



A Comprehensive Review on Ensemble Solar Power Forecasting Algorithms

Negar Rahimi¹ · Sejun Park¹ · Wonseok Choi¹ · Byoungryul Oh¹ · Sookyung Kim¹ · Young-ho Cho¹ · Sunghyun Ahn¹ · Chulho Chong¹ · Daewon Kim¹ · Cheong Jin² · Duehee Lee¹

Received: 29 June 2022 / Revised: 25 December 2022 / Accepted: 3 January 2023 / Published online: 12 January 2023
© The Author(s) 2023, corrected publication 2023

Abstract

With increasing demand for energy, the penetration of alternative sources such as renewable energy in power grids has increased. Solar energy is one of the most common and well-known sources of energy in existing networks. But because of its non-stationary and non-linear characteristics, it needs to predict solar irradiance to provide more reliable Photovoltaic (PV) plants and manage the power of supply and demand. Although there are various methods to predict the solar irradiance. This paper gives the overview of recent studies with focus on solar irradiance forecasting with ensemble methods which are divided into two main categories: competitive and cooperative ensemble forecasting. In addition, parameter diversity and data diversity are considered as competitive ensemble forecasting and also preprocessing and post-processing are as cooperative ensemble forecasting. All these ensemble forecasting methods are investigated in this study. In the end, the conclusion has been drawn and the recommendations for future studies have been discussed.

Keywords Ensemble methods · Solar forecasting · Cooperative ensemble forecasting

1 Introduction

Over the past decades, the installation and use of renewable energy resources have been increased due to environmental issues such as greenhouse gas emission and the reduction of fossil fuels [1]. Solar energy has been considered as one of the promising renewable energy resources that is able to supply global energy demand [2]. According to the International Renewable Energy Agency (IRENA), the top ten

countries with the highest installed solar PV capacity are shown in Fig. 1 [3].

According to Renewables 2015-Global Status Report, the global capacity of PV was increased from 3.7 GW to 7 GW between 2004 and 2007 and promoted to 40 GW by 2010. In 2015, the global installed capacity of solar PV was around 177 GW [4]. From 2016 to 2018, based on IRENA, there was a significant rise in this amount, and it was reached about 580 GW in 2019 [5]. It has been reported that in the

✉ Duehee Lee
hello.keanu@konkuk.ac.kr

Negar Rahimi
rahiminegar43@gmail.com

Sejun Park
alpheon6@konkuk.ac.kr

Wonseok Choi
hipive0301@konkuk.ac.kr

Byoungryul Oh
qkdlzldxn@konkuk.ac.kr

Sookyung Kim
spoint2000@konkuk.ac.kr

Young-ho Cho
jacobcho@utexas.edu

Sunghyun Ahn
ka4ka400@konkuk.ac.kr

Chulho Chong
jch9852@konkuk.ac.kr

Daewon Kim
kimdw010130@konkuk.ac.kr

Cheong Jin
cheong.jin@gmail.com

¹ Department of Electrical and Electronic Engineering, Konkuk University, Seoul, South Korea

² EINS S&C, Seoul, South Korea

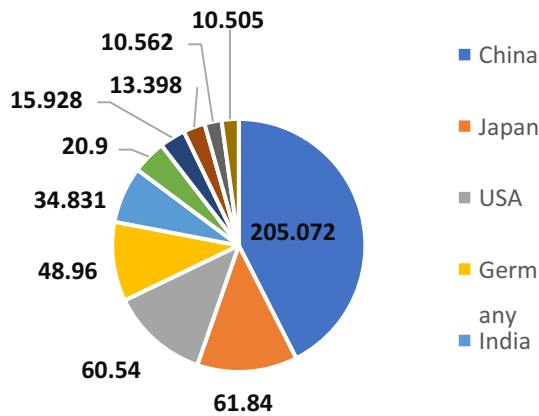


Fig. 1 Top 10 countries in 2019 based on total PV installed capacity (GW) [3]

first half of 2020, when the COVID-19 pandemic began, the renewable energy sector experienced a turbulent but in the second half of 2020, the PV sector set a new record for new capacity which bring the solar sector to its highest level of more than 800 Gigawatts [6].

However, the stochastic and unpredictable nature of solar energy led to a number of challenges such as voltage fluctuations and uncertainties in a power grid that make it difficult to maintain a balance between power generation and load demand [7]. For example, in cloudy days, the amount of solar irradiation received by PV modules has many fluctuations because of the cloud's movement, and the random fluctuations have a significant effect on the solar PV output. Various solutions including power system, scheduling battery reserves, demand response, backup generators, and dispatches have been proposed to prevent the above problems. However, several limitations exist in all these solutions. There is a restriction for the power decline rate of backup generators which caused by the unit ramp rate resulted in difficulties to meet the incremental power generation need. However, due to the storage capacity restrictions and the costs of battery reserves, massive-scale energy storage is still difficult to realize. In addition, there is a difficulty to achieve demand response technologies based on the lack information on behavior of electricity consumption by consumers. Besides, the efficiency of the above-mentioned solutions largely depends on forecasting accuracy (with different time horizons) [8].

There are two important aspects of accurate forecasting: reducing the negative effect of random PV power on the power grid and providing and predicting PV power output data for grid operators. Hence, there is a need to forecast the output power of solar systems for the efficient operation of the power grid. The optimal management of a power system and scheduling is important for estimation of the reserves. It should be noted that the solar forecasting becomes important

due to the substantial increase of solar power generation worldwide [9].

Recently, the ensemble forecasting was recommended for solar power forecasting. In the ensemble forecasting, many different predictions from different forecasting are averaged. Averaging predictions can reduce server biases when weather data in outliers, so it can avoid the worst predictions. However, it takes a lot of computational time to run the ensemble forecasting model since several forecasting models should be simulated at the same time. However, this computational time can be further reduced when the parallel computing environment is used. In this study, we organize several ensemble solar power forecasting algorithms.

For forecasting methods of PV systems, several review papers have been published during the last 5 years with different scopes. Their focus was ensemble methods, PV output power forecasting different PV forecasting methods, probabilistic forecasting in solar PV [10], hybrid models for solar radiation forecasting, post-processing in solar forecasting, different methods for forecasting solar irradiance. Hence, there is no updated review with a focus on ensemble methods only. Therefore, we reviewed the ensemble methods for solar irradiation forecasting which are divided into main categories competitive (data diversity and parameter diversity) and Cooperative forecasting (pre-processing and post-processing). The recent research papers that have been selected which have been published in the last four years (2018–2022) were reviewed. The above-mentioned papers focused on reviewing solar forecasting methods. In this paper, the focus was ensemble forecasting methods and their classifications in recent years.

For the ensemble forecasting, there are two topics, namely, solar power forecasting and solar irradiance forecasting which are known as solar forecasting. Meanwhile, they have strongly correlated each other and cannot be separated [11]. It should be noted that there is a strong impact of solar irradiance on the accuracy of solar power forecasts for the power production systems of various sorts. If a forecaster wants to achieve high-quality solar power forecasts, the ability to produce and use irradiance forecasts is essential. In other words, the best solar power prediction is always obtained by irradiance forecasts generation and conversion of those irradiance forecasts into power forecasts through a model chain. Therefore, there is not an intrinsic value for irradiance forecast [12].

2 Parameters Affecting Solar Power Forecasting

Forecast horizon, weather classification, error metrics, data processing affect the output of solar forecasting.

2.1 Forecast Horizon

The forecast horizon can be considered as the period of time in the future (time duration between actual and effective time) in which the forecasting should be done [13]. Forecasting horizon can be classified into four categories including (1) very short-term, (2) short-term, (3) medium-term, and (4) long-term [14]. These categories have summarized in Table 1 based on their forecasting period and related applications in solar energy systems. However, there is still no universal classification criterion [15].

Furthermore, the researchers usually prefer to use another categories to describe forecast horizon including intra-hour, intra-day, and day-ahead which have overlapping with short, medium, and long-term forecast horizons [16].

2.1.1 Intra-Hour

Intra-hour overlaps with very short-term and short-term horizon categories and it also shows the forecast horizon from a few seconds to an hour. It is used for operating regulation reserves, storage system optimization and ensuring grid quality and stability. Such prediction methods can be applied in high solar penetration areas such as island grids with spans of 1 to 6 h.

2.1.2 Intra-Day

This forecast horizon is used for 1–6 h and also overlaps with short and medium categories and its application is in electricity trading outside the standard grid and control of electric loads [13].

2.1.3 Day-Ahead

Day-ahead Forecasts spanning 6–48 h overlaps with long-term and medium horizons. Similar models have been used in unit commitment and utilities planning [13]. In general, before designing PV power forecasting model, the

appropriate forecast time horizon should be selected because the accuracy of a predicted model depends on the forecast time horizon [17]. It has been proved that by increasing the forecasting horizon, the accuracy of forecasting model (both single and ensemble models) will reduce. This is because the correlation between cloud cover and solar irradiance, which cannot be accurately predicted for long periods of time. In addition, for power system planning, long time horizon forecast is suitable while for PV output forecasting, intra-hour and intra-day forecast horizons work better [13]. In other words, the model performance decreases with long-time ahead forecasting (regardless of model type) while the performance would be increased with short time ahead forecasting [18].

2.2 Weather Classification

It is obvious that the PV output is directly related to solar irradiance and the accuracy solar irradiance forecasting models is strongly affected by meteorological factors such as cloud cover, temperature, humidity and wind speed. So, climate change and different weather types have significant effect on PV system output power. Therefore, to enhance the prediction performance, weather condition is an effective step especially for solar irradiance forecasting [19]. Typically, there are two types of weather condition including the normal (ideal) weather type (sunny days), and abnormal (non-ideal) weather types (rainy, foggy and windy days) [20]. According to some studies [19–21], the PV output power increases in the ideal weather condition (on sunny days) but it decreases in non-ideal conditions.

2.3 Error Metrics

In different steps of model development, evaluation error metrics are one of the important parameters. In these kinds of metrics, the comparison of the actual solar irradiance and predicted solar irradiance are considered [22]. It should be noticed that there are different units for performance of

Table 1 Different types of forecasting horizon and its applications [15]

Forecasting horizon	Forecasting period	Applications
Very short-term	From seconds to 1 min ahead	– Electricity market – Monitoring of real-time electricity dispatch – PV storage control
Short-term	Several hours ahead	Unit commitment
Medium-term	One week or within a month ahead	Scheduling of power systems maintenance (solar energy integrated power or conventional systems)
Long-term	Months or Years ahead	– Electricity pricing – Load forecasting

metrics while W/m² has been used as unit for the statistical error of solar radiation [18]. The most used evaluation metrics for statistical measures are listed as below:

2.3.1 Mean Bias Error (MBE)

Mean bias error (MBE): It shows the average bias of a forecasting model:

$$MBE = \frac{1}{N} \sum_{i=1}^N (y_{i(pred)} - y_{i(act)}) \quad (1)$$

The larger value of MBE shows the larger forecast bias. A positive value of MBE indicates over-forecasting, while a negative value means under-forecasting [23].

2.3.2 Mean Absolute Error (MAE)

Mean absolute error (MAE): This error is defined as the average of the absolute difference between forecasted and actual solar irradiance values. This metric is suitable for uniform forecast errors as equal weight will be given to all discrepancies in the data, and it is also used for both regression problems and evaluation of overall forecast accuracy. Generally, the smaller MAE is better in forecasting.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_{i(pred)} - y_{i(act)}| \quad (2)$$

where $y_{i(act)}$ is the actual solar irradiance, N is the total number of observations and $y_{i(pred)}$ is the predicted solar irradiance [22, 24].

2.3.3 Mean Square Error (MSE)

Mean square error (MSE): It is calculated by averaging the square of difference between the actual and predicted solar irradiance values [25].

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_{i(pred)} - y_{i(act)})^2 \quad (3)$$

2.3.4 Root Mean Square Error (RMSE)

Root mean square error (RMSE) is for calculating this metric, the square root for the average of the squared differences of predicted and actual solar irradiance values is considered. RMSE is known as the most appreciated performance evaluation metrics which the outliers in the data can be eliminated and identified by using this error [25]. Also, as RMSE emphasizes the larger errors, it can be used as the main error metric [26].

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{i(pred)} - y_{i(act)})^2} \quad (4)$$

2.3.5 Normalized RMSE (nRMSE)

Normalized RMSE (nRMSE) is the overall deviations of larger datasets nRMSE can be calculated.

$$nRMSE = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_{i(pred)} - y_{i(act)})^2}}{\bar{y}} \quad (5)$$

where (\bar{y}) denotes the mean of the actual solar irradiance [27].

2.3.6 Mean Absolute Percentage Error (MAPE)

Mean absolute percentage error (MAPE) is appropriate to evaluate uniform prediction errors like MAE however it can be calculated by the difference between each predicted and actual observation divide by the actual observed value.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_{i(pred)} - y_{i(act)}|}{y_{i(act)}} \quad (6)$$

2.3.7 Determination Coefficient (R^2)

Determination coefficient (R^2) is used to extract the information of the determination between the forecasted and the real values and the ranges is between 0 and 1 [28].

2.3.8 Skill Score (SS)

Skill score (SS) measures the performance comparison of a forecast model to a benchmark model is defined as S score and calculated as follows:

$$S = 1 - \frac{RMSE_{forecast\ model}}{RMSE_{benchmark\ model}} \quad (7)$$

A good forecast model has a SS score of 1. In addition, SS zero shows a model with forecast error equal to the benchmark model while a negative SS score is a forecast with higher forecast error [1].

2.4 Model Inputs

Inputs are considered as a key factor in forecasting and have a significant effect on prediction accuracy. Generally, inappropriate inputs can cause forecast errors in a system, for example, time delay, cost, and computational complexity [13]. In addition, the

correlation of the input and output values affects the performance of a forecasting model. Therefore, the correlation of PV power output with the different meteorological inputs, such as wind speed and direction, module temperature, atmospheric temperature, solar irradiance, and humidity is important [17]. Among these parameters, solar irradiance is the most significant input for the forecast and the accuracy of solar irradiance measurement affects the precision of solar power generation [29]. Demonstrated the highest influence in solar power generation related to the intensity of solar irradiance. In [30] a SVR-based forecasting model was proposed for PV power generation forecasting. In this study, the data of three different PV plants, in Malaysia, including the actual PV power generation data and meteorological data (wind speed atmospheric temperature and solar irradiance) were used and PV power output was received only from 8:00 AM until 19:00 PM. Also, PV output for a specific day and the pattern of solar irradiance was presented and the correlation coefficient ($R^2=0.9888$) was obtained. The correlation coefficient indicated that there is a strong correlation between solar irradiance and PV power output. It means that if PV power increases, then the irradiance will be increased, and vice versa. Other studies also showed that the PV power generation are correlated with other meteorological variables, such as clouds, module temperature, ambient temperature, and wind speed. Also, fluctuations in sunlight intensity caused by several parameters such as dissipation, cloud motion, deformation and birth which affect PV output [31]. In another study by [32], PV output variation with solar irradiance was analyzed during clear day, partially cloudy and cloudy day. Based on the results, in comparison with to partially cloudy and clear day, there was a high fluctuations degree with solar irradiance were observed in cloudy days.

Another input variables linked together are wind speed, ambient and module temperature. The temperature of the module and its level of efficiency depend on the ambient temperature and the amount of absorbed radiation, as well as the wind speed. When the ambient temperature increases (It is related to cloud cover), the module temperature rises and its efficiency also increases, while the wind speed reduces the module temperature and heat loss [33]. It is concluded that there was a weak correlation between atmospheric temperature and PV power output while there was an extremely weak correlation between wind speed and PV power output. In contrast, there was a strong correlation between module temperature and PV power output [30].

3 Classification of Solar Forecasting Algorithms

This section classifies solar forecasting algorithms. The solar forecasting algorithms are categorized into three main models, such as ensemble methods physical and statistical time-series.

Group one is statistical models, which are based on historical data as input for prediction and the internal behavior of the model does not have any effect on it [34]. In other words, to reconstruct the hourly irradiance and meteorological parameters statistical models can be used [35]. Different techniques are applied for statistical time series such as support vector machine and Markov chain, regression model, and artificial neural network [34].

The second group generated solar power output from external data such as temperature and solar irradiance and their physical relationship. The generation of irradiance forecast between many meteorological parameters would be achieved by using mathematical models. It can be said that the results are highly dependent on the accurate input data of the model [36]. This group can be categorized into two subsections of Empirical models and Numerical weather prediction (NWP) [37].

The last group is hybrid or ensemble methods which used the combination of physical and statistical methods with unique features to solve the limitations of an individual model. By these models, at the same time, the forecasting performance increases, and the model would be improved. Figure 2 shows these forecasting methods [38].

For forecasting methods of PV systems, several review papers have been published during the last 5 years with different scopes. Their focus was ensemble methods, PV output power forecasting [14, 32] different PV forecasting methods, probabilistic forecasting in solar PV [40], hybrid models for solar radiation forecasting [41], post-processing in solar forecasting [42], different methods

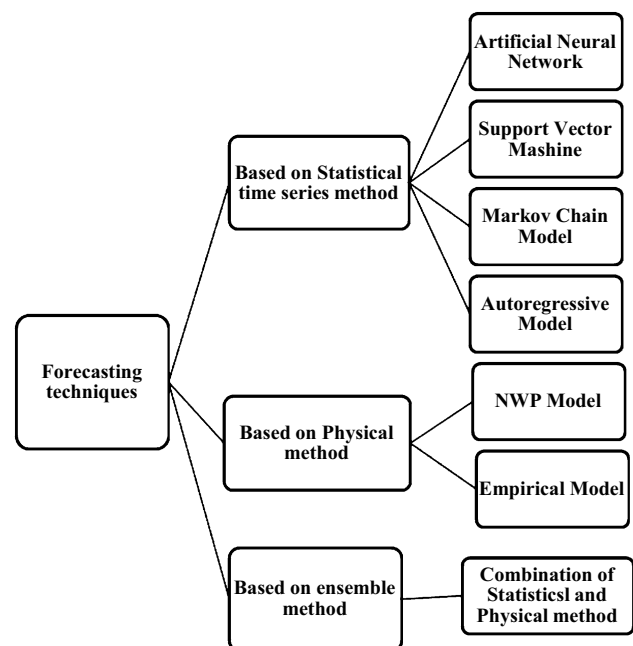


Fig. 2 Forecasting techniques based on three major models [39]

for forecasting solar irradiance [24]. Hence, there is no updated review with a focus on ensemble methods only. Therefore, we reviewed the ensemble methods for solar irradiation forecasting which are divided into main categories competitive (data diversity and parameter diversity) and Cooperative forecasting (pre-processing and post-processing). The recent research papers that have been selected which have been published in the last four years (2018–2022) were reviewed. The above-mentioned papers focused on reviewing solar forecasting methods. In this paper, the focus was ensemble forecasting methods and their classifications in recent years.

4 Ensemble Forecasting

The main concept of the ensemble technique is training ensemble members (base learners) and combining their prediction into a single output to obtain a better performance of a model [10]. Figure 3 shows the typical ensemble model construction. As can be seen in the Fig. 3, the training data is divided into many data sets initially and then several base learners are generated which can be run in a sequential or parallel format. Finally, the combination of the models is run on the base learners [43, 44].

On other words, Ensemble (hybrid, or combined technique) [45] is a machine learning-based method in which

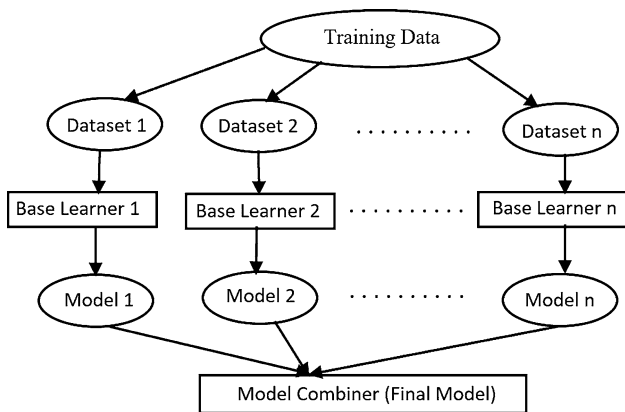


Fig. 3 General construction flow of an ensemble model

Table 2 The classification of ensemble forecasting models [49]

Categories	Perspectives	Methods
Competitive ensemble methods	–Data diversity –Parameter diversity	– Bagging, Boosting – Ensemble Kalman filter
Cooperative ensemble methods	–Pre-processing –Post-processing	– Wavelet decomposition (WD), empirical mode decomposition (EMD) – ARMA-GARCH, ARMA-ANN, ARIMA-SVM

multiple predictors are used to reach an aggregated decision in a better format of base predictors. The main advantage of an ensemble model is that it incorporates the suitability of its constituent techniques, which creates a stronger learning pattern. So, it can enhance the accuracy and strengths of individual methods by solving their weakness [46]. The ensemble methods can be classified into two main groups: competitive and cooperative [39]. In a competitive ensemble forecasting, the different base predictors train individually with the same or different data sets by using different parameters, and the prediction is obtained from the averaging the decisions of all base predictors [47]. Besides, in a cooperative method, the prediction task will be divided into several sub-tasks to select the appropriate predictors based on the characterization of the sub-tasks. Then, the sum of all the base predictors’ output will be the final decision of the cooperative method [48]. Table 2 shows the classification of the ensemble methods and their perspectives in this paper.

4.1 Competitive Forecasting Methods

As mentioned earlier, a competitive forecast model can construct individual forecast models for the formation of an ensemble forecast model by using multiple predictors using different initial conditions or parameters. Then, the results will be obtained from the average of the selected models [50]. The competitive ensemble methods are categorized into three different perspectives which are data diversity, parameter diversity and structural diversity [51]. Diversity is the most important feature of competitive ensemble forecasting. For instance, the outputs of base predictors would be similar if the sub-tasks are similar while the performance improvement of ensemble predictor will be marginal [50]. Some examples of competitive ensemble forecasting models are discussed in detail in the following sections.

4.1.1 Data Diversity

Data diversity is another category of cooperative ensemble forecasting method in which the forecasting system is fed by more than one input dataset [37]. As shown in the following equations, two variations have been used in Eq. (9) and (10). Equation (9) applies N predictors ($f_1(0) \dots f_N(0)$) for N input

datasets $(x_1 \dots x_N)$ for forecasting and the weighted average of all of them will be the final prediction value, but Eq. (10) uses only a single predictor for forecasting.

$$\hat{y}(t+h) = \frac{1}{N} \sum_{i=1}^N w_i f_i(x_i(t)) \tag{8}$$

$$w_i \geq 0 \tag{9}$$

$$\sum_{i=1}^N w_i = 1 \tag{10}$$

$$\hat{y}(t+h) = f(x_1(t), x_2(t), \dots, x_N(t)) \tag{11}$$

where \hat{y} is the predicted value and h is the forecast horizon. The most famous and common approaches are bagging and boosting for data diversity [48].

4.1.1.1 Bagging For the first time, bagging or bootstrap aggregation is proposed by Breiman [52] which helps decrease the forecasting model variance and avoid overfitting [51]. The main goal of Bagging is developing several estimators, and results are obtained by aggregation of individual estimators' results with some biasing. Initially, from the original datasets, several subsets of training data are created. Then, the samples are selected randomly by replacement of data samples which are called Bootstrap samples. In the last step, the final prediction is obtained by aggregation of all the bootstrap predictions [53]. This process is shown in the Fig. 4. One of the advantages of bagging is the error reduction of ensemble generation in the baseline predictors. It also can correlate its estimation with real datasets using estimates of test sets or cross-validation [54].

The most common algorithm for bagging is random forest (RF) which can be considered as an extension to the bagging concept, and it can be used for classification and regression. It is made up and trained a large number of decision trees (DT), called predictors and each one produces their own predictions that can create higher accuracy in prediction [55, 56]. In the RF algorithm, several decision trees are constructed through training samples (a subset of training

samples are injected into each tree randomly). The output of RF is obtained by voting of the decision trees. In the case of classification, the majority of voting is used to decide on the predicted result, but in the case of regression, the mean value of the predictions of all the estimators (predictors) are calculated [57]. Figure 5 demonstrates the procedure of random forest method graphically. This algorithm does not need complex calculations like Support Vector Machine (SVM) and ANN because the main variables that should be adjusted is the number of trees. Also, compared to ANN and SVM, training process is faster in the RF. In this algorithm, if sufficient trees are used, it is robust to noise and outliers [58].

The application of this RF can be extended to PV production forecasting. Tato et al. used real radiation measurements by combining simple radiation predictions to forecast the solar energy output for short-term temporal horizons [60]. A daily PV power generation forecasting model was proposed for North China in winter. The proposed forecasting model was based on the RF algorithm using weather measures [61]. The accuracy, extra trees (ET), computational cost, and stability of RF were investigated for predicting hourly PV generation output. In addition, their performance was compared with support vector regression (SVR). The performance evaluation of the model was performed by RMSE, MAE and R^2 and it was concluded that RF performed better on

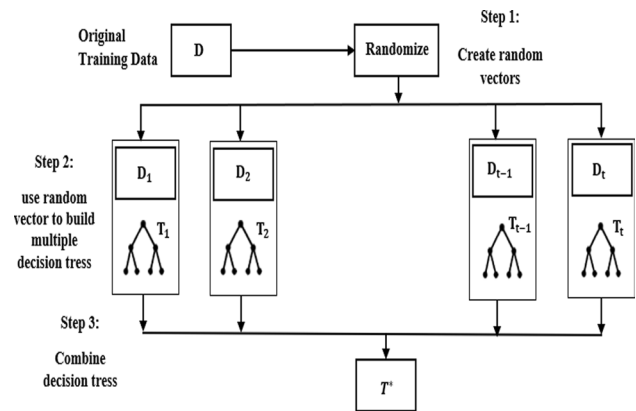
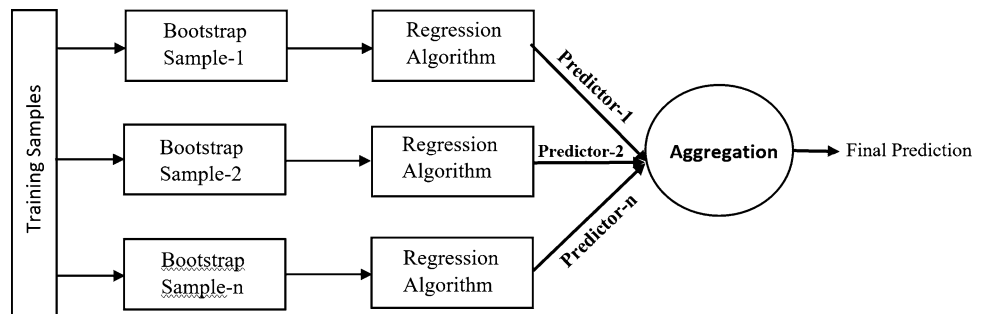


Fig. 5 The procedure of random forest method [59]

Fig. 4 Block diagram depicting the basic principle of Bagging ensemble learning



both testing datasets and training in comparison with other models [62].

4.1.1.2 Boosting To solve the classification and regression problems, boosting can be also used as a powerful learning strategy. It can be done by the combination of weak learners' output into a 'committee'. In addition, boosting improves the suitability of model to data through bias reduction [63]. In boosting process, several subsets can be created from the original dataset. Then it sequentially trains the predictors with the datasets, and at each iteration, smaller weights to the data points with smaller error and higher weights to the data points with larger error can be assigned by boosting. Finally, the results from the weighted average will be obtained [64]. Different algorithms can be used in bagging methods such as AdaBoost [65], extreme gradient boosting (XGBM) [64], gradient boosting machine (GBM) [66, 67], and light gradient boosting machine (LGBM) [68]. The recent advances in boosting algorithms and their applications in energy research such as solar and wind have been reviewed in [69]. They showed that how boosting algorithms are effective tools in the performance of prediction models.

Recently published studies are using bagging and boosting methods for forecasting solar irradiation in different regions. For instance, to improve forecasting accuracy in solar energy output, the bagging model was used in [64]. The proposed model as an ensemble method, used a based learner such as random forest, LGBMs, and XGB by addition of the past output data as new features. They found the bagging model was successful with higher model accuracy using past data features in comparison with a single model learner with default features [70].

In another study [44], the solar irradiance in five cities of Turkey, was estimated using bagging and boosting ensembles of ANN, SVR, and DT. Initially, base models (ANN, SVR, and DT) were created and examined by using 5 years of meteorological data. Then, both bagging and boosting ensembles methods of the base models were constructed and tested with the same data. The results were compared based on two evaluation metrics namely RMSE and R^2 . Based on their findings, the proposed model based on bagging and boosting methods improved ANN, DT, and SVR in the range of 4.6 and 14.6% in terms of RMSE. Several ensemble models to predict short-term solar irradiation were investigated in [71]. The models were RF, Boosted Trees, Generalized Random Forest, and Bagged Trees. The performance of these methods was validated via R^2 , and MAPE and compared with SVR and Gaussian process regression. Their result showed a consistent and reliable prediction of ensemble methods with high prediction performance in comparison with the individual regressors. In a recent study in 2022, a natural gradient boosting (NGBoost) algorithm was used for the short-term solar prediction of PV power systems

based on physical properties and human intuition. They found a 6% increase in RMSE with apply the most important features of the ensemble method [72]. The several tree-based methods of DT, GBM, XGBM, bagging, RF, and Cubist [73, 74] were used for forecasting solar irradiation on Jeju Island of South Korea. Also, the prediction performance was based on the comparison of MAE, RMSE, and nRMSE [75].

4.1.2 Parameter Diversity

In contrast to data diversity, parameter diversity is applied different parameter settings with the same dataset (x). The forecasts can be generated by using the following equation [71]:

$$\hat{y}(t+h) = \frac{1}{N} \sum_{i=1}^N f_i(x_i, \theta_i) \quad (12)$$

where h is the forecast horizon and \hat{y} is the predicted value. θ_i is the parameter for model f_i , $i = 1, \dots, N$. It should be noted that the combination in Eq. (11) uses equal weighting. In solar forecasting studies, once the initial conditions need to be perturbed, parameter diversity is a useful method as it considers numerical weather prediction models (NWP) [76, 77]. Kalman filter is the best method for NWP which estimates the "true" state of a dynamical system from noisy measurement data [76]. The Kalman filter processes in two stages including prediction and updating. In the first stage, there is a current state vector contains one or more state variables, and the state vector with a weighted average will be updated in the second stage. In addition, without additional past information the Kalman filter can run in real-time due to its recursive nature [78]. Recently, studies have been focused on solar forecasting using Kalman filtering.

Yang studied day-ahead NWP forecasts using Kalman filtering to forecast solar irradiation. The author used multiple Kalman filters to maintain the original day-ahead horizon [79]. Besides, an ensemble version of Kalman filter is called Ensemble Kalman Filter (EnKF) which can correct the forecasted value in real-time. It can be done by propagating the uncertainties in time [80]. In another study, the Ensemble Kalman filter (EnKF) and the state-space models (SSMs) in a short-term PV forecasting experiment which was effective to forecast solar irradiance [81].

Jiranantacharoen and Benjapolakul [82], used Kalman filter and Auto-Regressive Integrated Moving Average (ARIMA) for forecasting photovoltaic (PV) power generation. It should be noted that real-time measurement data is needed for Kalman filter to adjust forecast value. Then, they proposed a predictor model to apply in the forecasting process once the real-time measurement data is unavailable. In addition, to estimate the transition matrix for running

Kalman filter, ARIMA model was applied and the model performance by RMSE and SS.

4.2 Cooperative Forecasting Method

The cooperative ensemble forecasting is another type of forecasting method which divides a prediction task into several sub-tasks. These tasks can be solved individually. In addition, the output of one sub-task can be considered as input into another sub-task. The overall process of cooperative ensemble forecasting is distributed into two assignments. The first one is to realize the prediction task and the second are some assistant procedures such as parameter optimization, preprocessing, error correction, feature selection, and postprocessing.

Preprocessing includes reconstruction, decomposition, and transformation which should be performed before forecasting [49, 83].

4.2.1 Pre-Processing

Raw renewable energy data always has a variety of irregularities, such as fluctuations. These irregularities have nonlinear and non-static properties which deteriorate the performance of the forecasting. Therefore, many pre-processing techniques have been proposed to break down the renewable energy original signal into several components. These techniques have better behavior in terms of outliers and data variance. With the help of these data processors, the negative impact of irregularities on the accuracy of forecasting can be properly reduced [84]. In pre-processing forecasting method, the data is decomposed into finite numbers of subseries. It is noted that the performance of each subseries is better than the original data. Then, to forecast each subseries, a regressor and feature extractor is developed independently. In the next step, all the subseries is combined, and the forecasting results will be generated. In pre-processing forecasting, Empirical mode decomposition (EMD) and wavelet decomposition (WD) are two of the most widely used methods.

Time series of temperature data and solar irradiance include seasonal long-term behaviors, and daily information. Then, for training, it is suggested that to improve the forecasting model performance, the frequency contents of those signals can be used instead of the signal values. To do this, the forecasting models can be based on WD of the input data [85]. The WD deal with the solar irradiance fluctuations and used to the input data of a forecasting model and, which resulted in the accuracy improvement. Generally, an efficient solution have been made using the wavelet techniques represents resulted in the noise reduction in input datasets before to implement a prediction model [86]. Some recent studies about WD have summarized (based on the time) as follows.

A method combining the ANN and WD for the forecasting of the power output of PV power plants was presented by [87]. The solar irradiance, and meteorological variables such as wind speed humidity, and temperature were chosen as inputs of the ANN model as the ANN can address their nonlinear relationships. Then, WD is used to decompose output power of the PV plant resulted in the separation of the useful information from disturbances. To build models of the decomposed PV output power, ANNs were used. Next, various sky conditions include rainy, sunny, cloudy days and overcast were proposed for validation of model. Finally, the ANN applied to compare the presented method with the traditional forecasting method. The results show that the method needs lesser calculation time with a better forecasting accuracy (MAE of 10.34%, MAPE of 25.37%, and RMSE of 19.66% in rainy days).

Ref [88] was proposed a model called wavelet-coupled support vector machine (W-SVM) in global incident solar radiation forecasting based on the minimum and maximum temperature, sunshine hours, evaporation, precipitation and wind speed as the predictor variables. The merit of the W-SVM was benchmarked with the classical SVM model to achieve reliable results. Then, in sixteen months from 01-March-2014 to 30-June-2015 the data were divided into the test (35%) and train (65%) were set for daily forecasting in the three metropolitan stations (Townsville Aero, Cairns Aero and Brisbane City). The forecast was assessed by prediction errors (MAE, RMSE, MAPE and RMSE). Based on the obtained results, the W-SVM model outperformed better than the classical SVM model for daily forecasts using optimum input combinations.

By referring to Ref [89], the merits of wavelet-ANN models for solar radiation was evaluated. Four different architectures of ANN, namely: adaptive neuro-fuzzy inference system (ANFIS), generalized regression neural networks (GRNN), nonlinear autoregressive recurrent exogenous neural network (NARX) and multilayer perceptron (MLP), were used. The for the decomposition of the complex meteorological signals into relatively simple parts a wavelet analysis was used by using wavelet sub-series, and WD transformation algorithm. The ANN models were used to model the wavelet sub-series and reconstructed to estimate the original signal. Then, to model the global horizontal irradiation over Abu Dhabi city, four meteorological parameters were used including temperature, wind speed, relative humidity, and sunshine duration. The proposed approach was compared to ANN models and validated using different metrics such as R^2 , t-statistics RMSE, MAPE, and MBE. The results confirmed the proposed model improved the performance of the ANN with a maximum 6.84% in R^2 for MLP meanwhile GRNN had a minimum of 2.78% RMSE.

In [90], the information from raw data with better time–frequency resolutions was extracted and the WD was

applied with a bias compensation Random Forest (BCRF) to minimize the prediction error. First, eight decomposition layers in a stationary WD on all raw input features was conducted. Then, there was the lower frequency part of the original signal due to the higher level of decomposition. Next, to train a random forest regression model, the wavelet components and time index as input features were used. In fact, an additional model created by BCRF for prediction of bias to minimize the overall prediction error. Finally, Wavelet – BCRF technique was evaluated by some error metrics like RMSE, MAE and MAPE.

In another study [91], a multi-level WD based on day-ahead solar irradiance forecasting method was proposed. Initially, based on the weather conditions, the daily solar irradiance series were classified into different patterns. Next, the solar irradiance of the next day 24 h was forecasted for each weather pattern using decomposed data series at different WD levels. Then, to fuse the predictions into the final forecasting output, a data-driven fusion model corresponding to the weather pattern was applied. Simulations showed that the forecasting accuracy using different WD level data depends on the weather conditions (sunny, cloudy, and rainy days). In sunny days, solar irradiance reached its maximum level, and the day irradiance curves were relatively smooth but in cloudy and rainy days the value of solar irradiance was limited to lower level and there were more fluctuations on the irradiance curve. To evaluate the forecasting accuracy, two error indexes of RMSE and MAE were used. Generally, the forecasting accuracies all showed the trend of first increasing and then decreasing with WD.

It is usually difficult to design and implement forecasting for non-stationary and non-linear signals. In these cases, The Fourier decomposition can be used, but it does not give information about the time scale characteristics of the data. For this purpose, and in order to obtain the time-scale (time–frequency) information of the signal, we need a method that can extract the intrinsic modes embedded in the signal [92]. The EMD method is proposed for this purpose. In fact, the EMD is a signal analysis method that can indicate at what moment, at what frequency, and with what intensity it is present in a signal [93]. By EMD, there is the decomposition of complex signal into a linear combination of a limited number of intrinsic mode functions (IMFs) with different frequencies. Hence, each of the decomposed IMF components contains local characteristic signals of different time scales from the original signal [94]. Some recent studies about EMD have been summarized as follows.

In [95], a hybrid EMD and back-propagation neural network (BPNN) model was developed for photovoltaic power forecasting. Then, each IMF and each residue were used to

train and test the BPNN individually after decomposition of the time series data by EMD. The proposed EMD-BPNN model was evaluated with PV power output time series data. These data were collected from grid-connected photovoltaic power plants situated in Ghaziabad India at the 100-kW roof-top. Finally, for the performance evaluation of the developed model, the data set was divided into the four evaluation parameters (the symmetric mean absolute percentage error (sMAPE), MSE, MAPE and RMSE) and weekly data groups (W1-W4) with different forecast horizon of 1, 12 and 24 h ahead. Results indicate that the data decomposition greatly reduces the complexity and evaluation time of the back-propagation neural network.

In another study, a hybrid model with combination of ensemble empirical mode decomposition (EEMD) and variable weights was proposed to overcome the demerits of the EMD-ANN model. Therefore, for decomposition of the original PV power generation data, EEMD was used to obtain the residual component (RC) and multiple IMF components (IMF_1, IMF_2). Then, residual components and IMF components were divided into low-frequency, intermediate-frequency, and high-frequency sequences. Then for the prediction of these three sequences, variable-weight combination forecasting (VWCF) method was used. The final prediction results were calculated by summation of the three forecasting results. The total number of input variables were 13 but the first six influential variables were top net solar rad, surface thermal rad down, temperature, surface pressure, time, and relative humidity. RF algorithm was used for the determination of each variable impact. Then, MAE and MSE were used for the evaluation of the forecasting results. Based on the results, the prediction accuracy of each model (in terms of MAE) was reduced [96].

In [97], the pre-processing technique (EMD) was presented to decompose the data. The solar power output using the hybrid design of the SVR model was used the improved feature selection algorithm that resulted in the selection of the best input for the next processing. To improve the accuracy of the model, the proposed model design was set based on the SVR with PSO optimization. The results showed that the proposed algorithm performed better with an average of 14.55 (%) of MAPE and 0.95 (%) of nRMSE.

In [98], several multiscale decompositions in methods of time series analysis for one-hour global solar radiation. Initially, they calculated the time series of the Clear Sky Index. Then, EMD and EEMD methods are used to decompose obtained time series data. Next, the data was forecasted using a linear model and nonlinear models using the time scale fast fluctuation components. Finally, the results were

improved with a combined hybrid model using globally multiscale decomposition.

In [98], several multiscale decompositions in methods of time series analysis for one-hour global solar radiation. Initially, they calculated the time series of the Clear Sky Index. Then, EMD and EEMD methods are used to decompose obtained time series data. Next, the data was forecasted using a linear model and nonlinear models using the time scale fast fluctuation components. Finally, the results were improved with a combined hybrid model using globally multiscale decomposition. Table 3, which summarizes the above, is in the appendix.

4.2.2 Post-Processing

In general, PV data have time-series measurements [99]. In post-processing as a cooperative ensemble forecasting method, the forecasting will be performed using the time series consecutively by two or more predictors. Moreover, a time series data may have more than one characteristic which is suitable for one specific method. For instance, ANN is usually applied for modeling non-linear time series while ARIMA model is suitable for modeling linear time series [38]. The main advantage of post-processing methods is the improvement in the bias of the global solar irradiance forecasts without the need for a long-term historical data database [100]. Based on post-processing method, there are several cooperative ensemble forecasting models such as ARIMA-ANN, ARIMA-SVM, ARMA_ANN, ARIMA-GARCH, SARIMA-ANN, and ARMA-GARCH.

Based on the information, post-processing methods are often used to optimize the output of NWP models. It should be noted that detailed local weather characteristics generally cannot be resolved by NWP predictions hence the spatial resolution has been grown recently. By using the post-processing method, the forecasting will be improved slightly by consideration of the uncertainty of some forecasts even though most of the forecasting methods contain statistical components. Therefore, the final forecasting of data using this method will be a single value that can be valid for future studies [101].

In a study by David et al., two models of ARMA and GARCH were combined to provide probabilistic forecasts of solar irradiance. Meanwhile, to provide a framework that can be applied in an operational context easily, a recursive estimation of the parameters of the models has been set up. As they found, higher forecast accuracy has been found by the proposed method (ARMA and GARCH) in comparison to other machine learning-based techniques. In addition, this model gave additional information about the uncertainty of the forecasts which was easier to set up [102].

A solar radiation forecasting time series model was proposed for multi-hour forecasting (915 h ahead) and a small-scale solar radiation database (30- and 1-s scales) for 1 day (47,000 s ahead). In the first step, ARMA was used to predict future values of the global solar radiation time series. Then, a nonlinear autoregressive (NAR) neural network was applied for prediction purposes due to the nonlinearity in solar radiation time series. The results showed that the ARMA- NAR combination had better accuracy. The NRMSE of the hybrid model was equal to 0.2034 compared to 0.3241 for ARMA model and NRMSE equal to 0.2634 for NAR model [103]. An innovative hybrid model was proposed in three different cities in Morocco for forecasting the daily global solar radiation. There were three steps for estimation including 1) evaluating the linear aspect of the problem by ARIMA model, and 2) building an ANN model to model the residuals of the ARIMA. It was estimated that the error conditions for the ARIMA model can be forecasted by output from ANN model. The findings showed that the hybrid model (ARIMA-ANN) was more accurate in terms of MAPE, R2, RMSE, MBE, NRMSE, and TS [104]. Table 2 shows the findings of recent studies that use ensemble methods for solar forecasting.

5 Conclusion

In general, the reliability of solar power systems is affected by the dynamic nature of solar irradiance. Changes in sunlight intensity led to voltage and power fluctuations in solar power plants and disruption of power systems. A good way to deal with such problems is to predict solar irradiance. Accurate forecasting is challenging and involves a variety of methods statistical, physical and ensemble forecasting methods. This paper has reviewed recently published studies (2015–2021) on solar irradiance forecasting using ensemble models including competitive and cooperative forecasting methods. The former has been divided into data diversity and parameter diversity and the latter has been categorized according to pre-processing and post-processing. In this paper, recent articles have been discussed according to each category. It can be concluded that ensemble models perform better than standalone ones. However, a hybrid models have a more complex structure, but they provide better accuracy. Based on the previous studies, Artificial Neural Network (ANN) and Space Vector Machine (SVM) are widely used with ensemble models (WD-ANN, EMD-BPNN and W-SVM) due to their ability in solving complex and non-linear forecasting models. Also, the performance of the ensemble models has been evaluated by some error metrics such as RMSE and MAE. It has been indicated that

EMD is more powerful, and it has more accuracy than WD in solar forecasting methods.

Moreover, in this article, the different model inputs, and their effects on the prediction of solar radiation have been discussed. Common inputs are solar irradiance, atmospheric and module temperature, wind speed and direction, and humidity. Among these, solar irradiance is most positively correlated with PV output. Solar irradiance is positively

correlated with temperature and negatively correlated with wind speed. Other weather variables have low correlation values.

Appendix

See Table 3.

Table 3 Summary of recent studies for solar forecasting using perspectives

Ref	Time ahead	Input variables	Output variable	Perspectives	Forecasting method	Error	Comparison
[53]	Daily	–Wind Speed –Wind Direction –Air pressure	Wind power	Data diversity	Bagging	RMSE= 10.6%	Bagging > ANN > k-NN
[61]	Daily	–Temperature –Air Pressure –Wind	PV power	Data diversity	Random Forest	MAPE= 8.5%	RF > GBDT
[63]	Daily	–Temperature –Relative humidity –Cloud cover –Precipitation	PV power	Data diversity	Boosting	RMSE= 9.48%	Boosting > AR
[65]	Hour	–Wind Speed	Wind power	Data diversity	Ada Boosting	MAE= 1.0581 MAPE= 7.63%	AB > ANN
[76]	Daily	–PV power	PV power	Parameter Diversity	Time series	MSE= 16.24	TS > SARIMA
[77]	Hour	–Irradiation	PV power	Parameter Diversity	SVR	MAE= 37.04	SVR > NAM > SP
[81]	Hour	–Irradiation –Temperature	Purchased PV power	Parameter Diversity	Polynomial Regression	MAPE= 10.51%	–
[87]	Daily	– Humidity – Temperature – Wind Speed	PV power	Pre-processing	WD-ANN	RMSE= 19.663% MAE= 10.349%	WD-ANN > ANN
[88]	Daily	– Sunshine hours – Temperature – Wind speed— evaporation— precipitation	Global solar irradiation	Pre-processing	W-SVM	RMSE= 2.317 MJ m^{-2} MAE= 1.819 MJ m^{-2}	W-SVM > SVM
[90]	6 h	– Solar irradiance – Temperature – PV output	PV power	Pre-processing	WD-BCRF	RMSE= 32.12% MAE= 20.64%	WD-BCRF > WD-SVM > RF
[96]	1 h	– surface thermal rad down – Top net solar rad – surface pressure – Humidity –Temperature	PV power	Pre-processing	EEMD-VWCF	RMSE= 0.0107 MAE= 0.0622	EEMD-VWCF > BPNN > RF
[95]	1,12 and 24 h	–	PV power	Pre-processing	EMD-BPNN	RMSE= 0.019 MAPE= 5.47%	EMD-BPNN > BPNN
[38]	10 min	– Irradiation	PV power	Post-processing	ANN	RMSE= 6%	ANN > ARIMA

Acknowledgements This work was supported by the Korea Electric Power Corporation (KEPCO) as “Study of Distribution Market Operator (DMO) Platform and Electricity Market” under Grant R19X002-02.

Funding Konkuk University Researcher Fund, 2021, (A)Duehee Lee, Korea Electric Power Corporation (KEPCO), R19X002-02, (A)Duehee Lee.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Yagli GM, Yang D, Srinivasan D (2019) Automatic hourly solar forecasting using machine learning models. *Renew Sustain Energy Rev* 105:487–498
2. Vyas S et al. (2022) Forecasting solar power generation on the basis of predictive and corrective maintenance activities. *arXiv preprint arXiv:2205.08109*
3. (IRENA) I.R.E.A. (2020) *Renewable Capacity Statistics 2020*. Abu Dhabi.
4. Zervos A (2015) *Renewables 2015 Global Status Report*.
5. REN21 (2021) *Renewables 2021 Global Status Report*. (Paris: REN21 Secretariat)
6. Notton G et al (2018) Intermittent and stochastic character of renewable energy sources: consequences, cost of intermittence and benefit of forecasting. *Renew Sustain Energy Rev* 87:96–105
7. Wang F et al (2020) A minutely solar irradiance forecasting method based on real-time sky image-irradiance mapping model. *Energy Convers Manage* 220:113075
8. Voyant C et al (2017) Machine learning methods for solar radiation forecasting: a review. *Renew Energy* 105:569–582
9. Van der Meer DW, Widén J, Munkhammar J (2018) Review on probabilistic forecasting of photovoltaic power production and electricity consumption. *Renew Sustain Energy Rev* 81:1484–1512
10. Yang D, van der Meer D (2021) Post-processing in solar forecasting: ten overarching thinking tools. *Renew Sustain Energy Rev* 140:110735
11. Zhang G et al (2022) Solar forecasting with hourly updated numerical weather prediction. *Renew Sustain Energy Rev* 154:111768
12. Ahmed R et al (2020) A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization. *Renew Sustain Energy Rev* 124:109792
13. Mellit A et al (2020) Advanced methods for photovoltaic output power forecasting: a review. *Appl Sci* 10(2):487
14. Kumari P, Toshniwal D (2021) Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance. *J Clean Prod* 279:123285
15. Hong T et al (2020) Energy forecasting: a review and outlook. *IEEE Open Access J Power Energy* 7:376–388
16. Das UK et al (2018) Forecasting of photovoltaic power generation and model optimization: a review. *Renew Sustain Energy Rev* 81:912–928
17. Singla P, Duhan M, Saroha S (2022) A comprehensive review and analysis of solar forecasting techniques. *Front Energy* 16(2):187–223
18. Wang F et al (2019) Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting. *Energy Convers Manage* 181:443–462
19. Chen C et al (2011) Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Sol Energy* 85(11):2856–2870
20. Gao M et al (2019) Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using LSTM. *Energy* 187:115838
21. Behera MK, Majumder I, Nayak N (2018) Solar photovoltaic power forecasting using optimized modified extreme learning machine technique. *Eng Sci Technol, Int J* 21(3):428–438
22. AISkaif T et al (2020) A systematic analysis of meteorological variables for PV output power estimation. *Renew Energy* 153:12–22
23. Kumar DS et al (2020) Solar irradiance resource and forecasting: a comprehensive review. *IET Renew Power Gener* 14(10):1641–1656
24. Kumari P, Toshniwal D (2021) Deep learning models for solar irradiance forecasting: a comprehensive review. *J Clean Prod* 318:128566
25. Pedro HT et al (2018) Assessment of machine learning techniques for deterministic and probabilistic intra-hour solar forecasts. *Renew Energy* 123:191–203
26. Fernández-Peruchena CM et al (2015) Increasing the temporal resolution of direct normal solar irradiance series in different climatic zones. *Sol Energy* 115:255–263
27. Lai CS et al (2021) A deep learning based hybrid method for hourly solar radiation forecasting. *Expert Syst Appl* 177:114941
28. Vignola F et al (2016) Effects of changing spectral radiation distribution on the performance of photodiode pyranometers. *Sol Energy* 129:224–235
29. Das UK et al (2017) SVR-based model to forecast PV power generation under different weather conditions. *Energies* 10(7):876
30. Sun Y et al. (2015) Research on short-term module temperature prediction model based on BP neural network for photovoltaic power forecasting. In: 2015 IEEE power & energy society general meeting. 2015
31. Raza MQ, Nadarajah M, Ekanayake C (2016) On recent advances in PV output power forecast. *Sol Energy* 136:125–144
32. Reddy SS (2017) Optimal scheduling of thermal-wind-solar power system with storage. *Renew Energy* 101:1357–1368
33. Gupta A, Gupta K, Saroha S (2021) A review and evaluation of solar forecasting technologies. *Mater Today: Proc* 47:2420–2425
34. Kim I (2017) Master Markov chain Monte Carlo and acceptance-rejection algorithms for synthesising short-term variations in the generation output of the photovoltaic system. *IET Renew Power Gener* 11(6):878–888
35. Panamtash H, Zhou Q (2018) Coherent probabilistic solar power forecasting. In 2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). 2018. IEEE
36. Rodríguez F et al (2018) Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control. *Renew Energy* 126:855–864
37. Sobri S, Koohi-Kamali S, Rahim NA (2018) Solar photovoltaic generation forecasting methods: a review. *Energy Convers Manage* 156:459–497

38. Gürtürk M, Ucar F, Erdem M (2022) A novel approach to investigate the effects of global warming and exchange rate on the solar power plants. *Energy* 239:122344
39. Li B, Zhang J (2020) A review on the integration of probabilistic solar forecasting in power systems. *Sol Energy* 210:68–86
40. Guermoui M et al (2020) A comprehensive review of hybrid models for solar radiation forecasting. *J Clean Prod* 258:120357
41. Polikar R (2012) Ensemble learning in ensemble machine learning: methods and applications. Zhang C, Ma Y, Eds. Springer: Berlin
42. Basaran K, Özçift A, Kılınc D (2019) A new approach for prediction of solar radiation with using ensemble learning algorithm. *Arab J Sci Eng* 44(8):7159–7171
43. Diagne M et al (2013) Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renew Sustain Energy Rev* 27:65–76
44. Arora I, Gambhir J, and Kaur T (2020) Solar irradiance forecasting using decision tree and ensemble models. In: 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA). 2020. IEEE
45. Bracale A, Carpinelli G, De Falco P (2016) A probabilistic competitive ensemble method for short-term photovoltaic power forecasting. *IEEE Trans Sustain Energy* 8(2):551–560
46. Ren Y, Suganthan P, Srikanth N (2015) Ensemble methods for wind and solar power forecasting—A state-of-the-art review. *Renew Sustain Energy Rev* 50:82–91
47. He Y, Wang Y (2021) Short-term wind power prediction based on EEMD–LASSO–QRNN model. *Appl Soft Comput* 105:107288
48. Gensler A, and Sick B (2016) Forecasting wind power—an ensemble technique with gradual cooperative weighting based on weather situation. In: 2016 International Joint Conference on Neural Networks (IJCNN). 2016. IEEE
49. Ribeiro MHDM et al (2022) Efficient bootstrap stacking ensemble learning model applied to wind power generation forecasting. *Int J Electr Power Energy Syst* 136:107712
50. Breiman L (1996) Bagging predictors. *Mach Learn* 24(2):123–140
51. Vaish J, Datta SS, and Seethalekshmi K (2020) Short term load forecasting using ANN and ensemble models considering solar irradiance. In: 2020 International Conference on Electrical and Electronics Engineering (ICEE3). 2020. IEEE
52. Ribeiro MHDM, dos Santos Coelho L (2020) Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Appl Soft Comput* 2020(86):105837
53. Srivastava R, Tiwari A, Giri V (2019) Solar radiation forecasting using MARS, CART, M5, and random forest model: a case study for India. *Heliyon* 5(10):e02692
54. Munawar U, Wang Z (2020) A framework of using machine learning approaches for short-term solar power forecasting. *J Electr Eng Technol* 15(2):561–569
55. Ali M et al (2021) Variational mode decomposition based random forest model for solar radiation forecasting: new emerging machine learning technology. *Energy Rep* 7:6700–6717
56. Meenal R et al (2022) Weather forecasting for renewable energy system: a review. *Arch Comput Methods Eng*. <https://doi.org/10.1007/s11831-021-09695-3>
57. Shaik, AB and Srinivasan S (2019) A brief survey on random forest ensembles in classification model. In: International Conference on Innovative Computing and Communications. 2019. Springer
58. Huertas Tato J, Centeno Brito M (2018) Using smart persistence and random forests to predict photovoltaic energy production. *Energies* 12(1):100
59. Meng M, Song C (2020) Daily photovoltaic power generation forecasting model based on random forest algorithm for north China in winter. *Sustainability* 12(6):2247
60. Ahmad MW, Mourshed M, Rezgui Y (2018) Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression. *Energy* 164:465–474
61. Persson C et al (2017) Multi-site solar power forecasting using gradient boosted regression trees. *Sol Energy* 150:423–436
62. Taieb SB, Hyndman RJ (2014) A gradient boosting approach to the Kaggle load forecasting competition. *Int J Forecast* 30(2):382–394
63. Liu H et al (2015) Comparison of four Adaboost algorithm based artificial neural networks in wind speed predictions. *Energy Convers Manage* 92:67–81
64. Fan J et al (2018) Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: a case study in China. *Energy Convers Manage* 164:102–111
65. Kim H, Lee D (2021) Probabilistic solar power forecasting based on bivariate conditional solar irradiation distributions. *IEEE Trans Sustain Energy* 12(4):2031–2041
66. Park J et al (2020) Multistep-ahead solar radiation forecasting scheme based on the light gradient boosting machine: a case study of Jeju Island. *Remote Sens* 12(14):2271
67. Tyrallis H, Papacharalampous G (2021) Boosting algorithms in energy research: a systematic review. *Neural Comput Appl* 33(21):14101–14117
68. Choi S, Hur J (2020) An ensemble learner-based bagging model using past output data for photovoltaic forecasting. *Energies* 13(6):1438
69. Lee J et al (2020) Reliable solar irradiance prediction using ensemble learning-based models: a comparative study. *Energy Convers Manage* 208:112582
70. Mitrentsis G, Lens H (2022) An interpretable probabilistic model for short-term solar power forecasting using natural gradient boosting. *Appl Energy* 309:118473
71. McCandless T, Dettling S, Haupt SE (2020) Comparison of implicit vs explicit regime identification in machine learning methods for solar irradiance prediction. *Energies* 13(3):689
72. Sharma A, Kakkar A (2018) Forecasting daily global solar irradiance generation using machine learning. *Renew Sustain Energy Rev* 82:2254–2269
73. Moon J et al. (2021) A Comparative analysis of tree-based models for day-ahead solar irradiance forecasting. In 2021 International Conference on Platform Technology and Service (PlatCon). 2021
74. Yang D, Dong Z (2018) Operational photovoltaics power forecasting using seasonal time series ensemble. *Sol Energy* 166:529–541
75. Kaur A et al (2016) Benefits of solar forecasting for energy imbalance markets. *Renew Energy* 86:819–830
76. Flynn H, Larsen G (2022) Investigating the application of Kalman Filters for real-time accountancy in fusion fuel cycles. *Fusion Eng Des* 176:113037
77. Yang D (2019) On post-processing day-ahead NWP forecasts using Kalman filtering. *Sol Energy* 182:179–181
78. Nilam B, Ram ST (2022) Forecasting geomagnetic activity (Dst Index) using the ensemble kalman filter. *Monthly Notices Royal Astron Soc* 511(1):723–731
79. Takeda H (2017) Short-term ensemble forecast for purchased photovoltaic generation. *Sol Energy* 149:176–187
80. Jiranantacharoen P, and Benjapolakul W (2019) Photovoltaic power generation forecast by using estimator model and kalman filter. In 2019 International Conference on Power, Energy and Innovations (ICPEI). 2019. IEEE

81. Hajirahimi Z, Khashei M (2022) Hybridization of hybrid structures for time series forecasting: a review. *Artif Intell Rev*. <https://doi.org/10.1007/s10462-022-10199-0>
82. Bokde N, Feijóo A, Kulat K (2018) Analysis of differencing and decomposition preprocessing methods for wind speed prediction. *Appl Soft Comput* 71:926–938
83. De Giorgi MG et al (2015) Error analysis of hybrid photovoltaic power forecasting models: a case study of mediterranean climate. *Energy Convers Manage* 100:117–130
84. Malvoni M, De Giorgi MG, Congedo PM (2017) Forecasting of PV Power Generation using weather input data-preprocessing techniques. *Energy Procedia* 126:651–658
85. Zhu H et al (2015) A power prediction method for photovoltaic power plant based on wavelet decomposition and artificial neural networks. *Energies* 9(1):11
86. Deo RC, Wen X, Qi F (2016) A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. *Appl Energy* 168:568–593
87. Hussain S, Alalili A (2017) A hybrid solar radiation modeling approach using wavelet multiresolution analysis and artificial neural networks. *Appl Energy* 208:540–550
88. Chiang P-H et al. (2017) Forecasting of solar photovoltaic system power generation using wavelet decomposition and bias-compensated random forest. In 2017 Ninth Annual IEEE Green Technologies Conference (GreenTech)
89. Zhen Z et al (2018) Multi-level wavelet decomposition based day-ahead solar irradiance forecasting. In 2018 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT). 2018. IEEE
90. Prasad R et al (2019) Designing a multi-stage multivariate empirical mode decomposition coupled with ant colony optimization and random forest model to forecast monthly solar radiation. *Appl Energy* 236:778–792
91. Zhang N et al (2022) Short-term PV output power forecasting based on CEEMDAN-AE-GRU. *J Electr Eng Technol* 17:1–12
92. Bokde N et al (2019) A review on hybrid empirical mode decomposition models for wind speed and wind power prediction. *Energies* 12(2):254
93. Yadav HK, Pal Y, Tripathi MM (2020) Short-term PV power forecasting using empirical mode decomposition in integration with back-propagation neural network. *J Inf Optim Sci* 41(1):25–37
94. Wang H, Sun J, Wang W (2018) Photovoltaic power forecasting based on EEMD and a variable-weight combination forecasting model. *Sustainability* 10(8):2627
95. Zhang W, Dang H, Simoes R (2018) A new solar power output prediction based on hybrid forecast engine and decomposition model. *ISA Trans* 81:105–120
96. Monjoly S et al (2017) Hourly forecasting of global solar radiation based on multiscale decomposition methods: a hybrid approach. *Energy* 119:288–298
97. Acharya SK, Lee J (2018) An Ensemble Learning Method for Solar Power Forecasting. In: Proceedings of Symposium of the Korean Institute of communications and Information Sciences, vol 8, no 3, pp 1310–1311
98. Blaga R et al (2019) A current perspective on the accuracy of incoming solar energy forecasting. *Prog Energy Combust Sci* 70:119–144
99. Yang B et al (2021) Classification and summarization of solar irradiance and power forecasting methods: a thorough review. *CSEE J Power Energy Syst*. <https://doi.org/10.17775/CSEEJPES.2020.04930>
100. David M et al (2016) Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models. *Sol Energy* 133:55–72
101. Mukaram MZ, Yusof F (2017) Solar radiation forecast using hybrid SARIMA and ANN model: a case study at several locations in Peninsular Malaysia. *Malays J Fundam Appl Sci Spec Issue Some Adv Ind Appl Math* 13:346–350
102. Belmahdi B, Louzazni M, Bouardi AE (2020) A hybrid ARIMA–ANN method to forecast daily global solar radiation in three different cities in Morocco. *Eur Phys J Plus* 135(11):1–23

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Negar Rahimi she is currently studying for a Ph.D. degree at Konkuk University. Her research interests are electricity market and solar power forecasting.

Sejun Park he is currently studying for a master's degree at Konkuk University's AI Energy System Laboratory. His research interests are electric vehicles, ESS, and prediction.

Wonseok Choi he has been in Konkuk University AI system Laboratory since 2020. His Majors are system operation optimization analysis and algorithm production.

Byoungryul Oh he received the B.S. and M.S. degrees in Electrical Engineering from Konkuk University, Seoul, South Korea in 2019 and 2022 respectively. His research interests include data science and distributed optimization utilized in distribution system and electricity market.

Sookyung Kim she received the B.S. and M.S. degrees in electrical engineering from Konkuk University, Seoul, South Korea, in 2019 and 2021 respectively. Her research interests include forecasting and optimization utilized in power systems, smart grids, and electricity market.

Youngho Cho he received the B.S. degree from Konkuk University, Seoul, Korea, in 2022. He is currently working toward the Ph.D. degree from The University of Texas at Austin, TX, USA. His research interests include machine learning and data analytics for renewable energy, power system operations, and energy management systems.

Sunghyun Ahn he is currently studying at Konkuk University's AI Energy System Laboratory. His research interests are power system flexibility and real-time power system evaluation.

Chulho Chong he is currently studying for B.S. degree in Electrical and Electronics Engineering from Konkuk university. His research interests are machine learning, solar power forecasting, and capacity market.

Daewon Kim he is currently studying for B.S. degree in Computer Engineering from Konkuk university. His research interests are machine learning, solar power forecasting, and reinforcement learning.

Cheong Jin he received B.S degree in electrical engineering from Pohang University of Science and Technology, Pohang, South Korea, in 2006. His research interests are digital twin and artificial intelligence.

Duehee Lee he received the B.S. degree in electronic and electrical engineering from the Pohang University of Science and Technology, Pohang, South Korea, in 2004, and the M.S. and Ph.D. degrees in electrical and computer engineering from The University of Texas at Austin, Austin, TX, USA, in 2009 and 2015, respectively. He is currently an associate professor with the Electrical Engineering Department, Konkuk University, Seoul, South Korea.