Average Coverage Error |
| MADANET | Modified Adaptive Structural Learning of Neural Network | ResNet | Residual neural network |
| MOMVO | Multi-objective multi-verse optimization | DenseNet | Dense convolutional network |
| BPNN | Back propagation neural network | FE | Fuzzy entropy |
| CEEMDAN | Complete Ensemble Empirical Mode Decomposition with Adaptive Noise | SVM | Support vector machine |
| DGF | Double gaussian function | KLD | Kullback–Leibler divergence |
| GPR | Gaussian process regression | EM | Energy measure |
| ORELM | Outlier Robust Extreme Learning Machine model | SE | Sample entropy |
| MODWPT | Maximal overlap discrete wavelet packet transform | ReLU | rectified linear unit |
| FIG | Fuzzy information granulation | DGM | Discrete grey model |
| BDPMM | Bivariate dirichlet process mixture model | PSR | Phase space reconstruction |
| SARIMA | Seasonal autoregressive integrated moving average | RMSE | Root mean square error |
| ARIMAX | ARIMA model with an exogenous variable | MAE | Mean absolute error |
| MAPE | Mean absolute percentage error | SDE | Standard Deviation of Error |
| RBFNN | Radial basis function neural networks | LSH | Locality-sensitive hashing |
| MWNN | Morlet Wavelet Neural Network | IS | Interval sharpness |
| WNN | Wavelet Neural Network | RKF | Recurrent Kalman Filter |
| RBM | restricted Boltzmann machine | DHI | Diffuse horizontal irradiation |
| GA | Genetic Algorithm | GTI | Global tilted irradiance |
| MHHOGWO | Mutation and Hierarchy Harris hawks Optimization and grey Wolf Optimizer | SELU | Scaled exponential linear unit |
| AGO | Accumulated Generating Operation | BP | Back propagation |
| CRPS | Continuous ranking probability score | QR | Quantile regression |
| MLR | multi-response linear regression | BiLSTM | Bidirectional LSTM |
| TL | Transfer learning | SGD | Stochastic gradient descent |
| KELM | Kernel-based extreme learning machine | GD | Gradient descent |
| MOFEPSO | Multi-objective feasibility enhanced particle swarm | XGBF | eXtreme Gradient Boosting Forest |

## References

[1] Wang H, Lei Z, Zhang X, Zhou B, Peng J. A review of deep learning for renewable energy forecasting. Energy Convers Manag 2019;198:111799.
[2] Ahmed R, Sreeram V, Mishra Y, Arif MD. A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization. Renew Sustain Energy Rev 2020;124:109792.
[3] Ahmad T, Zhang H, Yan B. A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. Sustain Cities Soc 2020;55:102052.
[4] Ozcanli AK, Yaprakdal F, Baysal M. Deep learning methods and applications for electrical power systems: a comprehensive review. Int J Energy Res 2020;44:7136–57.
[5] Shamshirband S, Rabczuk T, Chau K-W. A survey of deep learning techniques: application in wind and solar energy resources. IEEE Access 2019;7:164650–66.
[6] Mosavi A, Salimi M, Faizollahzadeh Ardabili S, Rabczuk T, Shamshirband S, Varkonyi-Koczy AR. State of the art of machine learning models in energy systems, a systematic review. Energies 2019;12(7):1301.
[7] Liu H, Mi X, Li Y. Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network. Energy Convers Manag 2018;166:120–31.
[8] Sweeney C, Bessa RJ, Browell J, Pinson P. The future of forecasting for renewable energy. Wiley Interdiscip Rev Energy Environ 2020;9(2):e365.
[9] Reda FM. Deep learning an overview. Neural Netw 2019;12(21).
[10] Voyant C, et al. Machine learning methods for solar radiation forecasting: a review. Renew Energy 2017;105:569–82.
[11] Akhter MN, Mekhilef S, Mokhlis H, Shah NM. Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. IET Renew Power Gener 2019;13(7):1009–23.
[12] Wang H, et al. Taxonomy research of artificial intelligence for deterministic solar power forecasting. Energy Convers Manag 2020;214:112909.
[13] Mellit A, Massi Pavan A, Ogliari E, Leva S, Lughi V. Advanced methods for photovoltaic output power forecasting: a review. Appl Sci 2020;10(2):487.
[14] Liu H, Chen C, Lv X, Wu X, Liu M. Deterministic wind energy forecasting: a review of intelligent predictors and auxiliary methods. Energy Convers Manag 2019;195:328–45.
[15] Marugán AP, Márquez FPG, Perez JMP, Ruiz-Hernández D. A survey of artificial neural network in wind energy systems. Appl Energy 2018;228:1822–36.
[16] Yu R, et al. Scene learning: deep convolutional networks for wind power prediction by embedding turbines into grid space. Appl Energy 2019;238:249–57.
[17] Zhang J, Tan Z, Wei Y. An adaptive hybrid model for day-ahead photovoltaic output power prediction. J Clean Prod 2020;244:118858.
[18] Lin Z, Liu X. Wind power forecasting of an offshore wind turbine based on high--frequency SCADA data and deep learning neural network. Energy 2020:117693.
[19] Kumari P, Toshniwal D. Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance. J Clean Prod 2021;279:123285.
[20] Zhu S, Yuan X, Xu Z, Luo X, Zhang H. Gaussian mixture model coupled recurrent neural networks for wind speed interval forecast. Energy Convers Manag 2019;198:111772.
[21] Sun Y, Szűcs G, Brandt AR. Solar PV output prediction from video streams using convolutional neural networks. Energy Environ Sci 2018;11(7):1811–18.

[22] Zhen Z, et al. Deep learning based surface irradiance mapping model for solar PV power forecasting using sky image. IEEE Trans Ind Appl 2020;56:3385–96.
[23] Hong Y-Y, Satriani TRA. Day-ahead spatiotemporal wind speed forecasting using robust design-based deep learning neural network. Energy 2020;209:118441.
[24] A. Ghaderi, B.M. Sanandaji, and F. Ghaderi, "Deep forecast: deep learning-based spatio-temporal forecasting", no. ii, 2017, [Online]. Available: http://arxiv.org/abs/1707.08110.
[25] Zhu Q, et al. Learning temporal and spatial correlations jointly: a unified framework for wind speed prediction. IEEE Trans Sustain Energy 2020;11(1):509–23. doi:10.1109/TSTE.2019.2897136.
[26] Liu Y, et al. Probabilistic spatiotemporal wind speed forecasting based on a variational Bayesian deep learning model. Appl Energy 2020;260:114259.
[27] Khodayar M, Wang J. Spatio-temporal graph deep neural network for short-term wind speed forecasting. IEEE Trans Sustain Energy 2018;10(2):670–81.
[28] Khodayar M, Mohammadi S, Khodayar ME, Wang J, Liu G. Convolutional graph autoencoder: a generative deep neural network for probabilistic spatio-temporal solar irradiance forecasting. IEEE Trans Sustain Energy 2020;11(2):571–83. doi:10.1109/TSTE.2019.2897688.
[29] Gensler A, Henze J, Sick B, Raabe N. Deep Learning for solar power forecasting—An approach using AutoEncoder and LSTM Neural Networks. In: Proceedings of the 2016 IEEE international conference on systems, man, and cybernetics (SMC); 2016. p. 2858–65.
[30] Ju Y, Sun G, Chen Q, Zhang M, Zhu H, Rehman MU. A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting. IEEE Access 2019;7:28309–18.
[31] Zhang Y, Li Y, Zhang G. Short-term wind power forecasting approach based on Seq2Seq model using NWP data. Energy 2020;213:118371.
[32] Qing X, Niu Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy 2018;148:461–8.
[33] Alzahrani A, Shamsi P, Dagli C, Ferdowsi M. Solar irradiance forecasting using deep neural networks. Procedia Comput. Sci. 2017;114:304–13.
[34] Peng Z, et al. A novel deep learning ensemble model with data denoising for short--term wind speed forecasting. Energy Convers Manag 2020;207:112524.
[35] Ghimire S, Deo RC, Raj N, Mi J. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. Appl Energy 2019;253:113541.
[36] Lee W, Kim K, Park J, Kim J, Kim Y. Forecasting solar power using long-short term memory and convolutional neural networks. IEEE Access 2018;6:73068–80.
[37] Wu W, Chen K, Qiao Y, Lu Z. Probabilistic short-term wind power forecasting based on deep neural networks. In: Proceedings of the 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS); 2016. p. 1–8.
[38] Zhang J, Yan J, Infield D, Liu Y, Lien F. Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model. Appl Energy 2019;241:229–44.
[39] Ray B, Shah R, Islam MR, Islam S. A new data driven long-term solar yield analysis model of photovoltaic power plants. IEEE Access 2020;8:136223–33.
[40] Sharadga H, Hajimirza S, Balog RS. Time series forecasting of solar power generation for large-scale photovoltaic plants. Renew Energy 2020;150:797–807.
[41] Pearson RK, Neuvo Y, Astola J, Gabbouj M. Generalized Hampel Filters. EURASIP J Adv Signal Process 2016(1):87 2016. doi:10.1186/s13634-016-0383-6.
[42] Zang H, et al. Application of functional deep belief network for estimating daily global solar radiation: a case study in China. Energy 2020;191:116502.
[43] Rana SS. Quadratic spline interpolation. J Approx Theory 1989;57(3):300–5.
[44] Lima MAFB, Carvalho PCM, Fernández-Ramírez LM, Braga APS. Improving solar forecasting using Deep Learning and Portfolio Theory integration. Energy 2020;195:117016.
[45] Lin Z, Liu X, Collu M. Wind power prediction based on high-frequency SCADA data along with isolation forest and deep learning neural networks. Int J Electr Power Energy Syst 2020;118:105835.
[46] Liu H, Chen C. Data processing strategies in wind energy forecasting models and applications: a comprehensive review. Appl Energy 2019;249:392–408.
[47] Liu H, Chen C. Multi-objective data-ensemble wind speed forecasting model with stacked sparse autoencoder and adaptive decomposition-based error correction. Appl Energy 2019;254:113686.
[48] Pang Z, Niu F, O'Neill Z. Solar radiation prediction using recurrent neural network and artificial neural network: a case study with comparisons. Renew Energy 2020;156:279–89.
[49] Hong Y-Y, Rioflorido CLPP. A hybrid deep learning-based neural network for 24-h ahead wind power forecasting. Appl Energy 2019;250:530–9.
[50] Liu Y, et al. Ensemble spatiotemporal forecasting of solar irradiation using variational Bayesian convolutional gate recurrent unit network. Appl Energy 2019;253:113596.
[51] Wang F, et al. Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting. Energy Convers Manag 2019;181:443–62.
[52] Wang K, Qi X, Liu H. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. Appl Energy 2019;251:113315.
[53] Yin H, Ou Z, Huang S, Meng A. A cascaded deep learning wind power prediction approach based on a two-layer of mode decomposition. Energy 2019;189:116316.
[54] Ospina J, Newaz A, Faruque MO. Forecasting of PV plant output using hybrid wavelet-based LSTM-DNN structure model. IET Renew Power Gener 2019;13(7):1087–95.
[55] Chen J, et al. Learning heterogeneous features jointly: a deep end-to-end framework for multi-step short-term wind power prediction. IEEE Trans Sustain Energy 2020;11(3):1761–72. doi:10.1109/TSTE.2019.2940590.
[56] Qureshi AS, Khan A, Zameer A, Usman A. Wind power prediction using deep neural network based meta regression and transfer learning. Appl Soft Comput 2017;58:742–55.
[57] Memarzadeh G, Keynia F. A new short-term wind speed forecasting method based on fine-tuned LSTM neural network and optimal input sets. Energy Convers Manag 2020;213:112824.
[58] Khodayar M, Wang J, Manthouri M. Interval deep generative neural network for wind speed forecasting. IEEE Trans Smart Grid 2018;10(4):3974–89.
[59] Chen Y, Zhang S, Zhang W, Peng J, Cai Y. Multifactor spatio-temporal correlation model based on a combination of convolutional neural network and long short--term memory neural network for wind speed forecasting. Energy Convers Manag 2019;185:783–99.
[60] Xiaoyun Q, Xiaoning K, Chao Z, Shuai J, Xiuda M. Short-term prediction of wind power based on deep long short-term memory. In: Proceedings of the 2016 IEEE PES Asia-pacific power and energy engineering conference (APPEEC); 2016. p. 1148–52.
[61] Liang S, Nguyen L, Jin F. A multi-variable stacked long-short term memory network for wind speed forecasting. In: Proceedings of the 2018 IEEE international conference on big data (Big Data); 2018. p. 4561–4.
[62] Hossain MS, Mahmood H. Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast. IEEE Access 2020;8:172524–33.
[63] Wang F, Xuan Z, Zhen Z, Li K, Wang T, Shi M. A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework. Energy Convers Manag 2020;212:112766.
[64] Qin Y, et al. Hybrid forecasting model based on long short term memory network and deep learning neural network for wind signal. Appl Energy 2019;236:262–72.
[65] Yu R, et al. LSTM-EFG for wind power forecasting based on sequential correlation features. Futur Gener Comput Syst 2019;93:33–42.
[66] Chai S, Xu Z, Jia Y, Wong WK. A robust spatiotemporal forecasting framework for photovoltaic generation. IEEE Trans Smart Grid 2020;11(6):5370–82.
[67] 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 2020;160:26–41.
[68] Huang C-J, Kuo P-H. Multiple-input deep convolutional neural network model for short-term photovoltaic power forecasting. IEEE Access 2019;7:74822–34.
[69] Yu C, Li Y, Bao Y, Tang H, Zhai G. A novel framework for wind speed prediction based on recurrent neural networks and support vector machine. Energy Convers Manag 2018;178:137–45.
[70] Niu Z, Yu Z, Tang W, Wu Q, Reformat M. Wind power forecasting using attention-based gated recurrent unit network. Energy 2020;196:117081.
[71] Zhou H, Zhang Y, Yang L, Liu Q, Yan K, Du Y. Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism. IEEE Access 2019;7:78063–74.
[72] Sorkun MC, Paoli C, Incel ÖD. Time series forecasting on solar irradiation using deep learning. In: Proceedings of the 10th international conference on electrical and electronics engineering (ELECO); 2017. p. 151–5.
[73] Wen L, Zhou K, Yang S, Lu X. Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting. Energy 2019;171:1053–65.
[74] Abdel-Nasser M, Mahmoud K. Accurate photovoltaic power forecasting models using deep LSTM-RNN. Neural Comput Appl 2019;31(7):2727–40.
[75] Shahid F, Zameer A, Mehmood A, Raja MAZ. A novel wavenets long short term memory paradigm for wind power prediction. Appl Energy 2020;269:115098.
[76] Li C, Tang G, Xue X, Chen X, Wang R, Zhang C. The short-term interval prediction of wind power using the deep learning model with gradient descend optimization. Renew Energy 2020;155:197–211.
[77] Hu T, Guo Q, Li Z, Shen X, Sun H. Distribution-free probability density forecast through deep neural networks. IEEE Trans Neural Netw Learn Syst 2019;31(2):612–25.
[78] Suresh V, Janik P, Guerrero JM, Leonowicz Z, Sikorski T. Microgrid energy management system with embedded deep learning forecaster and combined optimizer. IEEE Access 2020;8:202225–39.
[79] Hu Q, Zhang R, Zhou Y. Transfer learning for short-term wind speed prediction with deep neural networks. Renew Energy 2016;85:83–95.
[80] Khodayar M, Kaynak O, Khodayar ME. Rough deep neural architecture for short--term wind speed forecasting. IEEE Trans. Ind. Inform. 2017;13(6):2770–9.
[81] Jiao R, Huang X, Ma X, Han L, Tian W. A model combining stacked auto encoder and back propagation algorithm for short-term wind power forecasting. IEEE Access 2018;6:17851–8.
[82] Neo YQ, Teo TT, Woo WL, Logenthiran T, Sharma A. Forecasting of photovoltaic power using deep belief network. In: Proceedings of the Tencon 2017-2017 IEEE Region 10 Conference; 2017. p. 1189–94.
[83] Lin K-P, Pai P-F, Ting Y-J. Deep belief networks with genetic algorithms in forecasting wind speed. IEEE Access 2019;7:99244–53.
[84] Luo X, et al. Short-term wind speed forecasting via stacked extreme learning machine with generalized correntropy. IEEE Trans. Ind. Inform. 2018;14(11):4963–71.
[85] Hu H, Wang L, Lv S-X. Forecasting energy consumption and wind power generation using deep echo state network. Renew Energy 2020;154:598–613.
[86] Li Q, Wu Z, Ling R, Feng L, Liu K. Multi-reservoir echo state computing for solar irradiance prediction: a fast yet efficient deep learning approach. Appl Soft Comput 2020;95:106481.
[87] Sideratos G, Hatziargyriou ND. A distributed memory RBF-based model for variable generation forecasting. Int J Electr Power Energy Syst 2020;120:106041.

[88] Chen J, Zeng G-Q, Zhou W, Du W, Lu K-D. Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. Energy Convers Manag 2018;165:681–95.
[89] Hu Y-L, Chen L. A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and Differential Evolution algorithm. Energy Convers Manag 2018;173:123–42.
[90] Wang HZ, Wang GB, Li GQ, Peng JC, Liu YT. Deep belief network based deterministic and probabilistic wind speed forecasting approach. Appl Energy 2016;182:80–93.
[91] Wang H, Li G, Wang G, Peng J, Jiang H, Liu Y. Deep learning based ensemble approach for probabilistic wind power forecasting. Appl Energy 2017;188:56–70.
[92] Wang H, et al. Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. Energy Convers Manag 2017;153:409–22.
[93] Mishra M, Dash PB, Nayak J, Naik B, Swain SK. Deep learning and wavelet transform integrated approach for short-term solar PV power prediction. Measurement 2020;166:108250.
[94] Jiajun H, Chuanjin Y, Yongle L, Huoyue X. Ultra-short term wind prediction with wavelet transform, deep belief network and ensemble learning. Energy Convers Manag 2020;205:112418.
[95] Li F, Ren G, Lee J. Multi-step wind speed prediction based on turbulence intensity and hybrid deep neural networks. Energy Convers Manag 2019;186:306–22.
[96] Liu H, Duan Z, Chen C, Wu H. A novel two-stage deep learning wind speed forecasting method with adaptive multiple error corrections and bivariate Dirichlet process mixture model. Energy Convers Manag 2019;199:111975.
[97] Li P, Zhou K, Lu X, Yang S. A hybrid deep learning model for short-term PV power forecasting. Appl Energy 2020;259:114216.
[98] Dolatabadi A, Abdeltawab H, Mohamed YA-RI. Hybrid deep learning-based model for wind speed forecasting based on DWPT and Bidirectional LSTM network. IEEE Access 2020;8:229219–32.
[99] Hwangbo S, Nam K, Heo S, Yoo C. Hydrogen-based self-sustaining integrated renewable electricity network (HySIREN) using a supply-demand forecasting model and deep-learning algorithms. Energy Convers Manag 2019;185:353–67.
[100] Nam K, Hwangbo S, Yoo C. A deep learning-based forecasting model for renewable energy scenarios to guide sustainable energy policy: a case study of Korea. Renew Sustain Energy Rev 2020;122:109725.
[101] Yang H-F, Chen Y-PP. Representation learning with extreme learning machines and empirical mode decomposition for wind speed forecasting methods. Artif Intell 2019;277:103176.
[102] He X, Nie Y, Guo H, Wang J. Research on a novel combination system on the basis of deep learning and swarm intelligence optimization algorithm for wind speed forecasting. IEEE Access 2020;8:51482–99.
[103] Wang J, Li Y. Multi-step ahead wind speed prediction based on optimal feature extraction, long short term memory neural network and error correction strategy. Appl Energy 2018;230:429–43.
[104] Wang L, Tao R, Hu H, Zeng Y-R. Effective wind power prediction using novel deep learning network: stacked independently recurrent autoencoder. Renew Energy 2021;164:642–55.
[105] Yildiz C, Acikgoz H, Korkmaz D, Budak U. An improved residual-based convolutional neural network for very short-term wind power forecasting. Energy Convers Manag 2021;228:113731.
[106] Yan H, Wu Z. A hybrid short-term wind power prediction model combining data processing, multiple parameters optimization and multi-intelligent models apportion strategy. IEEE Access 2020;8:227126–40.
[107] Li Y, Wu H, Liu H. Multi-step wind speed forecasting using EWT decomposition, LSTM principal computing, RELM subordinate computing and IEWT reconstruction. Energy Convers Manag 2018;167:203–19.
[108] Pei S, et al. Wind speed prediction method based on empirical wavelet transform and new cell update long short-term memory network. Energy Convers Manag 2019;196:779–92.
[109] Liu H, Mi X, Li Y. Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network. Energy Convers Manag 2018;156:498–514.
[110] Gao B, Huang X, Shi J, Tai Y, Zhang J. Hourly forecasting of solar irradiance based on CEEMDAN and multi-strategy CNN-LSTM neural networks. Renew Energy 2020;162:1665–83.
[111] Liang T, Xie G, Fan S, Meng Z. A combined model based on CEEMDAN, permutation entropy, gated recurrent unit network, and an improved bat algorithm for wind speed forecasting. IEEE Access 2020;8:165612–30.
[112] Liu H, Mi X, Li Y, Duan Z, Xu Y. Smart wind speed deep learning based multi-step forecasting model using singular spectrum analysis, convolutional Gated Recurrent Unit network and Support Vector Regression. Renew Energy 2019;143:842–54.
[113] Zhang Y, Le J, Liao X, Zheng F, Li Y. A novel combination forecasting model for wind power integrating least square support vector machine, deep belief network, singular spectrum analysis and locality-sensitive hashing. Energy 2019;168:558–72.
[114] Mi X, Zhao S. Wind speed prediction based on singular spectrum analysis and neural network structural learning. Energy Convers Manag 2020;216:112956.
[115] Yan X, Liu Y, Xu Y, Jia M. Multistep forecasting for diurnal wind speed based on hybrid deep learning model with improved singular spectrum decomposition. Energy Convers Manag 2020;225:113456.
[116] Zang H, Cheng L, Ding T, Cheung KW, Wei Z, Sun G. Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning. Int J Electr Power Energy Syst 2020;118:105790.
[117] Fu W, Wang K, Tan J, Zhang K. A composite framework coupling multiple feature selection, compound prediction models and novel hybrid swarm optimizer-based synchronization optimization strategy for multi-step ahead short-term wind speed forecasting. Energy Convers Manag 2020;205:112461.
[118] Mi X, Liu H, Li Y. Wind speed prediction model using singular spectrum analysis, empirical mode decomposition and convolutional support vector machine. Energy Convers Manag 2019;180:196–205.
[119] Liu H, Mi X, Li Y. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. Energy Convers Manag 2018;159:54–64.
[120] Xiang L, Li J, Hu A, Zhang Y. Deterministic and probabilistic multi-step forecasting for short-term wind speed based on secondary decomposition and a deep learning method. Energy Convers Manag 2020;220:113098.
[121] Ma Z, et al. Application of hybrid model based on double decomposition, error correction and deep learning in short-term wind speed prediction. Energy Convers Manag 2020;205:112345.
[122] Chen G, Li L, Zhang Z, Li S. Short-term wind speed forecasting with principle–subordinate predictor based on Conv-LSTM and improved BPNN. IEEE Access 2020;8:67955–73.
[123] Woo S, Park J, Park J, Manuel L. Wind field-based short-term turbine response forecasting by stacked dilated convolutional LSTMs. IEEE Trans Sustain Energy 2020;11(4):2294–304.
[124] Wang K, Qi X, Liu H, Song J. Deep belief network based k-means cluster approach for short-term wind power forecasting. Energy 2018;165:840–52.
[125] Duan Z, Liu H. An evolution-dependent multi-objective ensemble model of vanishing moment with adversarial auto-encoder for short-term wind speed forecasting in Xinjiang wind farm, China. Energy Convers Manag 2019;198:111914.
[126] Aly HHH. A novel deep learning intelligent clustered hybrid models for wind speed and power forecasting. Energy 2020;213:118773.
[127] Yan J, Zhang H, Liu Y, Han S, Li L, Lu Z. Forecasting the high penetration of wind power on multiple scales using multi-to-multi mapping. IEEE Trans Power Syst 2018;33(3):3276–84.
[128] Li G, Xie S, Wang B, Xin J, Li Y, Du S. Photovoltaic power forecasting with a hybrid deep learning approach. IEEE Access 2020;8:175871–80.
[129] Wang J, Guo L, Zhang C, Song L, Duan J, Duan L. Thermal power forecasting of solar power tower system by combining mechanism modeling and deep learning method. Energy 2020;208:118403.
[130] Afrasiabi M, Mohammadi M, Rastegar M, Kargarian A. Multi-agent microgrid energy management based on deep learning forecaster. Energy 2019;186:115873.
[131] Wang K, Qi X, Liu H. Photovoltaic power forecasting based LSTM-convolutional network. Energy 2019;189:116225.
[132] Chang GW, Lu H-J. Integrating gray data preprocessor and deep belief network for day-ahead PV power output forecast. IEEE Trans Sustain Energy 2018;11(1):185–94.
[133] Gao M, Li J, Hong F, Long D. Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using LSTM. Energy 2019;187:115838.
[134] Zhang H, Liu Y, Yan J, Han S, Li L, Long Q. Improved deep mixture density network for regional wind power probabilistic forecasting. IEEE Trans Power Syst 2020;35(4):2549–60.
[135] Manero J, Béjar J, Cortés U. 'Dust in the wind...', deep learning application to wind energy time series forecasting. Energies 2019;12(12):2385.